



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE COMUNICAÇÃO E EXPRESSÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INGLÊS: ESTUDOS LINGÜÍSTICOS E
LITERÁRIOS

MANUELLA HINCHINGR PEREIRA

Developing an L2 Pronunciation Assessment Rubric for Classroom Use in Brazil:
Teachers' Insights on Criteria, Performance Levels, and Usability

FLORIANÓPOLIS

2026

Manuella Hinchings Pereira

Developing an L2 Pronunciation Assessment Rubric for Classroom Use in Brazil:
Teachers' Insights on Criteria, Performance Levels, and Usability

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Inglês da Universidade Federal de Santa Catarina como requisito parcial para obtenção do título de Mestra em Inglês: Estudos Linguísticos e Literários.

Orientadora: Prof^ª. Dr^ª. Rosane Silveira

Florianópolis

2026

Pereira, Manuella Hinchinger

Developing an L2 pronunciation assessment rubric for classroom use in Brazil : teachers' insights on criteria, performance levels, and usability / Manuella Hinchinger Pereira ; orientadora, Rosane Silveira, 2026.

159 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Comunicação e Expressão, Programa de Pós-Graduação em Inglês: Estudos Linguísticos e Literários, Florianópolis, 2026.

Inclui referências.

1. Inglês. 2. Avaliação da pronúncia em segunda língua. 3. Desenvolvimento de rubricas. 4. Avaliação em sala de aula. 5. Ensino de inglês. I. Silveira, Rosane. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Inglês: Estudos Linguísticos e Literários. III. Título.

Manuella Hinchinger Pereira

**Developing an L2 Pronunciation Assessment Rubric for Classroom Use in Brazil:
Teachers' Insights on Criteria, Performance Levels, and Usability**

O presente trabalho em nível de Mestrado foi avaliado e aprovado, em 10 de março de 2026,
pela banca examinadora composta pelos seguintes membros:

Profa. Gladys Quevedo-Camargo, Dra.

Universidade de Brasília (UnB)

Prof. Germán Zárate-Sánchez, Dr.

Western Michigan University

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado
adequado para obtenção do título de Mestra em Inglês: Estudos Linguísticos e Literários.



Documento assinado digitalmente
Alinne Balduino Pires Fernandes
Data: 30/03/2026 15:12:40-0300
CPF: ***.588.801-**
Verifique as assinaturas em <https://v.ufsc.br>

Coordenação do Programa de Pós-Graduação



Documento assinado digitalmente
Rosane Silveira
Data: 30/03/2026 13:12:44-0300
CPF: ***.936.839-**
Verifique as assinaturas em <https://v.ufsc.br>

Profa. Rosane Silveira, Dra.

Orientadora

Florianópolis, 2026.

ACKNOWLEDGEMENTS

First of all, I want to express my deepest gratitude to my supervisor, Prof. Rosane Silveira, whose confidence in me as a novice researcher made this thesis possible. Your professionalism, expertise, and kindness will stay with me as I continue my academic journey. It goes without saying that the more I learn from you, the more inspired I feel to become a better professional and person. Thank you for your constant support, for patiently replying to my long e-mails filled with endless questions, for reading my work with such care, and for offering guidance and words of encouragement during the most challenging moments of this process.

I am grateful to the other professors whose contributions were essential to the development of this thesis. In particular, I would like to thank Prof. Hanna Kivistö de Souza, from whom I learned so much about research methods and statistics in Applied Linguistics. Thank you for making every class engaging and memorable, especially through your creative (and much appreciated) examples involving cats. I am extremely thankful to Prof. Gladys Quevedo-Camargo and Prof. Germán Zárate-Sánchez for reading my work and offering important suggestions for improving it. Receiving feedback from researchers I deeply admire in my two favorite research areas was truly an honor.

I am thankful to my parents, Silvana and Francisco, for always believing in the transformative power of education and for doing everything within their reach to support my dream of dedicating myself to my studies. À minha mãe, obrigada por ser minha primeira professora e me ensinar o que é amor e dedicação a tudo aquilo que fazemos. Você é minha maior inspiração e, se sou quem sou hoje, é porque tive você para me guiar. Obrigada pelas incontáveis xícaras de chá de camomila que fez para mim, noite após noite, desde o começo do mestrado. O que de fato me acalmava não eram elas, mas sim sua presença e afeto.

I want to thank my friends for being such an essential part of my life, inside and outside academia, and for making my days more joyful. I am especially grateful to Elisa Eller, who supported and encouraged me since the very first day I shared my intention to apply for the Master's graduate program. I hope you know how much I value our friendship and all the adventures we have shared (and those yet to come). Nicolý Antunes, Leticia Batista, and Samantha Souza, thank you for being incredible women and friends who inspire me daily. I am also grateful for my colleagues from NUPFFALE, whose company made this journey less lonely, especially Gustavo Duarte and Crystoffer Zanchet. Thank you for our coffee meetings full of giggling moments and (sometimes) reflections on our studies.

Finally, I would like to express my gratitude to the English Graduate Program (PPGI) and to UFSC for the opportunity to pursue my academic studies. In addition, I want to thank FAPESC (Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina) for the financial support I received through a scholarship during the two years this piece of research was conducted, which was fundamental to its completion. Last but not least, a special thank you to all the amazing teachers who participated in the data collection and kindly contributed to the development of this study. Words may not be enough to express my gratitude to all of you.

“It may turn out that it is not rubrics per se (that is, rubrics as an assessment tool in a particular form), but the provision of focused learning goals, criteria, and performance descriptions in whatever form that supports learning and motivational outcomes for students. Even so, the fact that rubrics are an efficient, clear, and easily understood way to focus learning goals, criteria, and performance descriptions would recommend their use as one of, or even the primary, form in which to do this.”

(Susan Brookhart and Fei Chen, 2014)

RESUMO

Avaliar a pronúncia de aprendizes de língua inglesa como segunda língua, em contextos formais de ensino, pode ser desafiador para professores devido a lacunas em sua formação docente, ao tempo limitado para lidar com os componentes fonético-fonológicos da língua e às dificuldades em definir o que e como avaliar o progresso dessa habilidade linguística. Ainda assim, quando orientada por uma perspectiva formativa, a avaliação da pronúncia possui um papel importante ao fornecer informações sobre o desenvolvimento da produção oral de estudantes, as quais são capazes de orientar decisões pedagógicas mais sensíveis às necessidades identificadas ao longo do processo de ensino-aprendizagem. Nesse contexto, ao descreverem com clareza os critérios e os níveis de desenvolvimento esperados em cada momento da aprendizagem da pronúncia, rubricas podem contribuir para tornar a avaliação dessa habilidade mais transparente e confiável. Diante disso, este estudo visou desenvolver uma rubrica de avaliação formativa da pronúncia em segunda língua voltada à sala de aula e analisar as percepções de professores brasileiros de inglês em exercício sobre esse instrumento de modo a aprimorá-lo. A pesquisa adotou uma abordagem de métodos mistos e contou com a participação de 13 professores brasileiros de inglês em exercício. Os dados foram coletados por meio de três questionários destinados a investigar: (1) características sociodemográficas e experiências acadêmico-profissionais; (2) formação docente em avaliação de línguas e práticas de avaliação da pronúncia; e (3) percepções sobre a versão preliminar da rubrica, utilizada pelos participantes para avaliar três amostras de produção oral de aprendizes brasileiros, bem como para analisar a rubrica quanto aos seus critérios, níveis de desempenho e usabilidade, considerando a realidade educacional brasileira. Os dados quantitativos e qualitativos foram analisados por meio de estatística descritiva e análise de conteúdo, respectivamente. Segundo as sugestões dos professores, diversos aspectos da rubrica poderiam ser melhorados. Por exemplo, o número de níveis de desempenho da rubrica foi ampliado de quatro para seis e alguns descritores foram reformulados. De modo geral, a rubrica mostrou-se útil para apoiar a avaliação formativa da pronúncia e contribuiu para ampliar a conscientização dos professores acerca de aspectos específicos dessa habilidade. Além disso, os participantes tiveram uma visão positiva sobre o seu potencial de uso em sala de aula, apesar de destacarem limitações relacionadas ao tempo necessário para sua aplicação e à compreensão do instrumento por parte dos estudantes. Por fim, o estudo contribui para a área de desenvolvimento de rubricas e para a compreensão das necessidades de professores brasileiros de inglês em exercício no que se refere à avaliação da pronúncia em sala de aula.

Palavras-chave: Avaliação da pronúncia em segunda língua; Desenvolvimento de rubricas; Avaliação em sala de aula; Ensino de inglês.

ABSTRACT

Assessing the pronunciation of learners of English as a second language in formal teaching contexts can be challenging for teachers due to limitations in their teacher education, restricted time to address the phonetic-phonological components of the language, and difficulties in defining what and how to assess students' linguistic development. Nevertheless, from a formative perspective, pronunciation assessment can provide meaningful information about students' oral production learning, which can inform more sensitive pedagogical decisions throughout the teaching-learning process. In this context, as rubrics clearly describe the criteria and performance levels of each stage of pronunciation development, they may contribute to a more transparent and reliable assessment of this ability. Therefore, the present study aimed to develop a second language pronunciation rubric designed for formative classroom assessment and to investigate Brazilian in-service English teachers' perceptions of this instrument in order to improve it. The study involved 13 Brazilian in-service English teachers and adopted a mixed-methods approach. Data were collected through three instruments: (1) a background questionnaire, (2) an assessment survey designed to understand teachers' professional preparation in language assessment and their classroom pronunciation assessment practices, and (3) a rubric survey, in which teachers used the preliminary version of the rubric to assess selected speech samples and evaluated the instrument in relation to its criteria, performance levels, and usability within the Brazilian educational context. Quantitative and qualitative data were analyzed through descriptive statistics and content analysis, respectively. According to Brazilian English teachers' suggestions, several aspects of the rubric could be improved. For instance, the number of performance levels was expanded from four to six, and some level descriptors were reformulated. The findings reveal that the rubric was useful in supporting the formative assessment of pronunciation and contributed to increasing teachers' awareness of specific features of this ability. In addition, participants had a positive view of the rubric's potential for classroom use, although they highlighted limitations related to the time required for its implementation and students' ability to understand the instrument. Finally, the study contributes to the field of rubric development and to understanding the needs of in-service Brazilian English teachers regarding pronunciation assessment in the classroom.

Keywords: Second language pronunciation assessment; Rubric development; Classroom-based assessment; English teaching.

LIST OF FIGURES

Figure 1 – Rating process in performance-based assessment.....	27
Figure 2 – Holistic versus analytic scoring in rubrics	29
Figure 3 – Educational contexts where participants are currently working	38
Figure 4 – Age groups taught by participants.....	38
Figure 5 – Proficiency levels at which Brazilian teachers assess pronunciation.....	64
Figure 6 – Criteria used by Brazilian teachers to assess students’ pronunciation	65
Figure 7 – Instruments Brazilian teachers use to assess students’ pronunciation	67
Figure 8 – Importance Brazilian teachers give to pronunciation assessment in students’ grade	68
Figure 9 – Challenges Brazilian teachers face when assessing pronunciation in the classroom	70
Figure 10 – Brazilian teachers’ intended uses of the pronunciation assessment rubric	87
Figure 11 – Teachers’ suggestions on how to improve the pronunciation rubric for the Brazilian context	88

LIST OF TABLES

Table 1 – Participants’ background	36
Table 2 – Brookhart’s (2014) desired characteristics to write high quality criteria for formative rubrics	41
Table 3 – Brookhart’s (2014) desired characteristics to write high quality performance levels descriptors for formative rubrics	43
Table 4 – The preliminary version of the L2 pronunciation assessment rubric	47
Table 5 – Mapping specific student L2 pronunciation difficulties in rubric criteria.....	52
Table 6 – Brazilian teachers’ prior education in classroom-focused assessment and testing ..	60
Table 7 – Brazilian teachers’ prior education in assessment and testing purposes	62
Table 8 – Brazilian teachers’ education in content and concepts of language assessment and testing.....	63
Table 9 – Mean scores assigned to learners across the six pronunciation criteria of the rubric	72
Table 10 – Summary of participants’ perception of the rubric criteria	73
Table 11 – Summary of participants’ perception of the rubric performance levels and descriptors	78
Table 12 – The revised version of the L2 pronunciation assessment rubric	91

TABLE OF CONTENTS

1 INTRODUCTION	13
1.1 OBJECTIVES.....	15
1.1.1 Main objective.....	15
1.1.2 Specific objectives	15
1.2 RESEARCH QUESTIONS	15
1.3 SIGNIFICANCE OF RESEARCH	16
1.4 ORGANIZATION OF THE THESIS	16
2 REVIEW OF LITERATURE.....	18
2.1 LANGUAGE ASSESSMENT LITERACY, CLASSROOM-BASED ASSESSMENT AND FORMATIVE ASSESSMENT: AN OVERVIEW	18
2.2 SECOND LANGUAGE PRONUNCIATION ASSESSMENT.....	22
2.3 PERFORMANCE-BASED ASSESSMENT, SCALES/RUBRICS AND SCORING METHODS	26
2.3.1 Studies on Rubrics for Second Language Pronunciation Assessment.....	31
3 METHOD	35
3.1 STUDY CHARACTERIZATION	35
3.2 ETHICS REVIEW BOARD.....	35
3.3 PARTICIPANTS	36
3.4 INSTRUMENTS AND MATERIALS FOR DATA COLLECTION.....	39
3.4.1 Approach to Designing the Rubric.....	39
3.4.2 Rubric Criteria	40
3.4.3 Rubric Performance-Levels and Descriptors.....	43
3.4.4 Rubric Supporting Guide and Glossary	45
3.4.5 The Preliminary Version of the L2 Pronunciation Assessment Rubric	46
3.4.6 Background Questionnaire.....	52
3.4.7 Assessment Survey.....	53
3.4.8 Rubric Survey	54
3.5 PROCEDURES FOR DATA COLLECTION	56
3.6 PROCEDURES FOR DATA ANALYSIS.....	58
4 RESULTS AND DISCUSSION	60
4.1 BRAZILIAN ENGLISH TEACHERS' EDUCATION IN LANGUAGE ASSESSMENT AND CLASSROOM PRONUNCIATION ASSESSMENT PRACTICES	60
4.2 BRAZILIAN TEACHERS' PERCEPTION OF THE PRELIMINARY L2 PRONUNCIATION RUBRIC	71
4.2.1 Consistency in Scoring Results Using the Rubric.....	71
4.2.2 Rubric criteria.....	72
4.2.3 Rubric performance levels and descriptors	76
4.2.4 Rubric usability.....	82
<i>4.2.4.1 Ease of use</i>	<i>82</i>
<i>4.2.3.2 Reported challenges.....</i>	<i>84</i>
<i>4.2.3.3 Classroom applicability</i>	<i>86</i>

<i>4.2.3.4 Suggestions to improve the rubric and the supporting guide for the Brazilian context</i>	87
4.3 THE REVISED VERSION OF THE L2 PRONUNCIATION ASSESSMENT RUBRIC	89
5 CONCLUSIONS	98
5.1 PEDAGOGICAL IMPLICATIONS	100
5.2 LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH	102
REFERENCES	104
APPENDIX A - INFORMED CONSENT FORM	112
APPENDIX B - BACKGROUND QUESTIONNAIRE	117
APPENDIX C - ASSESSMENT SURVEY	119
APPENDIX D - RUBRIC SURVEY	124
APPENDIX E - SUPPORTING GUIDE	132
APPENDIX F - MESSAGES SENT TO PARTICIPANTS	156
APPENDIX G - INVITATION VIA SOCIAL MEDIA	158
APPENDIX H - DATA SPREADSHEETS	159

1 INTRODUCTION

Classroom assessment serves not only to measure students' development, but to enhance teaching and learning as well, thus assuming a central role in second language¹ (L2) education (Arter; McTighe, 2001; Scaramucci, 2006). When teachers are able to accurately collect and interpret evidence of students' progress, they can provide pedagogical support at all instances of the learning process and adapt teaching practices to address the identified needs. Nonetheless, as Earl (2003) argues, managing the position of an assessor in the classroom is extremely complex. Teachers are often expected to handle multiple, and at times conflicting, responsibilities at once (e.g., valuing students' individual identities and progress paths while presenting quantitative information for institutional accountability reports). Despite this great demand, many of them report feeling insufficiently prepared to carry out fair and effective assessment practices, a scenario that could be attributed to limited education and scarce opportunities for continuous professional development in language assessment (Brindley, 2001; Fulcher, 2012; Taylor, 2013; Turner; Purpura, 2016; Quevedo-Camargo; Scaramucci, 2018).

Under these circumstances, assessing distinct L2 linguistic abilities in foreign language settings represents an even greater challenge for teachers. Within these abilities, pronunciation is seen as especially difficult to assess in formal teaching contexts (Isaacs, 2018). Nevertheless, developing students' oral proficiency is essential in today's society, considering it prepares them to become competent L2 users² (Celce-Murcia *et al.*, 2010) when negotiating diverse communicative demands in personal, professional or social contexts. In the classroom, pronunciation assessment³ enables teachers and students to identify which phonological features have been successfully acquired and which still require more attention in order to enhance intelligibility (Levis, 2018), giving guidance on aspects that should be a priority during instruction. Consequently, it can provide pedagogical evidence to inform educational decisions

¹ In this thesis, no distinctions will be made between the terms 'second' and 'foreign' language.

² According to Cook (2007, p. 238), the "[...] unspoken assumption is that successful L2 learners are those that become like first language speakers". Nonetheless, scholars in the field of applied linguistics and second language acquisition seem to have moved past this postulation. Instead, contemporary perspectives perceive L2 proficiency as a "multicompetence" in which a learner is able to employ the language effectively and appropriately, considering distinct communicative settings and situations (Spada, 2007).

³ Some authors may attribute different meanings to what comprises assessing L2 pronunciation. I will adhere to Jones and Isaacs' (2021) broad definition of pronunciation assessment, which includes any procedure of methodically gathering data about a learner's pronunciation in order to define or reach a conclusion concerning his or her pronunciation ability. Thus, for instance, pronunciation assessment can be used independently for production and perception in the L2 classroom, as part of a diagnostic analysis, or associated as one scoring component of the speaking skill in oral tests.

involving from curriculum adaptation to the selection of resources and teaching strategies that are sensitive to learners' specific difficulties in pronunciation-related components.

While pronunciation assessment has great potential to support teaching and learning, in practice, it usually receives limited attention in L2 curricula and classrooms (Isbell; Sakai, 2022). According to Isbel and Sakai (2022), this can be attributed to several factors, including the time-consuming nature of oral assessment for teachers and the difficulty to adopt instruments that are reliable, valid, and effective for different pedagogical contexts. Additionally, as a result of the aforementioned limited professional preparation, L2 teachers often rely on intuition, prior experiences or personal preferences when judging students' oral performance, which may introduce differences to assessment results that are unrelated to the construct in focus (Knoch, 2017). Therefore, it is important to warrant classroom assessment practices that are not grounded on favoritism for a given pronunciation model or accent, but on theoretical and empirical findings that support assessing speech dimensions that are actually pertinent to communication and intelligibility. Adopting assessment instruments that lead teachers to more consistent and fair decisions about learners' development seems to be a promising response to these challenges.

For L2 pronunciation classroom assessments particularly, rubrics could be of great value. From a learning-oriented perspective, a rubric must include three components: "1. a set of criteria aligned to learning goals; 2. a performance scale typically consisting of three or four levels, and 3. descriptors that differentiate the levels of performance on the scale" (McTighe; Brookhart; Guskey, 2024, p. 2). Arter and McTighe (2001) state that rubrics define quality, foster descriptive assessment, clarify performance expectations, guide instructional efforts, support monitoring of learners' progress, encourages self-assessment, improve grading consistency, and suggest pathways for responsive action. There is extensive evidence in the literature that, when well-designed and implemented appropriately, rubrics can function as valuable assessment instruments to enhance the teaching-learning process (Arter; McTighe, 2001; Brookhart, 2024; Panadero; Jonsson, 2013). Nevertheless, the educational potential of rubrics depends on how prepared its users are and on the constant refinement of descriptors to guarantee usability (Harding, 2017).

In the existing literature, most studies investigating especially L2 pronunciation assessment rubrics have focused only on how they are used in to assess global measures such as intelligibility and comprehensibility (e.g., Martins; Silveira, 2025) or how to develop them for higher education classroom settings, with students who have distinct first languages (see Isaacs; Trofimovich; Foote, 2018). Consequently, there is still limited empirical evidence on

how to develop rubrics for formative purposes aimed at regular classroom, particularly in Brazil. As a starting point, it is necessary to investigate in-service teachers' perspectives on L2 pronunciation assessment rubrics, in order to comprehend how to develop and implement these instruments to better meet the needs they might report regarding instructional and professional constraints. To address this issue, the present study aims to create a preliminary analytic rubric for L2 pronunciation assessment and to refine it considering evaluations from in-service Brazilian English teachers.

1.1 OBJECTIVES

1.1.1 Main objective

To develop a preliminary analytic rubric to support in-service Brazilian English teachers to conduct formative classroom-based assessment of students' L2 pronunciation.

1.1.2 Specific objectives

1. To design an analytic L2 pronunciation rubric for formative classroom use in the Brazilian educational context;
2. To provide in-service Brazilian English teachers with training on how to use the preliminary version of the analytic rubric;
3. To collect in-service Brazilian English teachers' perceptions of the analytic rubric in terms of its criteria, performance levels, and usability;
4. To revise the preliminary L2 pronunciation analytic rubric considering in-service Brazilian English teachers' evaluations.

1.2 RESEARCH QUESTIONS

Considering the objectives previously established, two research questions (RQ) were formulated in order to guide this study:

RQ1: How do Brazilian teachers of English evaluate an L2 pronunciation assessment rubric in relation to its criteria, performance levels, and usability?

RQ2: How can the experiences and insights emerging from teachers' evaluation inform a revised version of the L2 pronunciation assessment rubric?

1.3 SIGNIFICANCE OF RESEARCH

The study is significant for three reasons. First, the field of L2 classroom pronunciation assessment has only recently started to receive scholarly attention (Isbell; Sakai, 2022). As a consequence, several aspects of pronunciation assessment conducted in small-scale contexts remain underexplored in research agendas, particularly pre-service and in-service teachers' education, cognition, and practices. Second, as Harding (2017) argues, developing assessment instruments that are empirically-grounded and co-designed with their intended users may improve the quality and reliability of L2 pronunciation assessment processes. Third, assessing students' oral production in the classroom, especially pronunciation, is extremely time-consuming and requires a lot of professional preparation from teachers to identify the phonological components worth addressing (Celce-Murcia *et al.*, 2010). Therefore, there is a need for well-designed and validated assessment instruments that, concurrently, were designed with the aid of its end-users so they are practical and accessible for diverse pedagogical contexts and can guide even teachers with limited experience in assessing learners' pronunciation in the classroom. In attempting to provide teachers with a rubric for this specific purpose, it may encourage them to effectively include pronunciation in classroom assessment practices and to promote more consistent and impartial assessment of students' L2 pronunciation. Ultimately, this research may collaborate to the expansion of pronunciation assessment studies in the Brazilian context, especially with regard to L2 teacher education and rubric development for oral production in classroom settings.

1.4 ORGANIZATION OF THE THESIS

This thesis is organized into five chapters: (1) Introduction, (2) Review of Literature, (3) Method, (4) Results and discussion, and (5) Conclusions. The first introductory chapter presented the main theme of the study and the research questions that will guide it. The second chapter addresses important discussions and findings from the fields of language assessment literacy, classroom-based assessment, pronunciation assessment, and rubric development, which provide the theoretical foundation for the study. The third chapter describes the methodological procedures adopted, including the developmental process of the preliminary

version of the L2 pronunciation assessment rubric and details of the study design, such as participants, instruments, and procedures for data collection and analysis. The fourth chapter presents and discusses the findings of the study in light of in-service Brazilian English teachers' evaluation of the rubric's criteria, performance levels, and usability. Additionally, it presents the revised version of the rubric after incorporating participants' suggestions. Finally, the last chapter delineates the conclusions drawn from the main findings and briefly discusses the pedagogical implications, limitations, and suggestions for future research.

2 REVIEW OF LITERATURE

This chapter presents the literature review that supports the rationale and existence of the present study. The first section introduces the concept of Language Assessment Literacy and gives an overview of what constitutes classroom-based and formative assessment. The second section will address some issues related to L2 pronunciation assessment, especially in regard to how complex it can be to be implemented in classroom contexts and the possible positive and negative impact it can have on teachers and learners due to factors such as language attitudes, assessment criteria, and so on. The third section considers the difference between scales and rubrics and describes design and scoring methods for rubric development. In the end, it narrows the focus to previous studies that have addressed rubrics for L2 pronunciation assessment, which is the study's main interest.

2.1 LANGUAGE ASSESSMENT LITERACY, CLASSROOM-BASED ASSESSMENT AND FORMATIVE ASSESSMENT: AN OVERVIEW

William (2013) perceives assessment as a “bridge” between teaching and learning. That is because, according to the author, assessment is the only way teachers can determine whether the instructional practices they had effectively led to the learning outcomes they originally intended students to achieve. Nonetheless, as Fulcher (2012, p. 125) points out, deciding “why, how, and when testing and assessment should (or should not) take place” is still a challenge for teachers within L2 classrooms. Therefore, the field of Language Assessment Literacy (LAL), with regard to teachers' assessment beliefs, principles and practices, has begun to receive considerable attention in the literature (Giraldo, 2021). LAL concerns the expertise required to design, adapt and implement valid and reliable language assessment procedures across distinct pedagogical contexts (Inbar-Lourie, 2008). As teachers are expected to measure students' ongoing learning in an L2, there is a growing need for them to develop LAL in order to apply effective and sensitive language assessment in the classroom (Vogt; Tsagari; Spanoudis, 2020).

Different scholars have attributed different definitions to what composes LAL. Historically, Stiggins (1991) was the first author to mention the concept of assessment literacy. In his definition, assessment literate teachers are prepared to hold “a basic understanding of the meaning of high- and low- quality assessment and are able to apply that knowledge to various measures of student achievement” (p. 535). On the other hand, Inbar-Lourie (2008) defined what it meant to be *language* assessment literate, arguing that assessing learners' linguistic knowledge requires an additional expertise concerning language-specific components.

Therefore, assessment literacy provides teachers with the ability to have general competencies related to assessment (e.g., designing a rubric and providing feedback), while language assessment literacy integrates this ability with knowledge regarding language acquisition and development (e.g., linguistic proficiency and abilities, etc.).

In Inbar-Lourie's (2013) words, LAL

[...] encompasses not only familiarity with tools and procedures for evaluating students' language abilities, but additional components, particularly the ability to render apt feedback so as to effectively direct learners toward setting and reaching future learning targets. Furthermore, assessment literate professionals are also aware of the ethical aspects involved in the assessment process and of the possible ramifications of putting assessment-based decisions into practice. (Inbar-Lourie, 2013, p. 1)

As teachers are responsible for measuring and indicating students' proficiency in the L2, there is an ever-increasing need for them to be acquainted and master the expertise that concerns language assessment. Consequently, developing LAL is of utter importance for them to be able to implement local assessment procedures inside the classroom, considering the particularities of the context they permeate and the specificities of assessing language learning. According to Bachman and Palmer (1996), each one of the four linguistic abilities in an L2 urges for distinct, specific assessment approaches and consequently demands considerable preparation from teachers. Yet, there is a great deal to be achieved when it comes to concrete, extensive LAL from them so that they are fully prepared and acquainted with such approaches. The aforementioned situation can be attributed to several factors, including how the area of language assessment is neglected in both pre-service and in-service teacher education.

The scarcity of attention given to the field of language testing⁴ and assessment throughout teacher education programs is widely discussed in the literature, with several authors arguing that LAL should become a new paramount dimension of it (Giraldo, 2021; Vogt; Tsagari, 2014). When focusing especially on the Brazilian scenario, Scaramucci (2016) has a similar concern and stresses that much more needs to be accomplished. For instance, in order to investigate the curricula of federal institutions' undergraduate courses from Brazil, Quevedo-Camargo (2020) analyzed how many disciplines dealt with language assessment and were offered for pre-service English teachers. She found that most of them were elective courses provided to all undergraduate students together regardless of the individual teaching degree they were pursuing. Additionally, they tended to deal with assessment in a one-size-fits-all

⁴ According to Isbell and Sakai (2022, p. 03), "all tests are assessments, [but] not all assessments are tests". Thus, in this thesis, assessment refers to "[t]he process of collecting information to make an informed judgment for some purpose", while testing concerns any "pre-planned form of assessment that uses a standard format [...] in controlled conditions and typically yields numeric scores".

manner. This is a limitation because it disregards the specificities of language assessment and, as a result, underestimates the necessity to develop teachers' assessment literacy tailored to L2 pedagogy.

Further research conducted with pre-service L2 teachers from Brazil preparing to complete teacher practicum found that they felt unprepared to assess students and implement different types of assessment procedures in the classroom (Damacena; Quevedo-Camargo, 2021). While experience acquired with time is expected to aid and improve in-service teachers' assessment practices (Giraldo, 2021), previous studies (e.g., Fulcher, 2012; Scarino, 2017; Vogt; Tsagari; Spanoudis, 2020) present a parallel scenario, indicating that experienced teachers needs regarding assessment range from wanting more initiatives to foster LAL to hands-on practical moments to analyze, use and create assessment instruments that are context-responsive to different educational settings. Therefore, it is possible to conclude that pre-service and in-service teachers have limited experience with assessment and may need more support, either through continuous education opportunities or access to high-quality assessment instruments that they are easily implement in their classrooms.

Now that some of the issues regarding LAL and teacher education have been presented, it seems important to turn the discussion to the sphere of classroom-based assessment (CBA), which is at the heart of L2 teachers' responsibility and occupies a considerable amount of time in daily pedagogical practices. In contrast to large-scale assessment involving high-stake exams applied to a wide population of learners in decisive conditions, CBA encompasses assessment procedures that improve and inform teaching that intends to foster learning (Fox; Abdulhamid; Turner, 2012). For Hamp-Lyons (2007, p. 497), the purposes of large-scale and classroom assessment are opposite, once the former "[...] seeks the general, the common, the group identifier, the scaleable, the replicable, the predictable, the consistent, and the characteristic", while the latter "[...] seeks the special, the individual, the changing, the changeable, the surprising, the subtle, the textured, and the unique".

Thus, from a learning-oriented perspective, L2 teachers' attention during CBA should be directed towards the action of guaranteeing that all learners thrive to their fullest potential. In light of that, over the past decades, research in the realm of language assessment, particularly centered on CBA, experienced a significant transformation (Hamp-Lyons, 2007). Presently, a growing number of authors (Andrade, 2010; Bachman, 2013; Black; William, 1998; Earl, 2003; Scaramucci, 1999; William, 2013) started to advise that CBA should adopt formative practices, moving away from the perception of assessment with only a summative end. This perception shift challenges the well-established assumption that assessment serves to merely measure

students' knowledge, but presupposes that it can fundamentally advance it. In the Brazilian context, this distinction is equally evidenced in the *Base Nacional Comum Curricular*, a policy document primary for education in our country, and which will be examined in-depth in the next section.

Generally, formative assessment describes “where the learner is right now, where the learner needs to be, and how to get there” (William, 2013, p. 16). It can be used to elicit evidence of students' progress, guide teachers' decisions, adapt lesson planning and curricula, and ultimately better direct students towards development (Earl, 2003; Andrade, 2010). Therefore, instead of only attributing a grade to learners as in summative assessment, formative assessment is concerned with increasing student learning. Additionally, Gu and Yu (2020) include another important stage in CBA for formative purposes: students' engagement with the assessment cycle. For the authors, it is imperative that the assessment procedures and criteria used are transparent to students, enabling them to comprehend learning achievements and initiate effective change towards improvement with the support of the teacher. Given that formative assessment is concerned with the function that assessment assumes in the educational process, it can possess the form of formal tests or non-graded activities, as long as it provides meaningful, accurate and timely feedback for students and teachers (Luckesi, 2011).

Despite the pedagogical potential of CBA for formative purposes, this stance on assessment may not always be reflected in real classroom contexts. In practice, teachers may have difficulties implementing the principles that underlie formative assessment with students due to the tensions between having to fulfill institutional requirements and valuing learners' individual necessities, in addition to other issues related to limited professional opportunities to develop LAL (Scaramucci, 2016), as discussed previously. In Brazil, product-oriented assessment practices still remain dominant and, consequently, can be an issue during the implementation of formative, learning-oriented assessment. Thus, even in cases where official documents and L2 teachers recognize the importance of formative assessment to support learning, its implementation depends on institutional support, pedagogical preparation on language assessment and access to appropriate assessment instruments (Freitas *et al.*, 2009).

These difficulties do not affect only how assessment is implemented, they are capable of impacting how assessment influences teachers and learners' experience as well. In this regard, Turner (2001) discusses the widespread impact that assessment exerts on all educational processes. In the classroom context, every decision a teacher has concerning assessment, may yield positive or negative outcomes, producing what it is called as *washback*. Issues in one or more stages of the assessment cycle (including, for example, teachers' limited LAL or

discriminatory criteria) could lead to a negative washback that not only affects the fairness of assessment, but the beliefs and behaviors of learners about what they are capable to do and achieve. From an ecological perspective, Chong and Isaacs (2023) argue that assessment dynamics transcends the school boundaries and have vast social implications for learners. As an extremely complex process, L2 pronunciation assessment within classroom settings may intensify negative washback if not conducted properly. Accordingly, the following section explores the field of L2 pronunciation assessment.

2.2 SECOND LANGUAGE PRONUNCIATION ASSESSMENT

According to Celce Murcia *et al.* (2010), L2 pronunciation is primary to effective oral language communication. To define it, the construct of pronunciation includes segmental and suprasegmental phonological dimensions of speech. While the former consists of vowel and consonant sounds, the latter includes larger prosodic features such as stress, intonation, and rhythm (Isaacs, 2018). With respect to pronunciation assessment, throughout the decades, different second language acquisition theories and pedagogical approaches have affected how oral proficiency has been perceived in language curricula and large-scale exams (Kang; Kermad, 2018). In the past, pronunciation teaching and research used to focus on accuracy, but more recently, it has moved away from a native-centered model and started embracing the notion of intelligibility (Levis, 2005). An intelligibility perspective to pronunciation assumes that learners' speech production simply needs to be understandable to listeners, which introduces a more attainable pronunciation standard for students and aid to delineate priority areas for instruction and assessment.

While reviewing the state-of-the-art of the L2 pronunciation assessment field, Isaacs (2018) reports that it has begun to receive more attention only in the last decade after a long period of neglect in research focusing on large-scale exams and classroom contexts. According to the author, the leading motivation for this change was the large evidence (for a complete review see Saito; Plonsky, 2019) that pronunciation can develop with pedagogical intervention and that linguistic competence involved conveying one's message clearly, which is impracticable without phonological learning. For large-scale pronunciation assessment, important progress was achieved with the CEFR⁵ (Council of Europe, 2018) phonological

⁵ The Common European Framework of Reference for Languages (CEFR; Council of Europe, 2018) is an international European high-stake assessment framework used to indicate learners' L2 proficiency levels.

control scale, a reference instrument to assess oral proficiency that was updated to allude fundamentally to the intelligibility construct. Whereas studies of pronunciation assessment involve largely L2 speech global dimensions (i.e., intelligibility, comprehensibility and accentedness) and the use of statistics to ensure the reliability and generalizability of a certain test, CBA pronunciation research regarding teachers' cognition and practices, in addition to empirically-developed and validated assessment instruments to support pedagogical practice, are still under-investigated (Isbell; Sakai, 2022).

Considering that classroom contexts are unique and distinct from other assessment settings, several challenges emerge when attempting to implement effective pronunciation assessment for L2 students at this level. One of these challenges is the impracticability of relying on large-scale tests for learning- and diagnosis-oriented assessment practices. Isbell and Sakai (2022) argue that scoring systems usually used in large-scale tests are designed for specific purposes (i.e., placement), which is why they are often incompatible with the classroom assessment needs. Additionally, as the authors point out, it is unrealistic to expect teachers to reproduce them in everyday pedagogical practices. They stress that, although these two assessment systems may share certain similarities (e.g., they aim at gathering evidence of a learner's pronunciation in order to assign them to a particular proficiency level), they have different purposes. For instance, the primary concern of classroom assessment is pedagogical impact, given that in this context "[...] beneficial impacts of assessment use on teaching and learning (i.e., positive washback) are valued over reliability, and beneficial impacts are considered central to the validity of assessment practices" (Isbell; Sakai, 2022, p. 5).

With regard to L2 pronunciation, CBA supports a wide range of assessment procedures (e.g., different tasks and types of assessment) and enables teachers to provide learners with more accurate analyses of strong aspects in their oral production, as well as areas in need of development throughout the entire delicate process of acquiring a second phonological system. Moreover, through CBA it is possible to reconcile individual pedagogical support for learners with curricular demands and institutional requirements (Isbell; Sakai, 2022). Nonetheless, at the same time, teachers are responsible for planning, implementing and interpreting all phases of the assessment process. This includes deciding on a series of elements that may have educational impact and that require significant LAL development, as well as knowledge of the body of research and pedagogical practices of L2 pronunciation assessment. For example, according to Celce-Murcia *et al.* (2010), teachers must determine which oral features need to be assessed, given the specific group of learners, their L1, age, proficiency level, syllabus, time available, resources at disposal, and even the impositions of external agents. In addition to that,

they are expected to assess a great extent of speech samples, provide continuing evaluation moments with formative feedback, administer tests and establish the criteria through which learners' progress will be elicited and interpreted.

Given the huge responsibility and teachers have to carry out CBA, personal experience and judgments inevitably become part of the assessment process. While L2 pronunciation assessment for research purposes frequently relies on human ratings due to the high level of reliability they present during assessment, Kang and Kermad (2018) note that raters may influence the assessment quality and result as a consequence of preconceived notions, beliefs and preferences of what is desirable in one's speech. For instance, in the classroom, despite research having already placed intelligibility at the core of pronunciation instruction (see Levis, 2018), L2 teachers' inclination towards nativeness could still influence the teaching and assessment approaches implemented when they chose only high-prestige varieties as pronunciation models, thus contradicting well-established pedagogical recommendations. In a similar note, Leung (2013) argues that teachers' prior experiences and language-related attitudes are determining factors in the selection of criteria and procedures for L2 assessment at the classroom level. This scenario could create a discrepancy between the dimensions that should theoretically guide pronunciation assessment and what occurs in practice, thereby promoting unproductive and unfair outcomes that may increase the chances of negative washback (Isbell; Sakai, 2022).

Considering the previous discussion, it is important to understand how these issues are treated in the Brazilian context. In this regard, *Base Nacional Comum Curricular* (BNCC; Brasil, 2018) is the official document that has been guiding English teaching in regular education since 2018. To illustrate, formal education in Brazil is compulsory from the ages of six to seventeen and it is organized into two instructional stages in both public and private institutions: nine years of elementary education and three years of secondary education (Brasil, 1996). However, English is only a mandatory subject from the 6th grade onwards. Concisely, BNCC encourages formative assessment practices and places the acquisition of the communicative competence as critical in the era of English as a *Lingua Franca*. Despite that, although the document states that pronunciation, intonation, and stress should be considered in students' oral linguistic development (Brasil, 2018), it does not provide clear definitions of these constructs or offer explicit pedagogical guidance to teachers regarding how and according to which criteria they should be assessed.

Further intensifying the matter, several studies in the field show that Brazilian teachers feel unprepared to teach L2 pronunciation (Camargo, 2020; Costa, 2016; Haus, 2018), which is

anticipated to affect other domains of classroom practice as well, including how they see and treat the assessment of speech components. In this sense, the apparent inequality in teachers' education concerning pronunciation instruction coupled with a dearth of pedagogical guidelines may create space for partiality during the assessment process of students' oral production. In light of this scenario, Quevedo-Camargo and Garcia (2017) investigated teachers' assessment procedures during an L2 speaking test in a Brazilian federal institute. The authors found that teachers did not have any collectively defined criteria, as their interpretation of and emphasis on specific constructs (such as pronunciation) varied significantly. This resulted in contrasting scores assigned to students within the same proficiency level, which reveals important issues in assessment reliability. Accordingly, Nagle, Sachs and Zárate-Sández (2020) highlight the need of more professional development initiatives for L2 teachers focused on pronunciation assessment within CBA, so that teachers can be informed on theory when implementing assessment practices and adopting assessment instruments.

In classroom contexts, defining clear learning outcomes through the use of predefined criteria may contribute to greater transparency and reliability in the assessment process, which is able to foster positive washback for teachers and learners (Arter; McTighe, 2001). When assessing complex abilities such as L2 speech and pronunciation, rating scales and rubrics are one of the most investigated assessment instruments (Isaacs; Thomsom, 2013). They have been shown to offer practical means of assessing L2 speech with the reliability necessary. Nevertheless, several factors may affect the extent to which such instruments fulfill their intended purposes, as one of the greatest challenges in assessment concerns assuring that teacher-raters will interpret and use them as they were designed originally (Harding, 2017). Thus, the next section addresses issues related to scale/rubric terminology and development. Additionally, it reviews previous studies concerning scales for L2 pronunciation assessment, with particular attention to claims regarding the need of further research on the validity and usability of these instruments considering teacher-raters' perceptions and local demands.

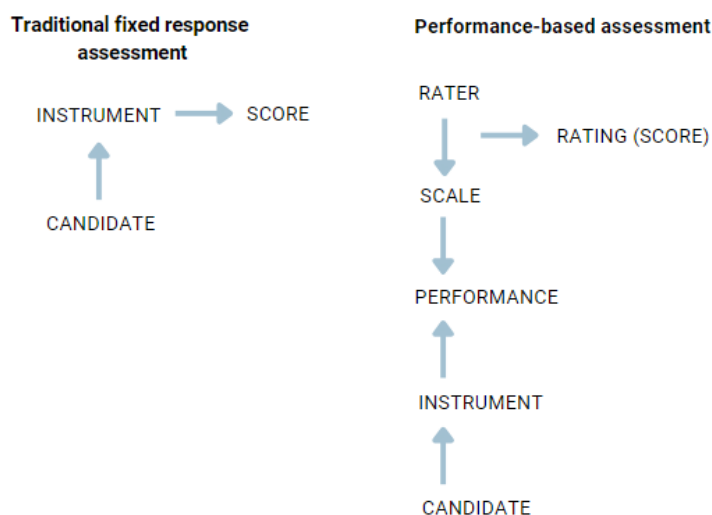
2.3 PERFORMANCE-BASED ASSESSMENT, SCALES/RUBRICS AND SCORING METHODS

With the advancement of communicative approaches to language teaching, scholars in second language acquisition and language assessment began to argue that learners should be assessed under conditions that resembled authentic communicative exchanges in real-world contexts (McNamara, 1996). At that moment, this paradigm shift introduced a more complex

system when compared to earlier scoring methods. According to McNamara (1996, p. 6), this is called performance testing or performance-based assessment. It refers to situations in which “the assessment of the actual performances of relevant tasks are required of candidates, rather than the more abstract demonstration of knowledge, often by means of paper-and-pencil tests”. Drawing on Wigglesworth (2008), Chinda (2013, p. 142) further explains that performance assessment differ from more traditional tests in that “(1) there is a performance by the candidate; (2) the performance is judged using an agreed set of criteria; and (3) there is a degree of authenticity of the assessment tasks”.

As a result of this new necessity, teacher-raters were now required to “[...] make more complicated judgments than the right-wrong decisions in multiple choice, true/false, error-recognition, and other item types of testing situations where the candidate’s responses can be marked as either “correct” or “incorrect”” (Chinda, 2013, p. 144). Within this context, a new dynamic emerged in the assessment process, namely the interaction between the teacher-rater and the rating instrument. In other words, to assess L2 learners’ performance, whether in large-scale assessment contexts or in classroom settings, raters and teachers now require a set of criteria that facilitates them to interpret, describe and indicate performance components necessary in distinct proficiency levels transparently. In most cases, particularly in the assessment of L2 oral production, this intermediary instrument has assumed the form of a scale. Figure 1 visually illustrates the differences between traditional and performance-based assessment approaches in terms of how scoring decisions are reached.

Figure 1 – Rating process in performance-based assessment



Source: Adapted from McNamara (1996, p. 120).

Scales have attracted interest in applied linguistics and have been extensively investigated within the field of L2 testing and assessment due to how versatile they are (Isaacs; Thomson, 2013). It is therefore important to discuss the different uses of the terms ‘rating scale’ and ‘rubric’ in the assessment and classroom-based assessment literature. According to Kuiken and Vedder (2021), scholars and practitioners from the North American tradition tend to use the expression ‘rubric’ more frequently, but it is often employed interchangeably with ‘scale’. In the CEFR (Council of Europe, 2018), an important framework to L2 assessment originated in Europe, the term ‘rating scale’ is adopted. In contrast, in studies concerned with CBA, instruments composed of criteria and descriptors that articulates learners’ success against a particular set of outcomes, are frequently referred to as ‘rubrics’ (see Chappuis *et al.*, 2012). Despite the fact that these terms are sometimes used as synonyms in research, some authors argue that rating scales and rubrics should not be treated as the same assessment instruments (Andrade; 2023; Brookhart; Chen, 2014; Dawson, 2017).

In this regard, Brookhart (2014) conceives rating scales and rubrics as distinct instruments with different purposes and uses. According to the author, a defining feature of rubrics compared to other assessment instruments is its performance level descriptors. In rubrics, the performance level descriptors are specified in a detailed manner, providing teachers with written evidence to justify why a particular student was placed at a given level considering their performance, as well as guidance on how learners can progress from one level to another (e.g., from “good” to “excellent”). Conversely, rating scales tend to rely on general descriptions that focus more on frequency or quality and they do not offer the same degree of detail needed to support users’ final assessment decision. Rating scales may be useful in other assessment contexts, but Brookhart (2014, p. 80) advocates for the use of rubrics in formative assessment practices, as they do not merely evaluate performance but, more importantly, “move learning forward”. In line with this perspective, the present piece of research (which aims to develop a classroom-based assessment instrument for formative purposes) adopts the term rubric. Nonetheless, throughout the literature review, the terms scale and rubric will be used in accordance with the terminology employed in each original study.

There are several issues concerning the design of rating scales and rubrics that need to be addressed. One of the first regards the two distinct types of scoring methods: holistic and analytic. As Arter and McTighe (2001) postulate, in holistic scoring, teachers are expected to assess learners’ performance as a single, unified entity. Thus, they assign only one score according to a general impression of what that learner was able to accomplish. In comparison, analytic scoring involves the separate assessment of multiple relevant performance dimensions,

as each criterion is appraised independently. The authors argue that these two scoring systems serve different intentions. For instance, holistic scoring in rubrics promotes faster evaluation that may be more suitable for simple or summative assessments, considering teachers only account for overall impressions. However, despite offering a faster analysis and being more practical for certain contexts, holistic rubrics have been criticized because of their restricted and controlled structure, particularly for classroom assessment oriented towards formative purposes, in which descriptive information is necessary and desired (Brookhart, 2014).

In contrast, analytic scoring rubrics may offer more advantages for teachers and learners in CBA contexts. To begin with, they are capable of generating more reliable assessment results once learners' performance is assessed in separate dimensions, which tends to lead to more consistent evaluations. Additionally, as a result of how comprehensive they are, analytic rubrics provide more precise and detailed information regarding students' positive and negative attributes in a given linguistic ability and, therefore, teaching practices can be adapted to address students' needs and improve learning. Furthermore, they further specify the elements that constitute quality in L2 use for students, teachers, parents, and other educational stakeholders (Arter; McTighe, 2001). Nevertheless, Chinda (2013) states that, because teachers are required to judge L2 performance across distinct criteria, analytic rubrics tend to demand more time for scoring and, at least initially, may be perceived as more complex and less easy to use compared to other assessment instruments available. Figure 2 exemplifies the differences between holistic and analytic rubrics.

Figure 2 – Holistic versus analytic scoring in rubrics

Holistic Rubric	
Performance level 5	
Performance level 4	
Performance level 3	✓
Performance level 2	
Performance level 1	

Analytic Rubric			
Performance levels	Criterion 1	Criterion 2	Criterion 3
5		✓	
4			
3	✓		
2			✓
1			

Source: Adapted from Arter and McTighe (2001, p. 25).

The second issue concerns the distinction between general (or generic) and task-specific rubrics (Arter; McTighe, 2001). As the names suggest, task-specific rubrics are

designed to assess students' performance on a single, well-defined task, whereas general rubrics can be used to assess complex abilities in a variety of tasks. For instance, a pronunciation assessment rubric can be considered a general rubric, as it may be applied to the assessment of learners' oral production in presentations, conversations, monologues, oral exams, and other speaking activities. The authors argue that task-specific rubrics may be useful when the primary intention is summative assessment. On the contrary, general rubrics are more effective in that they increase students' awareness of what constitutes quality learning and performance. Furthermore, Arter and McTighe (2001) point out that general rubrics encourage students to apply what they learned from one task to the succeeding ones, which is aligned with the principles of formative assessment. In addition, they can be collectively implemented in order to foster greater consistency in how students' performance is assessed among teachers of the same pedagogical context.

The third issue concerns how scales are developed (Chinda, 2013). In intuitively-developed scales, the main features of the instrument are defined according to what the developer considers most relevant for each criterion and descriptor, drawing on theoretical knowledge of the field and findings from previous research. Inversely, empirically-developed scales are evidence-driven, that means that they are outlined after actual performance samples of L2 learners performing tasks that the scale is intended to assess. In both instances, usually experienced teachers or assessment specialists assume the role of scale developers and later apply the instrument to its intended users, using the data collected to improve it. Furthermore, Kuiken and Vedder (2021) point out that a third approach to scale development that has been gaining increasing attention in the literature is the mixed-methods. In this approach, scale and rubric development combines qualitative procedures (such as the investigation of experts and end-users' perceptions) with quantitative procedures aimed at verifying inter-rater reliability.

Ultimately, the constructs of validity and reliability are of value to all assessment processes, and particularly so when designing a rubric. According to Roever and McNamara (2006), validity is a "multidimensional concept" that ensures the assessment represents the constructs it plans to measure, covers all the relevant aspects of those constructs, and assists sound interpretations in light of its results to guide pedagogical practices. Reliability, in turn, concerns the "degree to which the results of an assessment are dependable and yield consistent results across raters (interrater reliability), over time (test-retest reliability), or across different versions of the same test (internal consistency or interform reliability)" (Arter; McTighe, 2001, p. 181). There are multiple ways to strengthen the validity and reliability of a rubric. One of them involves the implementation of the instrument and the training of its users, since teachers

and the contexts in which they are part of play a primary role in influencing assessment practices (Kuiken; Vedder, 2021). Thus, a well-designed rubric, combined with sound interpretative practices of its results on the part of teachers, can be decisive in fostering positive washback and promoting better teaching and learning practices.

Beyond theoretical and methodological considerations, research have investigated the effects of rubrics on student learning outcomes. Brookhart (2024) conducted a systematic review to investigate the advantages of using rubrics with learners from basic education. The author analyzed 13 studies in total, in which 11 of them found that the use of rubrics in assessment practice was associated with improved student performance in areas such writing, mathematics, arts, science, and history. The other two studies, which did not find positive effects, introduced rubrics to students only one week before the assignment was due, leaving little or no time for them to become familiar with the instrument, an aspect that may help explain the results. Brookhart (2024) emphasizes that rubrics must be well-designed and, in particular, implemented as an integral part of the educational process to generate pedagogical gains. This includes, for example, sharing rubrics with students in advance and creating opportunities for them to actively engage with the instrument for self- or peer-assessment at different moments of learning.

To date, there appears to be a scarcity of studies that investigate the pedagogical potential of using rubrics for L2 pronunciation assessment in relation to basic education students' learning outcomes. That is, little is known about whether the implementation of a rubric throughout the assessment process actually leads to improvements in learners' oral production when faced with pronunciation-related communication difficulties or complex target sounds in the L2, for instance. Instead, existing research on the matter has focused on the development of rubrics for pronunciation in higher-education or large-scale testing L2 contexts and on the preferences of teacher-raters when using them. In light of this, the next subsection reviews studies that have addressed the use of rubrics in L2 pronunciation assessment, with particular attention to issues concerning design, usability and the perspective of teacher-raters during the refinement process of these instruments.

2.3.1 Studies on Rubrics for Second Language Pronunciation Assessment

When it comes to L2 pronunciation assessment, Isaacs and Thomson (2013) argue that scales can be very advantageous for CBA, since they support evaluating a wide range of oral samples from students across different types of activities and tasks. In addition, they can serve

different pedagogical purposes, such as supporting the provision of detailed feedback and giving clear guidance to assist students during self-assessment moments, for example. Nonetheless, several issues in this area still require further attention, especially when the aim is to develop L2 pronunciation rubrics for local classroom contexts in which teachers themselves act as assessors. In this regard, Zhong (2019) conducted a systematic review of 30 studies on rating scales for L2 pronunciation assessment. The author disclosed important themes that remain underexplored in the literature and divided them in two categories: scale design and scale users.

In the first category, Zhong (2019) identified problems in three domains of scale design. The first concerns the constructs embedded in the scale, as some constructs that are relevant to L2 pronunciation instruction and assessment are still misinterpreted (e.g., intelligibility and comprehensibility) and it is difficult to argue which one is more important to adopt as an aim considering the necessities of different contexts. The second domain involves the selection of scale criteria, given that the decision on which specific phonological features should be assessed remains open to debate, in scholars' views. Finally, the third problem regards the descriptor, length and format of the scale, elements that constitute this instrument and have a primary role in influencing how raters interpret and use it to assess L2 oral performance.

In the second category, Zhong (2019) addresses issues related to the user of rating scales, which includes both raters' characteristics and instrument usability. Regarding the former, the author explains how individual traits, personal perceptions and preferences can influence raters' judgments (e.g., prior teaching experiences, assessment beliefs, accent bias, and even their level of harshness when evaluating performance quality). The latter dimension concerns what raters perceive as problematic during the actual use of the scale and the kind of suggestions they have for its improvement. As a consequence, Zhong (2019) emphasizes the need for studies that investigate raters' perspectives and preferences during the design and development of scales, since they are its end users and, arguably, the most critical agents in the effective employment of these instruments.

Harding (2017, p. 12) investigated scale usability from the perspective of scale users, considering they are "[...] highly problematic to design and implement". The author argues that the adversities and uncertainties raters encounter during scale use should not be treated merely as implementation issues, but as important sources of evidence to inform and improve the development process of future instruments with similar purposes. He further stresses that such user-related considerations should be regarded as integral to the scale validation process. Considering this concern for raters' perspectives and the practical use of assessment

instruments, Harding (2017) examined the usability of the CEFR (Council of Europe, 2018) phonological control scale through a focus group of L2 teachers experienced with assessment. For the author, usability refers to the degree to which a scale is user-friendly to its intended audience.

The participants reported several problems related to the design and the content of the instrument. One issue concerned the scale's internal logic. According to participants' perception, it was rather inconsistent, as the importance of intonation for L2 speech, for example, was not present in all descriptor levels. Another issue involved the use of some words that were not explicit to describe pronunciation performance and required users to rely on external references for interpretation (for instance, *natural*). The findings of the study led Harding (2017) to develop two types of recommendation for future scale developers. Technical recommendations concern the structure and format of the scale (e.g., ideal number of performance levels), and construct recommendations relate to issues regarding criteria (e.g., pronunciation and fluency should be assessed together). Despite the relevance of the study, the author concludes stressing there is a need for more studies that investigate scale usability according to the perspective of specific groups of users, so that it will be possible to understand what design elements facilitate or hinder the use and interpretation of certain assessment instruments.

In this regard, Isaacs, Trofimovich and Foote (2018, p. 1) adopted a user-centered approach to scale development and created a rating scale to assess university students' comprehensibility level, defined as "the degree of perceived listener effort required for understanding L2 speech". Conceived as a formative instrument for L2 pronunciation assessment in monologue tasks, the scale aimed to inform pedagogical decisions and to identify specific phonological features that could help learners produce speech that would be perceived as easier to understand. To this end, the authors involved members of the intended user group in the development process. They analyzed data from several focus group discussions with teachers of English for Academic Purposes. After constant revisions, the initial version of the scale was expanded from three to six levels, and its descriptors underwent multiple modifications in order to better reflect participants' recommendations.

In addition, this study aimed to develop a rating scale that would not require learners to eliminate their L1 accent in order to reach the highest level of comprehensibility present in the scale. This reflects the authors' commitment to the principle surrounding contemporary L2 pronunciation research that states that it is possible for learners to maintain their linguistic identity while still being regarded as competent English users, even in higher education

contexts. Additionally, it shows that creating a structured assessment instrument for pronunciation assessment does not mean imposing a nativeness standard to learners. Finally, Isaacs, Trofimovich and Foote (2018) mention they plan to develop a user supporting guide to enhance the scale's usability for future users. This guide includes a glossary, explicit construct definitions, and examples of potential misuses of the instrument. The authors propose incorporating speech samples at each level of the scale to support more refined and continuous practice for teacher-raters.

In Brazil, research investigating teachers' perspectives on the use or development of rubrics or rating scales for L2 pronunciation assessment remains limited, particularly within CBA contexts. To the best of my knowledge, Martins and Silveira (2025) appears to be the only study to have explored how Brazilian L2 teachers' assessed learners' intelligibility and comprehensibility with the CEFR (Council of Europe, 2018) revised phonological control scale. Although the focus of the study was not classroom assessment, some qualitative comments indicate that teachers require additional practice in order to use the scale effectively when assessing students' pronunciation. Besides that, participants reported that having access to an instrument with clear performance descriptors guiding their judgment facilitated the complex process of assessing L2 pronunciation, which many of them perceived as challenging prior to the study.

Together, the studies reviewed in this chapter point to several issues that still require further investigation in research. First, there is a need to better understand Brazilian teachers' preparation for addressing language and pronunciation assessment, as well as how they perceive the use of rubrics for formative purposes in small-scale, classroom contexts with students in regular education. Second, important challenges remain regarding decisions about rubric design and construct, especially in relation to the selection of criteria and length of performance descriptors. Third, there is a need to test the rubric usability while recognizing the specific environments in which these instruments are implemented and the contextual constraints that their intended users may face. Accordingly, in an attempt to address these gaps, the present study aims to develop an analytic rubric for L2 pronunciation assessment in classroom settings and refine the instrument considering Brazilian English teachers' perceptions. The next chapter, therefore, explains the method employed in the study.

3 METHOD

This chapter outlines the method adopted in the present study. It first describes the study design and details ethical procedures that were adopted in accordance with national and institutional guidelines. Succeeding that, the chapter presents the study's participants and reports on the development process of the preliminary version of the L2 pronunciation analytic rubric, including the theoretical foundation that informed its creation. The chapter then presents the instruments used for data collection and explains how each one of them was designed. Ultimately, the procedures implemented for data collection and analysis are reported thoroughly.

3.1 STUDY CHARACTERIZATION

The present study adopted a mixed-method approach. Therefore, it integrates quantitative and qualitative procedures to understand Brazilian English teachers' perceptions of an L2 pronunciation rubric. As Dörnyei (2007) argues, with a mixed-methods, it is possible to have a more complete portrait of complex themes, especially when they are embedded in specific educational settings, which is the case with the phenomenon under investigation in this thesis. Furthermore, the study may align with the principles of action research (Banegas; Consoli, 2020), considering it aims to develop and refine an assessment instrument for classroom use in order to promote context-sensitive changes to the instrument's design that may aid teachers' classroom assessment practices.

3.2 ETHICS REVIEW BOARD

As the present study involves human beings, it was sent to the *Comitê de Ética em Pesquisa com Seres Humanos* from the Universidade Federal de Santa Catarina (CEPSH-UFSC) to receive the required ethical review. This process guarantees the integrity of the participants throughout the entire process of data collection and analysis. The CEPSH-UFSC approved this study under the number CAAE (*Certificado de Apresentação de Apreciação Ética*) 87404725.3.0000.0121 on May 7, 2025. Before accepting to join the study, participants read the *Termo de Consentimento Livre e Esclarecido* (TCLE), which detailed its method and other foreseeable physical or psychological discomforts they could experience. The TCLE will be referred to hereafter as the informed consent form (see Appendix A).

3.3 PARTICIPANTS

The participants were in-service Brazilian teachers of English as an L2 that held at least an undergraduate degree in English and were actively engaged in regular instructional practices at the moment of data collection. They were recruited online through a poster shared on NUPFFALE⁶'s Instagram profile. The image included a brief description of the study's purpose and a web address for reading the informed consent form, as well as for accessing the Background Questionnaire and the Assessment Survey (see Section 3.4). In addition, after confirming the participants' eligibility, the researcher encouraged them to invite other teachers who might be interested in participating, a strategy known as snowball sampling.

Of the 26 teachers who demonstrate interest to participate after the initial invitation, only 13 were considered qualified to be included in the study. Specifically, two teachers did not meet one of the prerequisites, which was having a finished English degree, and 11 teachers answered the Background Questionnaire and the Assessment Survey, but did not respond to the researcher's reminders to move on and complete the final stages of the study. Consequently, data from these 13 participants were not included. Therefore, the final sample consisted of 13 in-service teachers of English as an L2, whose information is visually presented in Table 1.

Table 1 – Participants' background

Category	Information
Gender	Men ($n = 7$, 53.8%); Women ($n = 6$, 46.2%)
Age	Mean = 34.1 years; $SD = 9.1$; Range = 23-48
Undergraduate degree	English and Portuguese teaching ⁷ ($n = 6$, 46.2%) English studies ⁸ ($n = 4$, 30.8%) English teaching ⁹ ($n = 3$, 21.1%)
Graduate degree	MA ($n = 6$, 46.2%) MA studies in progress ($n = 4$, 30.8%) Specialization ($n = 2$, 15.4%) None ($n = 1$, 7.7%)
Teaching experience	More than 15 years ($n = 3$, 23.1%) More than 10 years ($n = 3$, 23.1%) 6-10 years ($n = 3$, 23.1%) 1-5 years ($n = 3$, 23.1%) Less than 1 year ($n = 1$, 7.7%)

⁶ NUPFFALE (Núcleo de Pesquisa em Fonética e Fonologia Aplicada à Língua Estrangeira) is a Brazilian research group focused on the teaching and learning of speech in foreign language contexts. I am a member of this group and had access to its official Instagram account for data collection purposes.

⁷ Refers to *Licenciatura em Letras - Inglês/Português*.

⁸ Refers to *Bacharel em Letras - Inglês*.

⁹ Refers to *Licenciatura em Letras - Inglês*.

Prior experience with teacher	Yes ($n = 7$, 53.8%)
education pronunciation courses	No ($n = 6$, 46.2%)

Source: the author (2025).

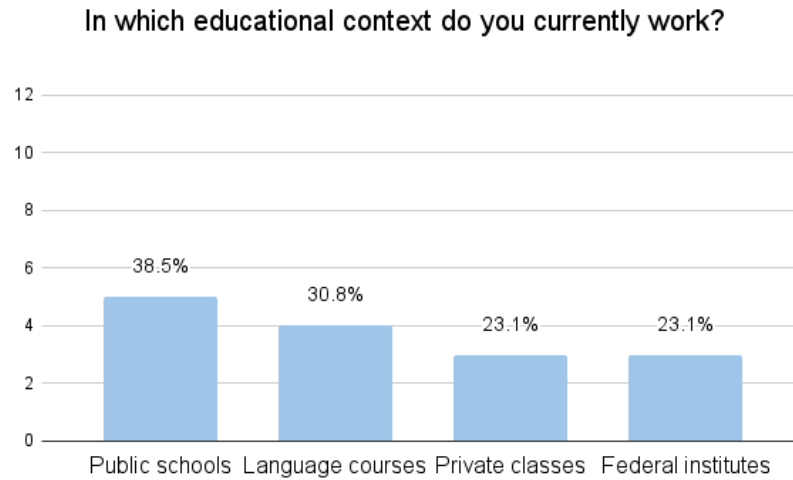
Notes: n = Participants within the sample; M = Mean value; SD = Standard deviation.

The participants were seven men (53.8%) and six women (46.2%), aged between 23 and 48 years ($M = 34.1$ years, $SD = 9.1$). Regarding undergraduate degrees, most of them ($n = 6$) held dual degrees in English and Portuguese teaching, while the remainder held a degree in English studies ($n = 4$) and English teaching ($n = 3$). Furthermore, with the exception of one participant, all of them reported having a graduate degree in the field of applied linguistics, including MA degrees ($n = 6$), MA studies in progress ($n = 4$), or specialization degrees ($n = 2$).

Participants' teaching experience profiles were diverse and included veteran and novice teachers. Three participants have been teaching for more than 15 years (23.1%), three for more than 10 years (23.1%), three for 6 to 10 years (23.1%), and three for 1 to 5 years (23.1%). Only one participant had less than one year of teaching experience (7.7%). In relation to teacher education initiatives for pronunciation instruction, more than half of the participants ($n = 7$, 53.8%) reported they had attended courses that addressed issues related to it. Six participants ($n = 6$, 46.2%) indicated no previous contact with any form of teacher education initiative in the field.

As shown in Figure 3 below, participants are part of different teaching contexts. In the public sector, five participants teach in elementary and high schools (38.5%) and three work in federal institutes (23.1%). As for the private sector, four participants are employed in language courses (30.8%) and three offer private classes independently (23.1%). Considering this was a multiple-response question, two teachers who contributed to the study reported teaching in two locations at the same time, namely, giving classes at language courses and offering private lessons.

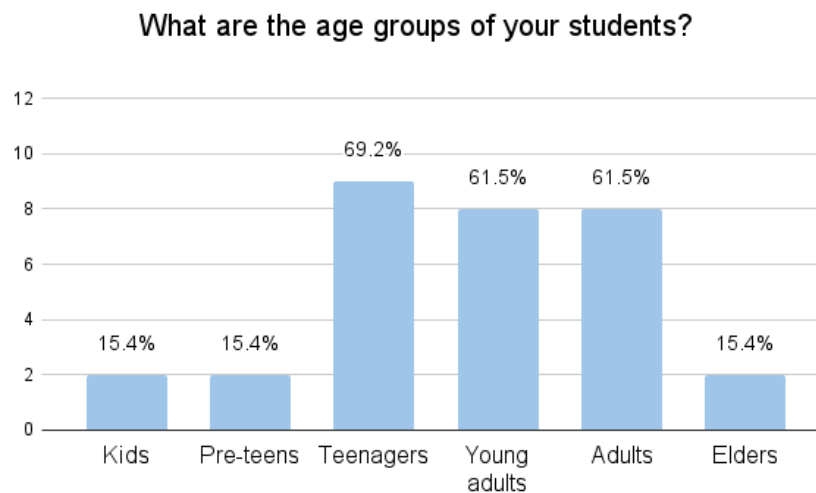
Figure 3 – Educational contexts where participants are currently working



Source: the author (2025).

For the age groups of the participants' students, Figure 4 illustrates that most teachers mainly teach teenagers ($n = 9$, 69.2%). Another considerable number indicated that they have young adults ($n = 8$, 61.5%) and adults ($n = 8$, 61.5%) students as well. In addition, some participants report they instruct children ($n = 2$, 15.4%), pre-teens ($n = 2$, 15.4%), and older learners ($n = 2$, 15.4%). In most cases, participants stated teaching more than one age group, which suggests that they usually deal with learners with different profiles and needs altogether.

Figure 4 – Age groups taught by participants



Source: the author (2025).

The next subsection will present the approach adopted to design the preliminary version of the L2 pronunciation rubric and the instruments used in the data collection process.

3.4 INSTRUMENTS AND MATERIALS FOR DATA COLLECTION

The three instruments implemented for data collection in this study include: i. a Background Questionnaire (see Appendix B), designed to outline participants' academic and professional experiences; ii. an Assessment Survey¹⁰ (see Appendix C), aimed at understand teachers' previous formal education in distinct aspects of language assessment in addition to their pronunciation assessment practices in the classroom; and iii. a Rubric Survey (see Appendix D), intended to investigate teachers' perception about the rubric developed in the study and how it could be improved for the Brazilian pedagogical context.

Nonetheless, it is important to note that this section presents the approach to designing the L2 pronunciation rubric, including the definition of its criteria and the development of its performance level descriptors. Although the rubric is not a data collection instrument in essence, it is presented here because it represents a central component of the study and functioned as a pedagogical device that participants were required to use to complete one of the surveys.

All instruments were piloted with a small sample of three English teachers from regular schools in Brazil. During this phase, participants used them in order to assess how clear and practical they were. They indicated some aspects that needed improvement in some instruments, which were all revised before data collection started. The suggestions from participants in the pilot affected mainly the format of some items (i.e., adopting more closed-ended, multiple-response questions instead of open-ended ones when appropriate). Data from the pilot were not included in the final analysis.

3.4.1 Approach to Designing the Rubric

When it comes to rubrics design for formative purposes in the L2 classroom, scholars (e.g., Arter; Chappuis, 2006; Brookhart, 2014; Chinda, 2013) defend the use of empirically-developed instruments. An empirical approach ensures that the assessment instrument is aligned with the expected performance and the needs of its intended pedagogical population end-users. In this process, the rubric developer (traditionally a researcher or a teacher) collects multiple samples from learners who belong to the same, or a similar context, and who are engaged in tasks comparable to those the rubric aims at assessing. Each sample is then assessed

¹⁰ McKay and Gass (2005, p. 92) describe that surveys are "one of the most common methods of collecting data on attitudes and opinions from a large group of participants", while questionnaires refer more to the written list of questions itself. In light of this distinction, I will use the term 'survey' for instruments designed to elicit participants' perceptions and beliefs and 'questionnaire' for instruments that primarily gather factual information.

individually, while the developer attempts to group them according to performance quality in order to produce detailed written reflections to justify why a particular sample was classified in a certain manner.

According to Arter and Chappuis (2006), at this stage, accuracy in the assessment results should not be the primary concern. The focus is on generating rich and descriptive statements that can support the development of the rubric's initial draft, particularly its criteria, as the developer begins to identify features that are important for quality performance. As more samples are analyzed, the developer may notice that a single performance often has strengths and weaknesses at the same time. For instance, a description of an L2 pronunciation performance at this planning phase might resemble the following: "This student produces all vowels correctly, but has difficulties with prosodic patterns at the word level, resulting in monotonous speech". This indicates that the construct in focus has distinct dimensions and justifies adopting an analytic rubric structure instead of a holistic one.

In line with this approach, the preliminary version of the rubric developed in the present study was created through a similar empirical process. The researcher selected approximately twenty speech samples from the Speech Accent Archive (Weinberger, 2025) belonging to Brazilian Portuguese learners of English as a L2 with different levels of proficiency and ages. None of them had experience living abroad. Initially, the samples were classified into only three categories: "proficient", "developing" and "beginner", as Arter and Chappuis (2006) recommends. The researcher wrote thorough descriptions for each one of the samples as they were classified and attempted to explain their placement grounded in identifiable speech features, as opposed to only using simple and evaluative words (i.e., great). Through this repeated process, recurring elements began to emerge in all descriptions, leading to a preliminary set of criteria. These criteria are presented and discussed in the following section.

3.4.2 Rubric Criteria

As Brookhart (2014, p. 24) argues, the criteria in a rubric should reflect the most significant "[...] characteristics of the learning outcome the performance is supposed to indicate". In other words, instead of focusing on features of a task (e.g., specific elements that students must include to complete it successfully), criteria should be centered towards relevant dimensions that meaningfully support student learning and serve as indicators of that ability. In this sense, criteria function as descriptors of what counts as valid evidence of students'

achievement in relation to the intended learning outcomes. Brookhart (2014) further emphasizes that, for rubric criteria to be considered sound and effective, they should meet some “desired characteristics”. Specifically, they should be appropriate, definable, observable, distinct, complete, and capable of representing a continuum of quality. The explanation for each characteristic is presented in Table 2.

Table 2 – Brookhart’s (2014) desired characteristics to write high quality criteria for formative rubrics

The criteria are...	Explanation
Appropriate	Each criterion represents an aspect of a standard, curricular goal, or instructional goal or objective that students are intended to learn.
Definable	Each criterion has a clear, agreed-upon meaning that both students and teachers understand.
Observable	Each criterion describes a quality in the performance that can be perceived (seen or heard, usually) by someone other than the person performing.
Distinct from one another	Each criterion identifies a separate aspect of the learning outcomes the performance is intended to assess.
Complete	All the criteria together describe the whole of the learning outcomes the performance is intended to assess.
Able to support descriptions along a continuum of quality	Each criterion can be described over a range of performance levels.

Source: Brookhart (2014, p. 25).

Considering the analysis of the speech samples described in the previous section, several elements important to the quality of learners’ performance were identified, which guided the identification of the rubric assessment criteria. In order to account for what contributed to facilitating or hindering Brazilian learners of English pronunciation in the L2, it was often necessary to refer to aspects such as *vowels*, *consonants*, *prosody*, *syllable structure*, *fluency* and *comprehensibility*. Consequently, the criteria included in the preliminary version of the rubric includes all of the aforementioned dimensions. These components are extensively discussed in the literature as fundamental to the development of an intelligible speech in the L2 (Levis, 2018). According to Derwing and Munro’s (2015, p. 379) definition, intelligibility refers to “the extent to which a speaker’s message is understood by a listener”.

In relation to each criterion present in the rubric, segmental accuracy refers to the precise articulation of vowels and consonants, which carry perceptual cues that are important for word recognition during speech production, especially when they are minimal pairs (Celce-Murcia *et al.*, 2010). In turn, prosody includes rhythm, stress, and intonation. According to Levis (2018), prosodic features are central to convey meaning as they reveal learners' communicative intention and support listeners' interpretation, which is seen as a fundamental factor to intelligibility. Additionally, syllable structure relates to the "phonotactic constraints of a language" (Brown, 2015), and it is a feature of speech that presents challenges for learners whose L1's syllabic patterns differ from those in the L2. This often leads to issues such as vowel epenthesis, consonant simplification, or stress misplacement, which can compromise listeners' ability to understand the utterance.

Moreover, fluency concerns how naturally the speech is produced. For example, even when individual sounds are articulated correctly, if they are paced too slow, too fast or filled with hesitations or repetitions, they can still hinder intelligibility (Thomson, 2015). Finally, comprehensibility refers to the perceived degree of effort required to understand someone (Derwing; Munro, 1997) and is, therefore, grounded in the listeners' own personal judgments. Consequently, factors such as familiarity with a particular variety or lexical and grammatical errors, may affect how comprehensible a learner is assessed. For this reason, intelligibility and comprehensibility are not always related (Derwing; Munro, 1995). A learner may produce speech that is intelligible (i.e., the listener understood all words produced) and yet be evaluated as difficult to understand due to other social and contextual variables that exceed pronunciation *per se* (Ballard; Winke, 2017; Isaacs; Trofimovich, 2012; Derwing; Munro, 2009).

To conclude, the speech dimensions identified through the analysis of Brazilian learners' oral production seems to be in line with Brookhart's (2014) guideline for high-quality rubric criteria discussed previously. Each criterion seems to represent relevant evidence of students' L2 pronunciation development and is supported in the literature as essential for the promotion of intelligible speech. In addition, the rubric is grounded on empirically-derived and theoretically-validated dimensions, which strengths its pedagogical relevance and methodological rigor. For the most part, the first proposed criteria for the rubric reflects all the desired characteristics Brookhart (2014) defends: they are sensitive to learning outcomes, can be defined, were perceptible in learners' performance, have distinct focus, but together provide a comprehensive representation of what constitutes L2 pronunciation development in a continuum of developmental levels.

3.4.3 Rubric Performance Levels and Descriptors

The discussion on how many numbers of performance levels a rubric should have remains ongoing. According to Brookhart (2014) and Arter and Chappuis (2006), the rubric developer should include as many levels as necessary to differentiate performance in relation to its quality. Nonetheless, some specific number of levels are already expected to be problematic for teacher-raters during assessment. For instance, a three-level rubric is not recommended in certain cases because, except for the best and the worst examples of performance, it may lead teachers to use only the middle level as a way to indicate students' development. In response to that, the authors suggest that rubric developers should aim for four to six levels when developing a rubric, although this number can be adapted according to the specificities of the context and the construct being assessed. Additionally, Isaacs, Trofimovich and Foote (2018) suggest including an “unable to assess” level in pronunciation scales, even if it is questionable whether many learners will actually be placed in it. This movement is supposed to encourage teachers to use the lowest level of the rubric, when necessary, instead of neglect it.

Regarding the performance levels descriptors, Brookhart (2014, p. 26) states that “[t]he most important aspect of the levels is that performance be described, with language that depicts what one would observe in the work rather than the quality conclusions one would draw”. Furthermore, she further proposes that rubric developers should start with the description of the level that represents the highest expected performance under ideal learning conditions. Once this reference point is defined, the other performance level descriptors should be articulated in relation to it. Again, Brookhart (2014) outlines several desired characteristics for the performance levels and descriptors of formative rubrics, stating that they should be descriptive (instead of evaluative), clear so as that teachers and students can comprehend the content, cover performance levels from beginner to proficient, be distinct from one another, accurately center the expected performance at the proper level, and maintain parallel descriptions and contents across all levels. An in-depth explanation of these characteristics is provided in Table 3.

Table 3 – Brookhart's (2014) desired characteristics to write high quality performance levels descriptors for formative rubrics

The descriptors of performance levels are...	Explanation
Descriptive	Performance is described in terms of what is observed in the work.
Clear	Both students and teachers understand what the descriptions mean.

Cover the whole range of performance	Performance is described from one extreme of the continuum of quality to another for each criterion.
Distinguish among levels	Performance descriptions are different enough from level to level that work can be categorized unambiguously. It should be possible to match examples of work to performance descriptions at each level.
Center the target performance (acceptable, mastery, passing) at the appropriate level	The description of performance at the level expected by the standard, curriculum goal, or lesson objective is placed at the intended level on the rubric.
Feature parallel descriptions from level to level	Performance descriptions at each level of the continuum for a given standard describe different quality levels for the same aspects of the work.

Source: Brookhart (2014, p. 28).

The characteristics discussed previously informed the development of the performance levels and the descriptors of the preliminary version of the rubric. Following the suggestion to include between four and six levels, the researcher adopted a five-level scale in the rubric, which includes an “unable to assess” level. The remaining four levels were considered sufficient to differentiate the spectrum of pronunciation performances seen in the samples analyzed during the design process¹¹. With respect to the descriptors, the descriptor of the highest performance level was drafted first, as it represents the expected learning outcome from students. The other levels were developed after it, through adaptations to illustrate a clear progression in the rubric. All descriptors were written using descriptive language and tried to focus on features of students’ L2 pronunciation that could be mentioned in words. Additionally, attention was paid for each descriptor to be clear, distinct from the other, and for them to maintain a consistent structure and content focus throughout all levels of the rubric.

3.4.4 Rubric Supporting Guide and Glossary

In assessment, teacher-raters training is of uttermost importance to ensure that an assessment instrument it is used in the same manner it should (Harding, 2017). Nonetheless, teachers often might not have access to the necessary preparation to effectively learn how to use ready rubrics, as participating in workshops or training sessions is frequently impractical or inaccessible (Jeong, 2015). As a result, they may rely on personal impressions or prior

¹¹ Studies reporting on the development of rubrics and rating scales show that the performance levels of these instruments tend to be upgraded as a consequence of teachers or experts’ preferences (see Chinda, 2013; Isaacs; Trofimovich; Foote, 2017; Galaczi *et al.*, 2011). This was another reason why I chose to start with a four-level rubric.

professional experiences to use these instruments. Isaacs, Trofimovich and Foote (2018) created a supporting guide to aid teachers who could use the comprehensibility scale the authors developed. Considering that, and that this study aims to develop an analytic rubric for in-service teachers that can be used in different pedagogical settings within the Brazilian educational context, it became imperative to design a supporting guide to enhance the usability and reliability of the rubric (see Appendix E). This guide is expected to provide teachers the autonomy to understand and implement the rubric independently, without requiring direct assistance from the researcher or other experts in the field.

The guide includes an overview of the role of rubrics in the L2 pronunciation assessment, possible contexts for implementation, and practical guidance on how to use the rubric before, during and after the assessment process, to mention a few. Additionally, a glossary is included to explain some important terms and concepts that will probably ease teachers' understanding. One particular relevant section focuses on common pronunciation difficulties and challenges Brazilian learners of English may encounter in speech production, with a specific focus on segmental sounds, which tend to be "more L1-specific than prosody patterns" (Isaacs; Trofimovich; Foote, 2018, p. 04). As discussed in the previous chapter, this part can be particularly relevant for English teachers from Brazil given the limited professional opportunities many of them received in this field (Buss, 2016). Therefore, the guide may serve as a pedagogical aid during rubric use and as a resource for teachers' continuing education as well.

To develop the section regarding pronunciation difficulties, which addresses the most frequent English segmental deviations of Brazilian learners, Zimmer, Silveira and Alves's (2009) meta-analysis was used to support its theoretical foundation. The authors identified prior studies investigating Brazilian learners' segmental production in English, including vowels, consonants, and syllable structure. For each difficulty addressed, there is an explanation detailing whether it stems from orthographic interference or from the dearth of that specific sound in the Brazilian Portuguese phonemic inventory. Moreover, there are examples to illustrate how these issues may affect intelligibility and communication and articulatory guidance on how to produce them. The section covers vowel contrasts such as /ɪ/ vs. /i/, /ʊ/ vs. /u:/, and /æ/ vs. /e/, known to be problematic for Brazilians, in addition to other difficult patterns, involving, for example, consonant replacement, final nasal vocalization, and syllable structure.

In order to share the resources developed in this study with participants, a dedicated Moodle page was created exclusively to host them (for more details see section 3.5). The page

was structured as a self-study environment and access was restricted to participants only. Its layout included a welcome section introducing the resources, an introductory video presenting the rubric and the supporting guide, and separate files containing the complete versions of the rubric (in an editable .docx format) and the supporting guide (in PDF format). In addition to that, the page hosted two speech samples, which were made available to encourage teachers' spontaneous and exploratory practice with the rubric. All resources could be downloaded. In general, the page centralized all learning resources in a single digital space and supported participants' independent training, which aided to address a frequent challenge reported in teacher education concerning assessment instruments and rubric use.

3.4.5 The Preliminary Version of the L2 Pronunciation Assessment Rubric

The first version of the L2 pronunciation assessment rubric designed to be used inside the classroom for formative purposes for Brazilian teachers and learners of English is presented below, in Table 4. Note that the rubric was written in Portuguese, the first language of its intended users – Brazilian English teachers.

Table 4 – The preliminary version of the L2 pronunciation assessment rubric

	Avançado 4	Progredindo 3	Começando 2	Básico 1	Não demonstrado 0
Vogais	As vogais são articuladas com clareza e precisão, sem deixar dúvidas quanto à palavra pronunciada. Há consistência na produção de todos os sons vocálicos que representam uma dificuldade para aprendizes brasileiros.	As vogais são geralmente articuladas com clareza, mas com substituições ou acréscimos ocasionais que não impactam a compreensão. Há pequenas variações na produção de sons vocálicos difíceis para brasileiros, mas essas raramente interferem na compreensão.	As vogais são articuladas de maneira compreensível, mas com substituições ou acréscimos frequentes que podem dificultar o entendimento da palavra pronunciada. Há inconsistência na pronúncia de sons vocálicos difíceis para brasileiros.	As vogais são compreensíveis às vezes, mas não há uma distinção clara entre os sons vocálicos, o que dificulta a diferenciação entre palavras semelhantes e prejudica o entendimento do que foi dito.	Ainda não é possível avaliar.

Consoantes	As consoantes sonoras são articuladas com clareza, permitindo fácil reconhecimento da palavra pronunciada. As consoantes surdas apresentam aspiração apropriada quando necessário (/p/, /t/ e /k/). Nenhum som é substituído por outro no lugar das consoantes esperadas.	A maioria das consoantes sonoras são articuladas com clareza, com apenas pequenas falhas que não comprometem a compreensão. A aspiração de consoantes surdas ocorre de forma consistente na maioria das vezes. Erros de substituição (‘tink’ ou ‘fink’ > ‘think’) acontecem com sons difíceis para brasileiros e em palavras que podem ser compreendidas pelo contexto.	Há dificuldade na articulação clara de consoantes sonoras e a aspiração em consoantes surdas ocorre de forma inconsistente ou não ocorre. Erros de substituição (‘hose’ > ‘rose’) acontecem com sons difíceis para brasileiros e em palavras que não são compreendidas mesmo com a ajuda do contexto.	As consoantes sonoras e surdas apresentam articulação pouco clara, sendo difícil entender qual palavra está sendo pronunciada devido a como uma consoante foi dita. A aspiração não é produzida. Substituições de sons consonantais por outros são comuns, dificultando significativamente a compreensão do que está sendo dito.	Ainda não é possível avaliar.
--------------------------	--	--	---	---	--------------------------------------

Prosódia (ritmo, entonação e acentuação)	Demonstra controle claro de padrões da fala. O acento tônico das palavras (word stress) está corretamente colocado, facilitando a compreensão (PHOtograph/photoGRAPHic). O uso do acento frasal (sentence stress) destaca com precisão as palavras-chave da frase. O ritmo da fala é natural e há o uso eficiente de sílabas fracas e fortes (I WANT to GO to the MARket). A entonação se adequa ao propósito comunicativo, indicando perguntas, surpresas, afirmações, etc.	Demonstra um controle geral dos padrões da fala. O acento tônico costuma estar bem posicionado, com algumas variações que não comprometem a compreensão (HOfel > hoTEL). Palavras-chave geralmente recebem destaque na frase, ainda que às vezes o ritmo soe menos natural ou com sílabas igualmente fortes. A entonação é usada de forma consistente para os propósitos comunicativos, mas com pouca variação.	Há tentativas de aplicar os padrões da fala. O acento tônico das palavras pode aparecer em posições inesperadas, o que pode dificultar a compreensão em alguns momentos. Palavras-chave para a compreensão da frase nem sempre se destacam. O ritmo tende a ser regular, sem contraste claro entre sílabas fortes e fracas. A entonação aparece de forma limitada, mas com potencial para desenvolvimento.	Os padrões da fala são reconhecidos poucas vezes. O acento tônico das palavras muitas vezes não segue o esperado, afetando a compreensão. As frases tendem a soar monótonas ou com ritmo irregular. A entonação é pouco explorada, dificultando perceber intenções (como momentos de pergunta ou ênfase).	Ainda não é possível avaliar.
Fluência e habilidade de auto-correção	A fala flui de forma contínua, com um número elevado de palavras considerando o tempo de produção. As pausas ocorrem em pontos apropriados	A fala flui geralmente de forma contínua, com um número consistente de palavras considerando o tempo de produção. Algumas pausas ou aumento da	A fala tem pausas frequentes que podem interromper o fluxo da comunicação, deixando-o mais lento ou confuso. O número de palavras é limitado	A fala é fragmentada, com pausas longas ou frequentes que dificultam o fluxo da comunicação. Poucas palavras são produzidas e o esforço para formar	Ainda não é possível avaliar.

	(como entre ideias ou frases) e não afetam a compreensão. O ritmo é mantido mesmo quando é necessário corrigir rapidamente algum engano. Auto Correções são feitas sem interromper a comunicação.	velocidade ocorrem em momentos inesperados, mas não afetam a comunicação. Há tentativas de correção de enganos, ainda que esse processo possa levar mais tempo e impactar o fluxo da fala.	para o tempo. Necessita de tempo para organizar o que quer dizer e para se autocorrigir em momentos necessários. OU A fala é muito rápida e é difícil acompanhar o fluxo da comunicação. O número de palavras é elevado para o tempo, mas a pronúncia das mesmas é geralmente equivocada. Necessita de tempo para se autocorrigir em momentos necessários.	frases é evidente. A autocorreção é rara ou ocorre de forma ineficaz.	
Formação silábica	Não há adição ou omissão de sons na formação silábica, mantendo o mesmo número de sílabas esperado. Encontros consonantais complexos (woRLD) para brasileiros e consoantes em posição final de palavra (hearT/spam/well) são produzidas adequadamente.	Poucas ocorrências de adição ou omissão de sons na formação de sílabas. Encontros consonantais e consoantes finais são produzidas corretamente, ainda que ocasionalmente haja simplificação dos sons ou aumento do número de sílabas esperado ('IStop' > 'stop' / 'milki' > 'milk').	Ocorrências moderadas de adição ou omissão de sons na formação de sílabas, aumentando ou diminuindo o número esperado. As mudanças nos encontros consonantais e em consoantes finais podem interferir na pronúncia ('IStriti' > 'street' / 'fin' > 'fine'), mas algumas partes são compreendidas com ajuda do contexto.	O número de sílabas é sempre muito maior ou menor que o esperado. Encontros consonantais são frequentemente simplificados e consoantes finais não são produzidas, comprometendo significativamente a compreensão mesmo com a ajuda do contexto.	Ainda não é possível avaliar.

Compreensibilidade	A fala é extremamente fácil de compreender.	A fala é fácil de compreender.	A fala demanda um pouco de esforço do ouvinte para ser compreendida.	A fala demanda muito esforço do ouvinte para ser compreendida.	Ainda não é possível avaliar.

Source: the author (2025).

To indicate the specific difficulties students had in each criterion of the rubric, a supplementary, optional chart adapted from Celce-Murcia *et al.* (2010) will accompany the instrument. Table 5, presented as it follows, illustrates its content.

Table 5 – Mapping specific student L2 pronunciation difficulties in rubric criteria

Se desejar, utilize este espaço para comentar uma dificuldade específica que tenha percebido na pronúncia do estudante avaliado:
Vogais
Vogais tônicas (stressed vowels):
Vogais átonas (unstressed vowels):
Consoantes
Iniciais:
Médias:
Finais:
Encontros consonantais (inicial ou final):
Prosódia (entonação e acentuação)
Entoação (perguntas ou afirmações):
Acentuação (das palavras ou frasal):
Fluência e habilidade de auto-correção
Formação silábica
Comentários adicionais

Source: Adapted from Celce-Murcia *et al.* (2010).

3.4.6 Background Questionnaire

Participants were required to answer an online background questionnaire (available in Appendix B) to warrant that they met the predetermined criteria to be involved in the study. This instrument was designed using the Google Forms platform, and the questions collected relevant information about teachers' age, undergraduate degrees, academic history, years of professional experience, teaching contexts, and students' age ranges.

The questionnaire was written in Portuguese to facilitate teachers' comprehension. Before answering any personal questions, participants were presented with a welcome message

and were able to read the informed consent form in its entirety. In this message, they were reminded of the voluntary nature of the study and of all the planned stages for data collection. The researcher's e-mail address was provided in case they had further questions. They could only proceed to the first section of the questionnaire after agreeing to participate.

The questionnaire consisted of eight items. The first questions were open-ended (e.g., name and age) and focused on demographic information, while the last ones were closed-ended and had multiple-choice options (for instance, "What are the age groups of your students?"). Participants were expected to complete this instrument in approximately 10 minutes.

3.4.7 Assessment Survey

The second instrument administered to participants was an online L2 assessment survey, available in Appendix C. This instrument was incorporated as the second section of the Background Questionnaire, primarily to reduce the time participants spent engaged with the study, considering the generally busy schedule of teachers. As with the first questionnaire, this one was designed using Google Forms and written in Portuguese. The survey aimed to collect data on Brazilian teachers' prior education in language assessment, as well as on the pronunciation assessment practices adopted in their classroom. Consequently, it was divided into two sections.

The first section contained questions adapted from Vogt, Tsagari and Spanoudis' (2020) study, conducted in Europe, which aimed to examine teachers' professional preparation in various domains of language assessment. These questions addressed classroom-focused assessment, purposes of assessment, the assessment of different L2 abilities, and important assessment constructs (i.e., validity and reliability). Given that the original instrument was written in English, all questions have been translated into Portuguese to enhance the comprehension of Brazilian participants.

Some adaptations were made to the instrument to ensure that it would better reflect the context of the present study. For instance, a question related to European assessment guidelines was removed and replaced with a question to explore teachers' knowledge of rubric development and use. The final version of this section consisted of 18 closed-ended questions, in which participants indicated, on a three-point Likert scale, whether they had received none, basic or advanced education in each item. Additionally, there was a final open-ended question for participants to clarify previous answers, if they wished.

The second section of the survey focused on teachers' practices regarding classroom pronunciation assessment. First, participants were questioned whether they assessed their students' pronunciation. Those who answered "yes" continued to the following questions, whereas those who answered "no" were allowed to leave the survey. This section included 10 closed-ended questions with multiple-choice options and one optional open-ended question. To illustrate, some of them investigated how confident they felt assessing pronunciation, the proficiency levels they usually assessed, the criteria adopted, the instrument used to operationalize these criteria, and the weight assigned to pronunciation in learners' oral production grades.

This survey was designed to be completed in approximately 15 to 25 minutes. The data collected through this instrument can provide a foundation to interpret how teachers evaluate the rubric and to identify the changes that need to be conducted in the instrument in order to better address some of the preferences and challenges teachers report, thereby indirectly supporting RQ1 (How do Brazilian teachers of English evaluate an L2 pronunciation assessment rubric in relation to its criteria, performance levels, and usability?) and RQ2 (How can the experiences and insights emerging from teachers' evaluation inform a revised version of the L2 pronunciation assessment rubric?).

3.4.8 Rubric Survey

The third instrument participants completed was a rubric survey (available in Appendix D). The purpose of this survey was to provide participants with the opportunity to apply the preliminary L2 pronunciation rubric to authentic speech samples from Brazilian learners of English, in order to investigate whether they would achieve comparable ratings of learners' pronunciation development. Furthermore, this instrument was intended to elicit teachers' perceptions on the quality of the rubric's design and suggestions they had for improving it. In line with the previous instruments, it was designed using Google Forms and presented in Portuguese to participants.

The survey had two complementary sections. In the first section, participants were requested to use the rubric to assess three speech samples from Brazilian university students, learners of English as an L2¹², who were instructed to read a short text without any planning time. Due to the learners' different proficiency levels (selected to distinctly represent beginner, intermediate, and advanced pronunciation profiles) and to control for order effects in the results,

¹² These samples were retrieved from a corpus used in Silveira and Martins (2020).

the speech samples were presented in a different sequence for each participant¹³. They could listen to each sample as many times as needed and were instructed to consult the rubric and assign a performance level for each criterion before proceeding to the next sample.

After completing the pronunciation assessment task, participants moved on to the second section, which aimed to explore teachers' perceptions of the rubric. This section was divided into three categories of rubric design and use: criteria, performance levels, and usability. Brookhart's (2014) desired characteristics¹⁴ to create high-quality rubrics served as the foundation to develop the questions related to rubric criteria and performance levels. To illustrate, the characteristic of *definability* (which suggests that a criterion should have a clear, definable meaning for the rubric's intended users) was operationalized as the statement "Each criterion is clear and easy to understand".

Participants indicated how much they agreed with each statement using a 10-point Likert scale, where 1 corresponded to "strongly disagree" and 10 to "strongly agree". This section included seven statements that addressed rubric criteria and five statements concerning performance levels. In addition, two additional statements were included to evaluate how realistic (in relation to time) and appropriate (in relation to teachers' professional development in pronunciation instruction) the performance levels were for the Brazilian context. All items in these categories were closed-ended questions. To complement these, three optional open-ended questions invited participants to comment on possible changes to the rubric's criteria, performance levels and descriptors. They could mention any issues they considered relevant, but which had not been previously included in the closed-ended questions.

The final category of this section focused on the usability of the rubric, drawing on Arter and McTighe's (2001) claim that usability is a central feature of good rubrics. This category comprised nine items, mainly closed-ended and multiple-choice questions, covering aspects such as the role of the rubric in facilitating the assessment process, whether participants would consider adopting it in their teaching practice, and the purposes for which they would use it, etc. In addition, there was one open-ended question that invited participants to suggest

¹³ Order effects are a variable important to control in L2 speech assessment, as they can influence teacher-raters' perception and assessment results (see Munro; Derwing, 2020). To control for that through Google Forms, I created six versions of the Rubric Survey. Each participant had access to a different version of the instrument, assigned to them randomly. Except for the order of the speech samples, all the other instructions and questions were the same.

¹⁴ Two desired characteristics ("being able to support descriptions along a continuum of quality" and "feature parallel descriptions from level to level") were not included in the survey, considering both involve the early stages of rubric development. At that phase, I had already verified that all criteria were capable of supporting a coherent progression of performance descriptors and that parallel descriptions were present in all levels. Consequently, these characteristics were not investigated in the present study.

improvements to other aspects of the rubric and the supporting guide. The estimated time to complete the survey ranged from 25 to 35 minutes.

3.5 PROCEDURES FOR DATA COLLECTION

Data collection for this study occurred at four moments. First, participants agreed to participate in the study after fully reading the informed consent form (Appendix A). Second, they completed the Background Questionnaire (Appendix B) and the Assessment Survey (Appendix C). Third, they were granted access to Moodle to study the rubric and its supporting guide (Appendix E). Finally, they completed the Rubric Survey (Appendix D). The data was collected entirely online, and the study was design having in mind that participants were part of different educational contexts and, as a result, did not have the same amount of time available. Thus, efforts were made to ensure that all teachers had a comparable experience with the rubric and that participant engagement could be fulfilled and managed digitally. All procedures are detailed below.

Initially, as aforementioned, a digital recruitment poster was shared on NUPFFALE's Instagram account. The poster contained an illustration related to assessment and pronunciation, as well as a short description of the aim of the study, the researcher's contact, and a link to the online instruments hosted on Google Forms. Upon accessing them, participants were presented with the informed consent form, which described what was expected from them if they accepted to integrate the study and other important ethical information (for example, rights during participation, possibility of withdrawal at any time, etc.). They could proceed to answer the Background Questionnaire only after selecting the option "I have read the informed consent form and I agree to participate in the study".

Participants first completed the Background Questionnaire, which collected demographic and professional information such as age, undergraduate degree, teaching experiences, students' age and so on. After that, they were automatically redirected to the Assessment Survey without the need to open a new page. This survey investigated teachers' education on language assessment and their current classroom practices regarding pronunciation. The Assessment Survey was incorporated as the second section of the same Google Form to reduce the effort required from participants, considering the length of the instruments used in the data collection, as well as to increase the chances of teachers completing them thoroughly.

The initial answers from the Background Questionnaire were analyzed to verify which participants were, in fact, qualified to be recruited to the study. Only participants who held a complete undergraduate degree in English and who were currently involved with teaching practices were contacted through e-mail or phone, depending on the method indicated as preferred. In this first contact message, participants received instructions on how to access a Moodle page created specifically for the study. Moodle was chosen due of its integrated monitoring features, which facilitated verifying teachers' interaction with the rubric and the supporting guide at an individual level. At this stage, the researcher was able to access each participant's profile and consult the 'reports' section, in particular the 'logs' function, which generated detailed records of the users' actions within the platform (e.g., what resources they viewed and which files they downloaded).

To welcome the participants to the Moodle page and explain what they would be required to do next, the researcher recorded an introductory video in Portuguese. The video gave an overview of the present study, the rubric and the supporting guide, in addition to demonstrating briefly how to apply the rubric in practice and offering recommendations for teachers' self-study. Participants were instructed to first watch the welcome video and only then open the pedagogical resources developed. Along with the video, the Moodle learning environment counted with two speech samples from Brazilian learners of English as an L2, retrieved from the Speech Accent Archive¹⁵. Teachers did not need to actually assess these samples if they did not want, they were provided only to give participants more time to use the rubric independently before answering the final instrument of the data collection, the Rubric Survey.

After participants received the first message on how to access the Moodle page presented previously, they had one week to create a profile on the platform and view the resources there to demonstrate they were still interested in participating. Those who did not reply were excluded from the study. After that, to be conscious of teachers' time and understanding they needed to know the rubric really well before completing the final data collection instrument, participants were given another additional week to explore the rubric and use it at any time they preferred. It is important to note that each one of them had a different deadline to use Moodle, considering they entered the study on different days, depending on the day they first saw the recruitment poster or received an invitation from a peer, for example.

¹⁵ The Speech Accent Archive website can be accessed at the following link: <https://accent.gmu.edu/>.

In order to manage participants' timeline closely, the researcher created a spreadsheet containing information about each teacher and the point in which they were in the planned stages of the study. This spreadsheet was updated daily and recorded the exact moment a participant first accessed the Moodle page, the resources they interacted with in the digital environment, and how much time they still had left until they needed to answer the Rubric Survey. When the one-week period participants were given was due, the researcher contacted them again to send the Rubric Survey. Teachers were instructed to answer the survey using a computer and headphones in an attempt to reduce the impact of external noise on the results, considering they were required to assess recorded speech samples. Data collection lasted approximately a month and a half.

3.6 PROCEDURES FOR DATA ANALYSIS

To answer the research questions proposed in the present study, the information gathered from the Background Questionnaire, the Assessment Survey, and the Rubric Survey was analyzed. The first moment of data analysis involved preparing the data. First, to comply with ethical guidelines from CEPESH-UFSC, participants' personal names were replaced with codes (e.g., P1 stands for the first teacher who completed the data collection), which will be used to identify them from now on. Second, the data, previously stored in Google Forms, was downloaded and organized into Microsoft Excel spreadsheets. The original responses were deleted from the online platform and transferred to an external flash drive belonging to the researcher.

The Background Questionnaire and the Assessment Survey consisted primarily of closed-ended questions. Therefore, they were analyzed with descriptive statistics, used to describe "general tendencies" in the data (Dörnyei, 2007, p. 213). Frequencies and percentages were generated to outline teachers' academic experiences and setting of pedagogical practice, as well as how formally instructed they were in relation to language assessment and the pronunciation assessment practices they had. The data was later displayed in figures or tables to be visually presented. This part of the analysis was important to underpin the interpretation of participants' perceptions and engagement with the rubric developed.

On the other hand, the Rubric Survey consisted of three different sections that yielded different types of data, which had to be analyzed accordingly. To begin with, the results of teachers' assessment of the speech samples using the rubric were converted to numbers. For instance, 0 represented the lowest performance level in the rubric (namely, the "unable to

assess” level) and 4 the highest one. Learners’ performance on each criterion was calculated using means and standard deviations to determine whether participants were able to differentiate between the pronunciation developmental levels they possessed and to understand the extent to which they had comparable perceptions.

Additionally, the results of teachers’ impressions of the rubric criteria and performance levels (Brookhart, 2014) that were collected through a 10-point Likert scale, viewing participants’ level of agreement with each statement, were analyzed using means and standard deviations as well. Thus, the highest the mean and the lowest the standard deviation, the more participants agree with a given statement present in the survey investigating certain aspects a high-quality rubric should exhibit. For the questions related to rubric usability (Arter; McTighe, 2001), frequencies were generated as they were mostly closed-ended. In all quantitative analysis, JASP (version 0.95.4.0) was used.

Lastly, participants’ open-ended answers were analyzed through content analysis, according to Dörnyei’s (2007) proposal. The data collected was read multiple times to be understood and coded considering recurring patterns that begin to appear. After that, each comment was attributed to the categories of interest of the present study (i.e., rubric criteria, performance levels and usability). For comments that were related to the usability category, different issues were identified after analyzing the participants’ responses. Primarily, these related to ease of use, classroom applicability, reported challenges, and suggestions for improvement. The grammar errors present in the excerpts were corrected and they were translated into English to improve readability¹⁶.

Accordingly, the next chapter presents and discusses the results of the study derived from the aforementioned data analysis procedures.

¹⁶ Some of the participants’ open-ended answers had to be adapted for two reasons. First, to highlight the aspects most relevant to answer the research questions of the present study. Second, to respect the character limit of this thesis. To read them in full, see Appendix H.

4 RESULTS AND DISCUSSION

This chapter presents and discusses the results of the study. Section 4.1 reports on the analysis of participants' prior education in language assessment and their pronunciation assessment practices, providing contextual information to assist in the interpretation of teachers' evaluations and experience with the rubric. Section 4.2 addresses the research questions. In this sense, it first examines the extent to which Brazilian in-service English teachers applied the pronunciation assessment rubric consistently, and then explores their perceptions of the rubric in light of rubric criteria, performance levels, and usability for classroom application. In each section, quantitative findings are presented and discussed first, followed by the qualitative analyses.

4.1 BRAZILIAN ENGLISH TEACHERS' EDUCATION IN LANGUAGE ASSESSMENT AND CLASSROOM PRONUNCIATION ASSESSMENT PRACTICES

The first set of results was collected through the Assessment Survey and they are concerned with teachers' self-reported education in distinct domains of language assessment. Understanding teachers' professional preparation in assessment is of great importance to interpret how they later engage with the pronunciation rubric developed and to identify possible needs presented that could be used to inform its refinement. The questions were adapted from Vogt, Tsigari and Spanoudis' (2020) study and encompass three dimensions of Language Assessment Literacy: classroom-focused assessment practices, purposes of assessment, and conceptual knowledge of language assessment and testing.

With respect to the first dimension, classroom-focused assessment, participants reported varied levels of professional preparation in activities that are associated with frequent assessment practices at a small-scale level. As it is possible to see in the group results presented in Table 6, teachers seem to possess greater, advanced instruction in preparing tests to be applied in the classroom (46.2%) and in providing feedback for students considering the results gathered in an assessment event (46.2%). These findings suggest that participants are more qualified in classroom assessment practices that require hands-on and active abilities from them.

Table 6 – Teachers' prior education in classroom-focused assessment and testing

Classroom-focused language assessment and testing	None (%)	Basic (%)	Advanced (%)
--	---------------------	----------------------	-------------------------

Preparing classroom tests	0.0 (<i>n</i> = 0)	53.8 (<i>n</i> = 7)	46.2 (<i>n</i> = 6)
Using ready-made tests from textbook packages	30.8 (<i>n</i> = 4)	53.8 (<i>n</i> = 7)	15.4 (<i>n</i> = 2)
Providing feedback to students based on information from tests/assessment	23.1 (<i>n</i> = 3)	30.8 (<i>n</i> = 4)	46.2 (<i>n</i> = 6)
Using self- or peer-assessment	30.8 (<i>n</i> = 4)	38.5 (<i>n</i> = 5)	30.8 (<i>n</i> = 4)
Using informal, continuous, non-test type of assessment	30.8 (<i>n</i> = 4)	38.5 (<i>n</i> = 5)	30.8 (<i>n</i> = 4)
Creating and using rubrics	61.5 (<i>n</i> = 8)	23.1 (<i>n</i> = 3)	15.4 (<i>n</i> = 2)

Source: the author (2025).

Notes: *N* = 13.

However, to contrast the more expert preparation reported previously, teachers revealed uneven degrees of education in most of the remaining activities investigated. Participants' formal knowledge of self-assessment and peer-assessment, as well as informal and continuous assessment in the classroom, is reportedly none (30.8%) or basic (38.5%). Although participants may have had different interpretations of what exactly constitutes *basic* teacher education, it can be assumed, nonetheless, that most of them have limited preparation in implementing alternative and non-traditional forms of classroom assessment, as the same tendency has been found in other studies (e.g., Lan; Fan, 2019; Vogt; Tsagari; Spanoudis, 2020).

One of the most relevant results within this dimension concerns the creation and use of rubrics for L2 classroom assessment. In this respect, 61.5% of participants reported they have received no formal education on how to design and implement rubrics, which represents the highest percentage in all activities investigated for this first knowledge domain. Considering that this piece of research is interested in teachers' perceptions of a pronunciation rubric, this finding is of importance. From the results, it is possible to assume that participants may need more professional preparation in the theoretical and practical aspects of rubrics and that their experience with such instruments was quite limited before the study.

Regarding the second dimension, the purposes of assessment and testing, teachers once again showed different levels of formal preparation in each one of the themes explored, as presented in Table 7. Participants indicated the highest level of assessment knowledge on using assessment to understand what needs to be taught and learned (38.5% reported advanced expertise and 46.2% reported at least an elementary preparation). The results suggest that most participants may be able to conceive assessment beyond its traditional summative function and to see it as a “driving force¹⁷” to promote sensitive decisions that enhance the teaching-learning process (Scaramucci, 2006).

Table 7 – Teachers’ prior education in assessment and testing purposes

Purposes of assessment and testing	None (%)	Basic (%)	Advanced (%)
Assigning grades	23.1 (n = 3)	61.5 (n = 8)	15.4 (n = 2)
Finding out what needs to be taught/learned	15.4 (n = 2)	46.2 (n = 6)	38.5 (n = 5)
Placing students onto courses, programs, etc.	46.2 (n = 6)	30.8 (n = 4)	23.1 (n = 3)

Source: the author (2025).

Notes: N = 13.

When it comes to assigning grades, most teachers reported that they had received minimal education in how to grade students. Thus, despite grading practices consisting of an everyday activity in teachers’ lives and something they should do in an informed and principled manner, it seems that most participants are not adequately prepared to do it. Additionally, the least developed knowledge concerns using assessment for placement purposes (46.2% participants reported having no preparation in this regard). This finding is not unexpected, given that participants’ assessment duties possibly tend to be more classroom-oriented than administrative, as they are mainly part of public schools and private language courses.

The third and last domain analyzed focused on conceptual knowledge of language assessment and its content. Table 8 visually presents the group results, which reveal the most considerable limitation in teachers’ education and professional development in language assessment. In nearly all practices, participants’ answers indicate only a foundational instruction on how to assess important L2 abilities in classroom contexts, including writing

¹⁷ In the original, “mola propulsora da aprendizagem-ensino”.

(46.2%), speaking (61.5%), reading (53.8%), listening (38.5%), pronunciation (53.8%), and grammar/vocabulary (46.2%).

Table 8 – Teachers' education in content and concepts of language assessment and testing

Content and concepts of language assessment and testing	None (%)	Basic (%)	Advanced (%)
Assessing writing	30.8 (<i>n</i> = 4)	46.2 (<i>n</i> = 6)	23.1 (<i>n</i> = 3)
Assessing speaking	7.7 (<i>n</i> = 1)	61.5 (<i>n</i> = 8)	30.8 (<i>n</i> = 4)
Assessing reading	30.8 (<i>n</i> = 4)	53.8 (<i>n</i> = 7)	15.4 (<i>n</i> = 2)
Assessing listening	30.8 (<i>n</i> = 4)	38.5 (<i>n</i> = 5)	30.8 (<i>n</i> = 4)
Assessing pronunciation	15.4 (<i>n</i> = 2)	53.8 (<i>n</i> = 7)	30.8 (<i>n</i> = 4)
Assessing grammar/vocabulary	23.1 (<i>n</i> = 3)	46.2 (<i>n</i> = 6)	30.8 (<i>n</i> = 4)
Establishing reliability of tests/assessment	46.2 (<i>n</i> = 6)	46.2 (<i>n</i> = 6)	7.7 (<i>n</i> = 1)
Establishing validity of tests/assessment	46.2 (<i>n</i> = 6)	30.8 (<i>n</i> = 4)	23.1 (<i>n</i> = 3)
Using statistics to study the quality of tests/assessment	69.2 (<i>n</i> = 9)	30.8 (<i>n</i> = 4)	0.0 (<i>n</i> = 0)

Source: the author (2025).

Notes: *N* = 13.

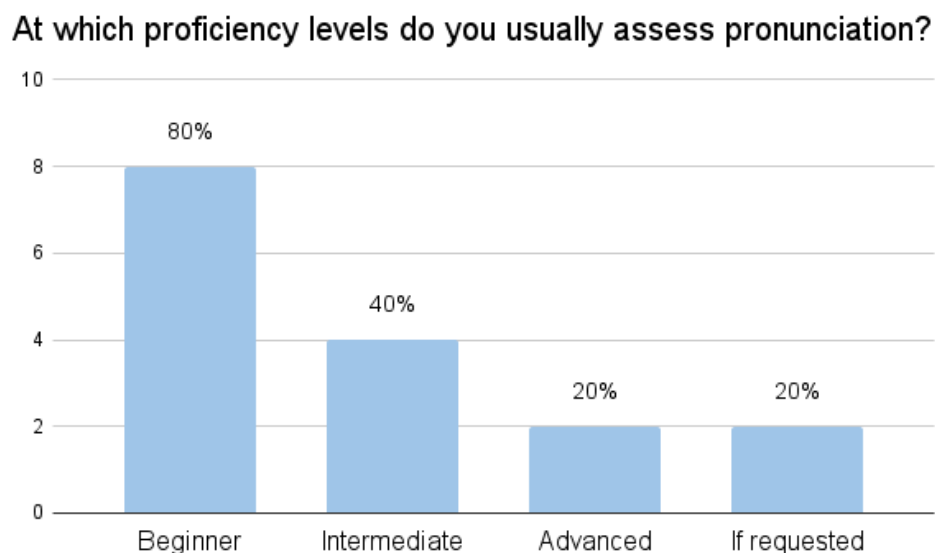
Teachers' previous formal preparation to assess oral performance are of particular interest to the present study. As it is possible to see, most participants report a restricted understanding of how to assess speaking and pronunciation in the classroom. This means that they possibly have some understanding about aspects related to L2 speech assessment, but may not have been extensively prepared to address them with the validity and reliability required,

considering it is one of the most difficult abilities to assess (Brown; Abeywickrama, 2018). Therefore, it seems that the group possesses different levels of education and an oral assessment-related competence is restricted to only the four participants who indicated advanced preparation in this regard.

Additionally, the results point to significant deficiencies in participants' assessment education regarding assessment properties. Most teachers reported they do not have any formal instruction on how to determine reliability (46.2%), validity (46.2%), or in using statistical measures to evaluate the quality of assessment activities or tests (69.2%). Ultimately, the results discussed so far suggest that teachers are more prepared to conduct and implement practical classroom-oriented assessment practices (such as applying tests and providing feedback) than to use concepts such as reliability and validity to aid them in improving the conditions and outcomes of the assessment practices they have (Lan; Fan, 2019).

The second set of the Assessment Survey results presents teachers' self-reported pronunciation assessment practices. Three participants (23.1%) stated that they do not assess pronunciation in the classroom and, therefore, did not need to answer the other questions that explored assessment practices in depth. For this reason, the analysis presented here considers only the 76.9% participants ($n = 10$) who actually assess students' pronunciation. Figure 5 presents the results of the first question, which investigated the level of proficiency of the learners assessed. As can be seen, the most frequent responses were beginner (80%), intermediate (40%), and advanced (20%). In addition, two participants (20%) mentioned that they only assess pronunciation when students themselves request it.

Figure 5 – Proficiency levels at which teachers assess pronunciation

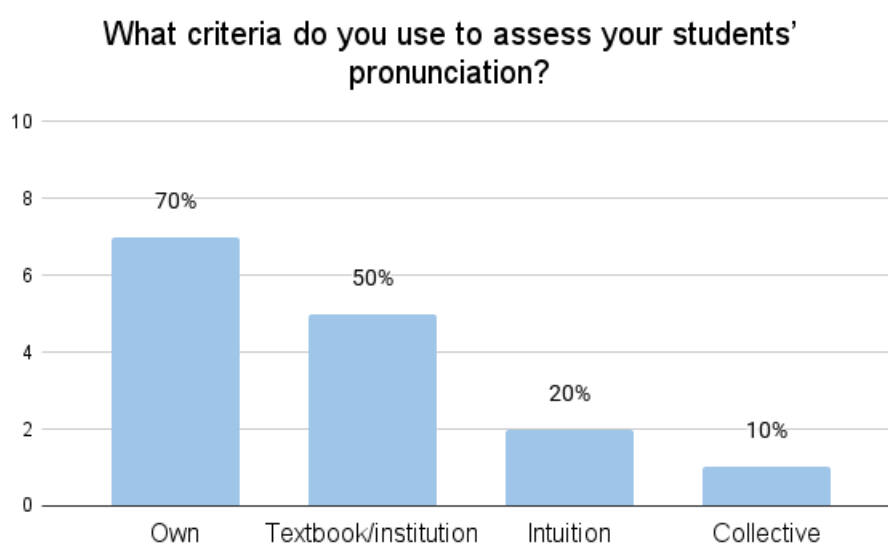


Source: the author (2025).
Notes: $N = 10$; Multiple-choice question.

As evidenced, participants more frequently assess students at the early proficiency levels. It is possible to presume that the practice is more frequent with beginners because, at this development stage, they tend to show more noticeable difficulties during L2 speech production (see Suzukida; Saito, 2022), which teachers may consider worth of special attention. Yet, the fact that teachers' pronunciation assessment practices are not consistent across all proficiency levels may further indicate that pronunciation is not given the same importance as other linguistic components. In this regard, Levis (2022) notes that, for a long time, L2 research disregarded pronunciation, something that certainly affected teacher education programs and, consequently, the way teachers were taught to perceive it.

In relation to the criteria teachers adopt to assess pronunciation in the classroom, Figure 6 shows that most participants rely on self-developed criteria (70%). Other frequently selected answers were institutional or textbook-defined criteria (50%) and criteria defined collectively with colleagues from the same educational setting (10%). In addition, it is important to highlight that two teachers (20%) reported using intuition to assess students' pronunciation. This implies that these teachers do not have predetermined standards of what constitutes quality in L2 pronunciation before the assessment event and, possibly, do not share a set of criteria with their students.

Figure 6 – Criteria used by teachers to assess students' pronunciation



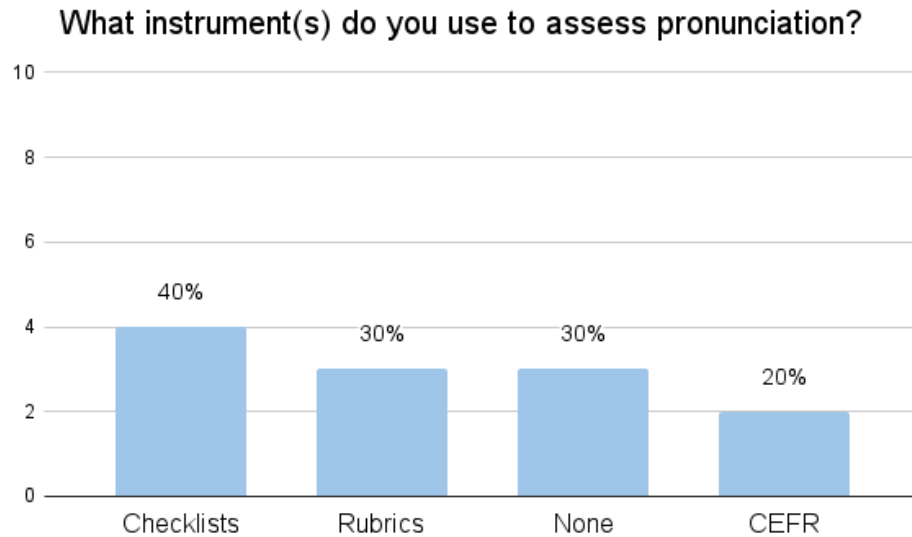
Source: the author (2025).
Notes: $N = 10$; Multiple-choice question.

While it is true that teachers should have autonomy to adapt the criteria used to specific contextual needs or learning outcomes, these results call attention to the need for Brazilian educational policy documents, such as *Base Nacional Comum Curricular* (Brasil, 2018), to provide teachers with clearer instructions on how to address pronunciation assessment in the classroom. This finding endorses what Quevedo-Camargo and Garcia (2017) found when they investigated the oral assessment practices of Brazilian teachers: they had different perceptions about which aspects of students' L2 speech were most important and adopted distinct criteria to assess it, which directly impacted the reliability of the results, since students within the same proficiency level received contrasting grades.

Furthermore, the reliance of some participants on intuition to develop assessment criteria for pronunciation, although less frequent, is consistent with their self-reported limited formal education in pronunciation assessment for classroom contexts, as discussed in the previous results regarding participants' experience with language assessment content and concepts. In these cases, the learning expectations teachers have may become less transparent to students and prevent them from fully participating in the assessment process. And, to intensify the scenario, pronunciation may be evaluated rooted in teachers' pre-existing linguistic beliefs and preferences rather than dimensions that have been shown to promote L2 intelligibility (Bøhn; Hansen, 2017).

Regarding the assessment instruments teachers use for pronunciation assessment, it is possible to see in Figure 7 that checklists (40%) and rubrics (30%) were the most frequently mentioned. In addition, a third of participants indicated that they do not use any formal assessment instrument (30%), and two teachers mentioned using the CEFR (20%) phonological control scale (Council of Europe, 2018). The results indicate that teachers do not assess pronunciation through only one type of instrument and that how they operationalize criteria and provide feedback differs significantly. This is yet another finding that stresses the importance of more initiatives to formally prepare Brazilian English teachers on how to address pronunciation assessment and the development of policy guidelines to support them in adopting a standardized approach to dealing with this issue.

Figure 7 – Instruments Brazilian teachers use to assess students' pronunciation



Source: the author (2025).
Notes: $N = 10$; Multiple-choice question.

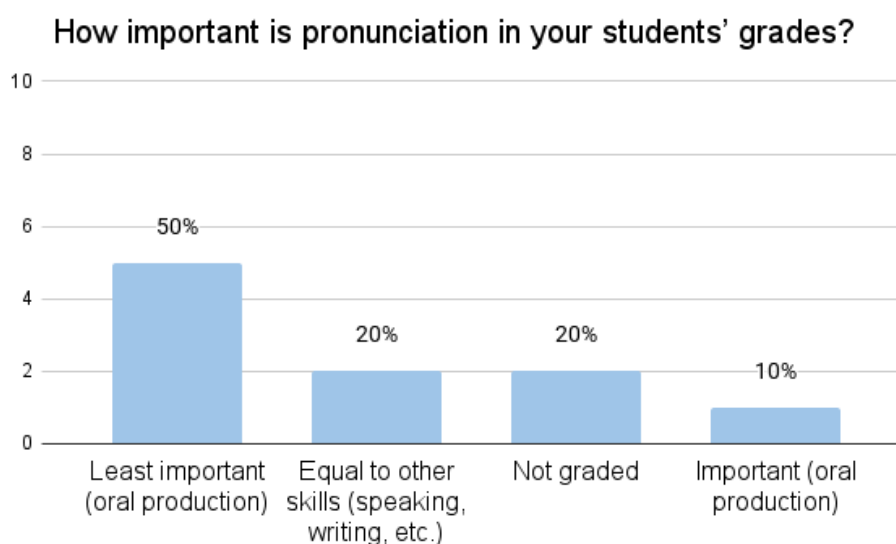
Additionally, as only three participants indicated using rubrics to assess pronunciation, the findings support the previous assumption that the largest part of the group may have approached the rubric developed in the present study with little prior experience in using analytic instruments for oral production assessment practices in classroom contexts. As checklists were the instrument participants mainly use according to the results, it is equally important to consider some of the limitations associated with their adopting for formative assessment procedures, as well as in the assessment of complex linguistic abilities, which is the case of L2 pronunciation (Isaacs; Thomson, 2013).

To illustrate, checklists are primarily used to verify whether desired elements have been included and completed in a given activity. According to Brookhart (2014), they are appropriate “[...] for situations in which the learning outcomes are defined by the *existence* of an attribute, not its quality” (p. 78, author’s emphasis). One great limitation of checklists, therefore, is that they do not specify what counts as great or poor performance, something that is fundamental in transparent and learning-oriented assessment practices. Consequently, in the assessment of pronunciation, they would indicate to students only that certain elements were or were not present in their oral performance. Nonetheless, when it comes to pronunciation, the mere presence of a feature does not guarantee that it was pronounced in a clear way that, in turn, would be consistent with learners’ actual linguistic ability or degree of intelligibility.

In a question investigating teachers’ grading practices, pronunciation was given varying levels of importance. Half of the participants (50%) reported that pronunciation

constitutes one of the least important criteria in students' oral production grade, while only one teacher (10%) considered it to be as important as other components of oral production (grammar, vocabulary, syntax, etc.). Two participants (20%) indicated that pronunciation has the same weight as other linguistic abilities included in students' grades (e.g., speaking, writing, listening, and reading), and the other two participants (20%) reported that they assess pronunciation, but do not assign grades to it. Figure 8 presents the results.

Figure 8 – Importance Brazilian teachers give to pronunciation assessment in students' grade



Source: the author (2025).

Notes: $N = 10$; Single-choice question.

Despite its value for effective communication in oral production (Celce-Murcia *et al.*, 2010), pronunciation does not seem to receive the same priority in Brazilian teachers' grades as grammar, for instance. Indeed, Isaacs (2018) argues that, compared to other linguistic abilities, pronunciation is set aside in many instructional and assessment contexts. One possible explanation for this result may be related to a widespread assumption that improved proficiency results in greater pronunciation control, which sometimes leads teachers to believe that pronunciation does not need to be explicitly taught (Levis, 2022) and, consequently, would not need to be assessed with the same rigor as other abilities.

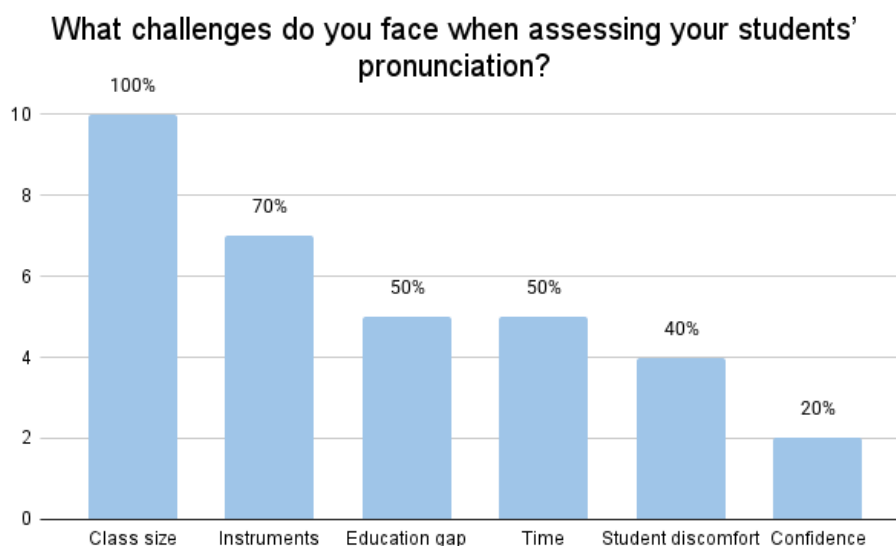
With regard to teachers' orientation towards intelligibility or nativeness (Levis, 2005) in assessment, most participants (80%) indicated that the primary intention they have when assessing students' pronunciation is to address features that may hinder intelligible communication. The findings appear to be consistent with Bøhn and Hansen's (2017) study, indicating that L2 teachers tend to adopt an intelligibility-oriented perspective to pronunciation

assessment. On the other hand, two teachers (20%) still mentioned that accent reduction and accuracy are the focus of the assessment criteria they draw on. In this regard, Levis (2020) holds that nativeness may be an impractical and unnecessary principle to guide pronunciation instruction and assessment, since learners can produce speech that is intelligible to listeners even if they still exhibit L1-accented features.

In another question exploring teachers' use of assessment results to inform and reorient teaching practices, most of the participants said that only 'sometimes' (50%) or 'rarely' (30%) lesson plans are adapted in view of the students' needs perceived after an assessment event. Just two teachers (20%) indicated that they 'always' use the information collected to improve pedagogical practices and better meet learners' pronunciation difficulties. It is possible to conclude then that the information which pronunciation assessment yields is not consistently integrated into the teaching and learning cycle. As Knoch (2017) describes, pronunciation assessment has pedagogical value for guiding syllabus design and a more productive use of class time to address students' difficulties with complex L2 oral aspects, but it is often not used to its full potential.

Ultimately, in relation to some of the challenges Brazilian teachers face when assessing pronunciation in the classroom, all possible options of the question were selected. As evidenced in Figure 9, teachers perceive class sizes (100%) and insufficient pronunciation assessment instruments for the Brazilian context (70%) as the two most important difficulties they experience in pedagogical practice. Other challenges noted were limited teacher education in phonetics, phonology, and language assessment (50%), time constraints (50%), concerns about embarrassing students (40%), and low levels of confidence to deal with the assessment of pronunciation (20%).

Figure 9 – Challenges Brazilian teachers face when assessing pronunciation in the classroom



Source: the author (2025).

Notes: $N = 10$; Multiple-choice question.

Since all participants indicated class size as one of the greatest challenges they face when assessing pronunciation, it is clear that this is certainly the element that most impacts how they approach pronunciation assessment in the classroom. Most participants in the study are part of public schools and federal institutes, where they probably have to manage from twenty to thirty students with different profiles and needs, in addition to overcoming the limited time allocated for English instruction. In this scenario, teaching and assessing pronunciation within a formative approach may be effortful, considering, as Celce-Murcia *et al.* (2010) describes, it demands them to plan, listen to hundreds of students (either in real time or through collected speech samples), provide individual feedback, use the information to orient future teaching practices, and so on.

The second most mentioned challenge concerns access to instruments that are context-responsive to assess the pronunciation of Brazilian students. According to Arter and Chappuis (2006), developing assessment instruments is time-consuming and can be difficult for teachers dealing with multiple students every day. Therefore, they should focus on creating analytic instruments only if they are too specific (i.e., aimed at assessing a particular activity) or cannot be adapted from other sources. For complex abilities, the authors argue that teachers should have access to already existing and validated assessment instruments, as well as be qualified to decide whether they are relevant, reliable, and well-designed for a given pedagogical context. Still, for that to happen, teachers need sufficient professional preparation on the construct they

plan to assess and on language assessment, which half of the participants (50%) identically mentioned as a frequent challenge faced.

Another challenge reported was related to teachers' fear of embarrassing students after assessing their pronunciation in formal contexts. Pronunciation learning is often seen as anxiety-inducing for L2 students (for a complete discussion, see Çelik *et al.*, 2025). In the same manner, Saito (2021) argues that correcting students' pronunciation in the classroom is a concern many teachers have and they tend to avoid giving feedback to students unless the pronunciation deviations they show are unintelligible, most of the time in an attempt to reduce negative feelings associated with the classroom. It would be interesting to investigate students' feelings regarding classroom pronunciation assessment in order to understand how they perceive it and, through which practices, this perception (if negative) could be attenuated.

Now that the prior formal experience of in-service Brazilian English teachers with language assessment and their classroom pronunciation assessment practices have been presented and discussed, it is important to understand how they evaluate the preliminary version of the L2 pronunciation assessment rubric. Thus, the next section reports on how consistent they were in assessing the speech samples with it. After that, the discussion turns to how they perceived the rubric in relation to its criteria, performance levels, and usability.

4.2 BRAZILIAN TEACHERS' PERCEPTION OF THE PRELIMINARY L2 PRONUNCIATION RUBRIC

4.2.1 Consistency in Scoring Results Using the Rubric

This section presents the data collected through the first part of the Rubric Survey. Teachers were requested to assess three Brazilian learners of English using the preliminary version of the L2 pronunciation assessment rubric developed in this study. The aim of this section was to investigate whether participants were already consistent in how they interpreted and applied the instrument to differentiate the proficiency levels represented in the speech samples chosen. To that end, the level of pronunciation performance assigned to learners was converted into numbers and mean scores were calculated for each learner across all criteria of the rubric.

As shown in Table 9, mean scores increased from learner 1 (beginner), to learner 2 (intermediate), to learner 3 (advanced) in all six criteria of the rubric. The results indicate that participants were able to consistently use the rubric to distinguish Brazilian English learners with different levels of pronunciation proficiency. Furthermore, they offer additional evidence

that participants had comparable interpretations of the performance levels descriptors, as well as that the descriptors may have helped them identify and indicate learners' stage of development in L2 pronunciation with greater reliability.

Table 9 – Mean scores assigned to three learners across the six pronunciation criteria of the rubric

Rubric criterion	Learner 1	Learner 2	Learner 3
Vowel	1.2 (0.4)	2.3 (0.6)	3.3 (0.5)
Consonant	1.2 (0.4)	2.1 (0.5)	3.1 (0.5)
Prosody	1.2 (0.5)	2.0 (0.6)	3.2 (0.7)
Fluency and self-repair	1.1 (0.7)	2.2 (0.8)	3.3 (0.6)
Syllable structure	1.3 (0.7)	2.1 (0.6)	3.3 (0.5)
Comprehensibility	1.6 (0.5)	2.7 (0.7)	3.4 (0.5)

Source: the author (2025).

Notes: $N = 13$; Mean (Standard Deviation).

At a more individual level, it is possible to perceive that learner 1 (who represents a beginner level), received the lowest mean score according to teachers' assessment in all the criteria present in the rubric. They ranged from 1.1 (fluency) to 1.6 (comprehensibility). And, as expected, learners 2 and 3 were placed in higher performance levels of the rubric. While learner's 2 mean scores of pronunciation development ranged from 2.0 (prosody) to 2.7 (comprehensibility), learner 3's performance quality reflected that of levels 3 and 4, as her assigned mean scores spanned from 3.1 (consonant) to 3.4 (comprehensibility). The standard deviations remained low in all assessments ($SD = 0.4-0.8$), suggesting that teachers' perceptions, when rubric-mediated, were related.

Note that, despite teachers' different experiences in language assessment and formal preparation to assess with rubrics (as reported previously), which could have influenced this positive outcome, they were still able to achieve a consistent assessment result regarding learners' pronunciation, possibly due to using the rubric (Jeong, 2015). This would be one of the most important impacts of introducing rubrics into assessment practices, since they guide users' interpretation and specify expectations of quality performance with comprehensive descriptors (Brookhart, 2014), which can help even teachers with limited expertise in language

assessment or in the assessed construct to have corresponding views on a student's learning stage.

4.2.2 Rubric criteria

To investigate participants' perceptions of the rubric criteria, the second section of the Rubric Survey included questions that explored the rubric's resemblance with Brookhart's (2014) desired characteristics for high-quality criteria. According to these characteristics, rubric criteria should be appropriate, definable, observable, distinct from one another, and complete. Table 10 presents the means and standard deviations for each statement with which teachers were requested to indicate agreement. More than one statement was created to address the desired characteristics of 'appropriate' and 'definable'. For this reason, an aggregated mean was calculated. In all statements, participants' ratings ranged from 7 to 10. The table includes an overall mean and standard deviation in order to offer a general overview of teachers' perceptions of the rubric criteria as a whole.

Table 10 – Summary of teachers' perception of the rubric criteria

Brookhart's (2014) desired characteristics for rubric criteria	Survey question	Mean (SD)	Overall mean per principle
Appropriate	The criteria are relevant to assess whether the student's pronunciation is intelligible.	9.6 (0.6)	9.4
	The criteria are in accordance with important educational documents found in curricular guidelines or language proficiency frameworks.	9.3 (0.7)	
Definable	Each criterion is clear and easy to understand.	9.6 (0.6)	9.4
	The criteria can be explained easily to other teachers and students.	9.3 (0.9)	
Observable	What each criterion describes could be perceived in the speech samples assessed.	9.5 (0.6)	9.5
Distinct from one another	Each criterion assesses a different aspect of pronunciation.	9.6 (0.8)	9.6

Complete	Taken together, the criteria include the main aspects of an intelligible L2 pronunciation.	9.9 (0.2)	9.9
		Overall <i>M</i> :	9.5
		Overall <i>SD</i> :	0.7

Source: the author (2025).

Notes: $N = 13$; Results on a 10-point Likert scale.

The first two questions addressed the ‘appropriate’ characteristic. Brookhart (2014) suggests that each criterion should represent a relevant aspect of the learning outcome, which for the rubric of the present study, represents the speech dimensions that L2 learners need to develop to achieve an intelligible pronunciation. Participants had a positive evaluation of how the rubric criteria resemble this characteristic ($M = 9.4$). The first question focused on the importance of the criteria within an intelligibility-oriented perspective, and it achieved a high level of agreement ($M = 9.06$; $SD = 0.6$). The second statement investigated whether the criteria were consistent with educational guidelines and standards for L2 pronunciation assessment. Although not actually low, this statement yielded one of the lowest mean scores of this section ($M = 9.3$; $SD = 0.7$). It seems that future versions of the rubric or the guide should more explicitly state how it aligns with national and international curriculum demands (e.g., how the rubric’s criteria could be correlated with BNCC (Brasil, 2018)).

The third and fourth questions investigate how ‘definable’ the criteria were. This characteristic is primary for teachers and learners to have a single definition of the criteria, as Brookhart (2014) argues. Once again, participants’ evaluation of this characteristic was high ($M = 9.4$). The level of agreement received in the third question ($M = 9.6$; $SD = 0.6$) suggests that participants believe that the meaning of each criterion was clear enough for them to understand. Nonetheless, the rates of the fourth question, which explored whether the criteria could be easily explained to students or even other teachers, show that they do not agree to the same extent ($M = 9.3$; $SD = 0.9$). This perception is mentioned again throughout the data analysis and further discussed in section 4.2.4, as teachers’ later comments indicate that they would need to explain the rubric criteria in detail and address it during class time for students to understand it.

In this respect, Brookhart (2014) emphasizes that students must comprehend the criteria of the rubrics adopted in the classroom. Therefore, they should be able to define them, especially when the assessment purpose is formative, since the criteria elucidate learning expectations and help them direct their efforts to achieve progress (Panadero; Andrade;

Brookhart, 2018). When it comes to pronunciation, the criteria may not be easy to understand from the beginning, given how complex certain speech dimensions can appear for students learning formally about them for the first time. As a result, it is important that they receive explicit pedagogical mediation to fully comprehend the oral features the rubric aims to assess. Despite the time required to do so, the efforts should be worth it, once the criteria are precisely what we expect students to develop when learning an L2. This reinforces Andrade's (2005) claim that the formative use of rubrics to improve learning do not depend only on how well designed they are (i.e., clarity), but on teachers' professional preparation to implement them throughout the educational process.

The fifth question aimed to examine whether the criteria could be 'observable' in learners' oral production. According to Brookhart (2014), well-developed rubrics have criteria that are related to a quality in students' performance that teachers can directly identify, see, or hear. The author states that this is especially important in abilities where intuitive, non-measurable elements might influence teachers' perceptions, such as student effort, for example. Participants seem to strongly agree ($M = 9.5$; $SD = 0.6$) that the rubric criteria could be perceived in learners' pronunciation during the assessment. This suggests that the speech dimensions (vowels, consonants, prosody, etc.) selected to compose the preliminary set of criteria of the rubric were identified in students' pronunciation.

The sixth question, related to the characteristic stating that criteria in a rubric should be 'distinct from one another', received high mean rates ($M = 9.6$; $SD = 0.8$). During the developmental process of rubrics, Arter and Chappuis (2006) highlight that criteria can go through many moments of improvement until it is decided which elements should be separated (e.g., they are too relevant to be together) or incorporated into the same dimension (for instance, they actually measure an identical aspect). In previous studies involving rubric or scale development for pronunciation, this is a recurrent issue; for example, participants in Harding's (2017) study preferred that segmental articulation and fluency were merged in the same criterion. Nevertheless, the results from this question indicate that participants agree that the six criteria of the rubric were distinct pronunciation features and that, for the time being, they are adequately separated.

The final characteristic Brookhart (2014) proposes for high-quality rubric criteria indicates they should be 'complete'. The question investigating the extent to which the criteria of the rubric developed in the study related to this characteristic generated the highest mean rate in all the statements evaluated in this dimension ($M = 9.9$; $SD = 0.2$). It is possible to conclude that, considering the strong level of agreement teachers demonstrated, the six criteria included

in the rubric are sufficient to assess all the relevant dimensions of L2 pronunciation. In addition, it seems that no other critical components were missing. In fact, 11 participants did not comment in the optional open-ended questions inquiring about the need to include or exclude a criterion in the rubric.

Only two participants left comments on the open-ended question of this section of the survey, where they could mention any aspects they believed were important in relation to the rubric criteria. The comments, although not entirely related to how the criteria can be improved, seemed relevant to be included in the results, as they offer important insights into different issues for rubric use and pronunciation assessment. In the first comment that will be discussed, one of the participants highlighted how complete the criteria are and, at the same time, stressed that he would need to adapt the number of criteria to implement the rubric with his students, as they may be too extensive. It should be noticed that this participant teaches mainly younger learners in a public-school context.

This rubric is too complete. I will use it with my seventh-grade students, but I will address comprehensibility and intonation/fluency. I believe I can provide feedback to the group as a whole on more “problematic” consonants/vowels, and give individual attention to intonation and rhythm, which are just as important as individual sounds¹⁸. (Participant 5, Rubric Survey).

The possibility of adapting the rubric criteria is mentioned in the supporting guide. Teachers are not expected to assess all criteria at the same time in every assessment event. Instead, they are encouraged to use professional judgment to select a specific criterion according to students’ needs, set learning outcomes, level of proficiency, or the pronunciation issues identified in learners’ production after a diagnostic assessment, for instance. In this sense, the main intention of the rubric is to include a comprehensive set of criteria that can inform teachers’ practices, while still granting them all the necessary flexibility for classroom implementation. One possible way to illustrate such flexibility is adapting Isaacs, Trofimovich and Foote’s (2018) proposal to include dotted lines separating each criterion in the rubric layout, which aids users to understand, through visual cues, that they can be adapted.

As for the second comment, P8 suggested including another criterion to assess “learner’s behavior”, in particular aspects associated with anxiety and nervousness during oral production. While this suggestion goes beyond the scope of the present study, which does not consider affective factors or individual differences in L2 assessment, it is relevant to report it

¹⁸ Adapted comment in Portuguese: “Essa rubrica é completa demais. Para meus alunos do sétimo ano usarei ela, mas tratarei da compreensibilidade e entonação/fluência. Creio que posso trazer para o grupo como um todo o feedback de forma mais geral em consoantes/vogais mais “problemáticas” e dar atenção no individual em entonação e ritmo que são tão importantes quanto sons individuais”.

as it demonstrates a common challenge in pronunciation assessment. According to Isaacs (2018), it is well-documented in the literature how difficult it is for teacher-raters to separate pronunciation performance from other linguistic and non-linguistic aspects during assessment. While anxiety can influence learners' L2 oral production to a certain degree, it is not part of the pronunciation construct itself and should not influence how learners' pronunciation performance is perceived.

While it is difficult to assess the level of anxiety and nervousness in pre-recorded audios – thus making it impossible to observe the speaker's behavior – a separate criterion for evaluating aspects related to the speaker's behavior would be interesting¹⁹. (Participant 8, Rubric Survey).

In general, it seems that the rubric criteria possess all the desired characteristics that Brookhart (2014) proposes. The participants of the study demonstrated a high level of agreement with all the questions appraised ($M = 9.5$). The question that received the best rates was related to how complete the criteria of the rubric are, which suggests that this is an aspect of the rubric that does not need to be improved for now. However, three important points should be considered: 1. future versions of the supporting guide could include how the rubric criteria can found in BNNC's (Brasil, 2018) requirement for learners' oral production development in English learning; 2. the possibility to adapt the rubric criteria should be made more explicit for teachers in order to facilitate its implementation in the classroom; and 3. students' understanding of the rubric criteria is essential and should be encouraged.

4.2.3 Rubric performance levels and descriptors

In order to explore participants' perceptions of the rubric performance levels and descriptors, the questions from the third section of the Rubric Survey aimed to understand how they related to Brookhart's (2014) desired characteristics for high-quality performance levels and descriptors. To the author, performance levels and descriptors should be descriptive and clear, cover the whole range of performance, be distinguishable from one another, and center the expected performance at an appropriate developmental stage. Table 11 presents the mean scores and standard deviations for each statement teachers evaluated. Individual responses to the statements ranged from 6 to 10. The table includes the overall mean and standard deviation

¹⁹ Original comment in Portuguese: “Por mais que seja complicado avaliar o nível de ansiedade e nervosismo de áudios já gravados - sendo assim sem possibilidade de ver o comportamento do falante - seria interessante um critério a parte para avaliar aspectos referentes ao comportamento do falante”.

to offer a general indication of teachers' collective view on the rubric's performance levels and descriptors.

Table 11 – Summary of participants' perception of the rubric performance levels and descriptors

Brookhart's (2014) desired characteristics for rubric performance levels	Survey question	Mean (SD)
Descriptive	The rubric descriptors are detailed and precise and, because they are grounded in observable evidence, teachers do not need to rely on personal judgments.	9.5 (0.5)
Clear	The rubric descriptors are clear, which means that both teachers and students can easily understand what is expected at each performance level.	9.3 (1.1)
Cover the whole range of performance	The rubric covers from the lowest to the highest levels of pronunciation proficiency, describing student performance at different stages of development.	9.3 (0.9)
Distinguished among levels	Each performance level is well-defined and distinct from one another; hence, teachers can unquestionably indicate a student's developmental stage with the rubric.	8.2 (1.0)
Center the target performance at the appropriate level	The description of the student's expected pronunciation performance level is placed at a corresponding level in the rubric. Thus, the level provides an accurate representation of their ability.	8.9 (0.9)
Overall <i>M</i> :		9.0
Overall <i>SD</i> :		1.1

Source: author (2025).

Notes: *N* = 13; Results on a 10-point Likert scale.

The first question aimed at comprehending how 'descriptive' the rubric performance levels descriptors were. According to Brookhart (2014), the descriptors are one of the most important aspects of analytic rubrics, as they detail exactly what to expect from students' performance in the ability measured. This level of detail is what aids teachers to not rely as much on personal impressions when assessing and students to understand which actions they should have to improve learning. Teachers seem to strongly agree that the rubric descriptors are well-articulated and explicit to the point that they reflect different speech production performance with just the exact degree of precision and universality ($M = 9.5$; $SD = 0.5$).

The second question investigated whether the rubric descriptors were 'clear'. It is important that not only teachers understand what they mean with accuracy, but students as well

(Brookhart, 2014). This is certainly one of the most challenging aspects of rubric development, especially for a construct such as pronunciation, given how difficult it is to reconcile a complete and precise description of speech dimensions with a linguistic style that students comprehend. For this statement, teachers' level of agreement varied a lot more ($M = 9.3$; $SD = 1.1$), despite still yielding high mean rates. One possible explanation for this finding, as later qualitative comments confirm, is that teachers probably found that students may not understand the descriptors of the rubric easily and need to receive guidance to interpret them.

The third question addressed whether the rubric scale 'covers the full range of performance', considering a rubric should represent learners' performance in every learning stage (Brookhart, 2014). Participants' rates were still high ($M = 9.3$; $SD = 0.9$), which suggests that they agree that the performance levels are able to represent the learning continuum from beginner to advanced L2 pronunciation development. It is possible to assume that, for the most part, the expected pronunciation learning progression was represented through the rubric scale. Nonetheless, despite the positive evaluation of the question, some later comments from participants revealed that the performance levels could be refined in order to reflect how they believe pronunciation progresses (see further discussion at the end of this section).

The fourth question focused on whether the rubric is 'distinguished among levels'. The rates for this question were the lowest in this dimension ($M = 8.2$; $SD = 1.0$). While it is not a low mean generally, when compared to other results, it suggests that teachers may have had difficulties finding a clear separation between the levels. Level distinction issues in rubric development have been reported in other studies. For example, Harding (2017) and Isaacs, Trofimovich and Foote (2018) found that users prefer that scales have comparable descriptors in all performance levels (i.e., if one element is mentioned in one level as important to the performance quality, it should be present in all of the other ones). Yet, this characteristic may blur what differentiates one level from the other which, in turn, confuse users and reduce clarity.

The final question explored if the rubric performance levels were able to 'center the target performance at the appropriate level'. Once again, participants' level of agreement on whether the rubric possessed this important characteristic was a little lower compared to other questions ($M = 8.9$; $SD = 0.9$). Possibly, teachers were uncertain about the correspondence between the performance level descriptors and the names assigned to them (e.g., '*básico*'). It is, indeed, complex to capture students' learning development in L2 pronunciation through single words that can be interpreted correspondingly. Thus, it will probably be necessary to refine the descriptors and reflect on how they were labeled in order to enhance clarity for teachers.

In addition to questions related to Brookhart's (2014) desired characteristics for high-quality rubrics, two additional statements were included to explore how practical the performance levels descriptors were for the Brazilian context. Attention was given to two difficulties frequently noted in the L2 pronunciation assessment literature (i.e., time and teacher education limitations). The first question was related to the appropriateness of the performance levels, given the amount of time Brazilian teachers have to assess students' pronunciation and oral production in the classroom. As it yielded the lowest mean rate and the highest standard deviation of the results analyzed ($M = 7.9$; $SD = 2.0$), teachers seem to disagree that the rubric's length, at least without the proper adaptations, is realistic for classroom implementation.

The differences in teachers' answers to the previous question may be due to the instructional and professional realities they have. For instance, while some participants teach in public schools with large class sizes, others are part of language courses or offer private classes. In the latter, they probably have more time and flexibility to adopt detailed assessment instruments, such as an analytical rubric, as well as time to provide formative and individual feedback. In a study conducted in Spain, which investigated teachers' perception of introducing a math rubric to classroom assessment, Gallego-Arrufat and Dandis (2014) found that the time required to use rubrics is often seen as one of the primary reasons why teachers are reluctant to implement them, particularly in public schools.

The second additional question aimed at comprehending if the performance levels were accessible to teachers who are not experts in pronunciation assessment or in phonetics and phonology. Once again, teachers' answers demonstrate lower levels of agreement with the statement ($M = 7.9$; $SD = 1.3$). This suggests that participants believe that other teachers, who have a more limited professional development in the field, may not be able to fully understand the rubric criteria and descriptors. Therefore, it is imperative that the supporting guide receives more attention in future research, as it is one possibility to prepare teachers with varied levels of experience to confidently learn more about pronunciation assessment and how to use the rubric.

Moving on to teachers' qualitative comments, they complement the previous findings and reveal how teachers perceive the performance levels in more detail. For the most part, participants noted some design elements that they considered positive and mentioned others that could be improved, such as descriptor clarity and level distinction, for example. To begin, one element that seems to have helped teachers during the assessment process was the examples included in the descriptors. One teacher said, "I especially liked the examples given in the level

descriptors, as they help to find similarities during the assessment²⁰ (P10). Therefore, although they were not specific, the examples possibly supported rubric users to interpret the descriptors and to differentiate the levels more easily.

In relation to that, two other participants stated they believe including more examples in the descriptors would improve the rubric, especially in the levels where they were not present before. For instance, P7 notes that including “[...] more common examples in the ‘vowels’ criterion could make the assessment easier²¹”, while P6 inquires: “Could you provide an example of syllable structure in the ‘*básico*’ level? I could not classify a speaker because I did not have a practical example of what would characterize them as one or the other²²”. Hence, the presence of examples elucidating what to expect from students’ performance in a given level seems to be capable of facilitating teachers’ judgment and, including more of them as a result of the suggestions seen here, may enhance its usability for Brazilian teachers.

Another comment concerned the usefulness of the level “not demonstrated”. One participant suggested to removing it: “I was unsure about the ‘*não demonstrado*’ level, perhaps I would exclude it²³” (P7). It is relevant to note that this level was included in the rubric due to findings from a previous empirical study reporting on the process of a scale validation with English for Academic Purposes teachers (Isaacs; Trofimovich; Foote, 2018). Although it is not expected to be used frequently, the authors state that its presence should serve two purposes. First, account for situations in which learners do not produce any kind of speech. Second, to encourage teachers to use the entire rubric, preventing them from grouping all learners’ performance in the center of the scale so as not to appear too severe in their evaluation.

With regard to the performance levels, participants suggest including additional levels to create a greater distinction in the rubric, particularly at the lower-end of the scale. A comment from P7 states: “I would add another level between ‘*básico*’ and ‘*começando*’. Perhaps the labels could also be reconsidered at these levels²⁴”. On the other hand, P9 reports: “I had difficulty classifying his [the student’s] performance as ‘*começando*’ and ‘*progredindo*’, it

²⁰ Original comment in Portuguese: “Gostei especialmente dos exemplos dados nas descrições dos níveis, pois ajudam a encontrar semelhanças na hora de avaliar”.

²¹ Adapted comment in Portuguese: “Se incluir exemplos mais comuns na seção vogais poderá deixar a avaliação mais fácil”.

²² Original comment in Portuguese: “Você poderia trazer um exemplo na parte de formação silábica na seção básico? Não soube classificar um falante por não saber exatamente um exemplo prático do que o caracterizaria como um ou outro”.

²³ Adapted comment in Portuguese: “Fiquei na dúvida em relação ao nível não demonstrado, talvez eu excluísse esse último nível”.

²⁴ Adapted comment in Portuguese: “Eu adicionaria mais um nível entre o básico e o começando. Talvez a nomenclatura também poderia ser repensada nesses níveis”.

seems that the gap between them is too big²⁵”. This perception indicates that the rubric may not be sufficiently nuanced yet when it comes to beginner performance. These findings are not unexpected, considering rubrics and scales often go through countless adaptations in the number of performance levels to have a more accurate reflection of the desired learning progression in the ability measured (e.g., Arter; Chappuis, 2006; Isaacs; Trofimovich; Foote, 2018; Galaczi *et al.*, 2011; Harding, 2017).

Finally, one comment addressed some possible challenges students may face in understanding the performance level descriptors: “When explaining it to my students, I am confident that a more detailed explanation would be necessary, as they would not always understand the examples²⁶” (P5). When considering his comment in light of the previous findings, this is another piece of evidence that endorses the interpretation that, in teachers’ view, the rubric is not completely supportive for student use yet and that they could use an adapted or mediated version, for example. As discussed previously, only when teachers and students, together, are able to engage with and apply the rubric does it become capable of promoting enhanced learning (Arter; McTighe, 2001).

Through rubric descriptors, learners can understand the defining qualities of an ability and actively put effort into improving, as they have written explanations on how to do so. It is important that they are not only introduced to rubrics, but ultimately able to interpret and use them independently (Brookhart, 2014; Arter; McTighe, 2001; Arter; Chappuis, 2006; Andrade, 2005). At the same time, there is a tension in writing descriptors that are easy to understand without simplifying them to the point where the rubric loses its quality in relation to descriptive details, which are crucial for fairer pronunciation assessment practices. Although some attempts have been made to illustrate how this can be achieved with writing rubrics, in which learners produced writing samples to elucidate how they interpreted performance levels or reformulated teacher-generated descriptors using simpler words (Brookhart, 2014), it remains unclear how effective these strategies are for pronunciation rubrics.

Considering all of this, it is possible to conclude that the rubric performance levels and descriptors resemble Brookhart’s (2014) desired characteristics, having in mind teachers’ perceptions discussed in the beginning of the section, although some aspects need further revision. Additionally, teachers’ open-ended comments clarified other aspects that would be

²⁵ Adapted comment in Portuguese: “Eu tive dificuldades em classificar seu desempenho em começando e progredindo, parece que o salto é muito grande”.

²⁶ Adapted comment in Portuguese: “Ao explicar para os alunos tenho certeza absoluta que teria que explicar de forma mais detalhada, os exemplos de referência nem sempre seriam entendidos por eles”.

relevant to be considered in future changes to the performance levels. More precisely, it seems pertinent to include more examples, to expand the rubric scale to encompass a greater number of performance levels, and to reconsider the names assigned to them. In relation to rubric implementation, the time required to do it and students' ability to use the rubric are the two most frequently mentioned concerns teachers have.

4.2.4 Rubric usability

4.2.4.1 *Ease of use*

The fourth section of the Rubric Survey focused on understanding teachers' views on rubric usability. The first question of the section concerned whether using a rubric facilitated the assessment of pronunciation, according to participants' own perceptions. Twelve teachers (93.2%) stated that using a rubric simplified "a lot" the assessment process, and only one participant (7.07%) indicated that "it made some aspects easier, but others more difficult". With such a positive finding, it is possible to conclude that teachers see that having the rubric to mediate the assessment made assessing pronunciation a less effortful experience. One of the teachers who believes using the rubric was something that enhanced her judgment process explains:

I found it incredible to have such specific guidance during the assessment. I first assessed each audio without the rubric and then with it, and I realized that, without the specific criteria, I had been more "lenient" precisely because I did not have a framework to follow. But, listening to the audios again with the rubric, I felt relieved to find descriptors and examples that matched what I was hearing²⁷. (P10, Rubric Survey).

Although it was not requested that she assessed first with a rubric and then without it, her comment elucidates not only a possible engagement with the rubric beyond the study, but the importance of rubrics to provide teachers with explicit criteria and detailed performance levels that aid them to rely less on intuitive reasoning. In fact, Brookhart (2014, p. 4) argues that "[...] the operating principle [of rubrics] is you match the performance to the description rather than "judge" it", which corresponds to the experience this teacher reports. Therefore,

²⁷ Original comment in Portuguese: "Eu achei incrível ter um direcionamento tão específico para avaliar. Avaliei cada áudio primeiro sem a rubrica e depois com ela, e percebi inicialmente que, sem os critérios específicos, tinha sido mais "boazinha" justamente por não ter um modelo a seguir. Mas, ao ouvir novamente com a rubrica junto, me senti aliviada por encontrar descrições e exemplos que iam ao encontro do que eu estava ouvindo".

using a rubric seems to have guided her towards a more descriptive assessment process and eased this, for many, complex mission.

In contrast, the only participant who reported contradictory perceptions regarding rubric use stated that the rubric made her pay attention to pronunciation dimensions she would usually disregard. Resembling P10, she believes that using a criterion-referenced instrument made her “more specific” on relation to criteria, which was positive. This teacher particularly mentions that, because of the rubric, she was able to see the importance of prosody and syllable structure to enhance pronunciation intelligibility. In this sense, it seems that using rubrics can initiate some sort of teacher reflection in them that may promote more informed pedagogical and assessment practices in the future, as Kennedy and Shiel (2022) argue.

I felt that the rubric helped me be more specific in my assessment criteria, which made me more rigorous, but it also added more time to the assessment (which would have made the process more difficult, given the well-known challenges teachers face regarding time). Normally, I did not consider prosody or syllable structure as much, but these are indeed important factors that can help in providing feedback and subsequent instructional guidance for students. However, I had doubts assessing the learners because I had little contact with the rubric and subject matter to feel truly confident in my judgment²⁸. (P7, Rubric Survey).

Despite that, P7 notes that the time required to use and implement rubrics would be the aspect that made her consider it less valuable. As mentioned beforehand, alternative forms of assessment instruments, such as analytic rubrics, entail more time to be applied due to how comprehensive they are (Arter; McTighe, 2001), which can be intensified during oral production assessment, notably already time-consuming for teachers (Celce-Murcia *et al.*, 2010). Nevertheless, it can be argued that using rubrics outweigh these issues, considering they can offer meaningful information about students’ pronunciation development and aid L2 teachers in determining which oral features are worth addressing in lessons, one of the greatest challenges they face in formal instructional settings (Darcy; Ewert; Lidster, 2012).

²⁸ Adapted comment in Portuguese: “Achei que a rubrica me ajudou a ser mais específica nos critérios de avaliação, então isso me fez ser mais rigorosa, mas incluiu mais tempo na avaliação (isso que teria dificultado o processo, pois se sabe dos desafios do professor no quesito tempo). Normalmente eu não considerava tanto a prosódia ou a formação silábica, mas de fato são fatores importantes que podem ajudar na construção de um feedback e de uma posterior orientação ao aluno. No entanto, tive dúvidas em avaliar os alunos, pois tive pouco contato com a rubrica e com o tema para de fato sentir-me segura avaliando”.

4.2.3.2 Reported challenges

In an open-ended question of the Rubric Survey, in which teachers were free to comment on any aspect they wished, participants reported several challenges related to rubric use and classroom pronunciation assessment in Brazil. The first issue relates to how teachers feel when they have to classify a learner's performance using the rubric when this performance falls between two closely related level. P5 says: "Sometimes we find ourselves torn between two levels. The rubric is here to facilitate the process, but you [the teacher] have to decide. At this point, some degree of personal judgment is important"²⁹.

According to Brookhart (2014), the difficulty this teacher reports is expected even in well-designed rubrics, given that descriptor interpretation can never be entirely "low-inference". The aim of rubrics is to have descriptors that are as detailed and explicit as possible, but they do not replace teachers' part in the assessment. Therefore, while it is true that rubrics describe to users what to expect from students' learning outcomes and constrain high degrees of subjective viewpoints, teachers' interpretation and professional judgment remain necessary in the entire process. They are the ones responsible for deciding, for instance, how frequent or salient a pronunciation feature is before learners are placed in a given performance level.

At the same time, it is important to point out that participants have a limited experience with using rubrics and only practiced for a short time. These factors could have led to a greater sense of uncertainty and less confidence in how they assessed because, as Fulcher (2015) argues, users need time to perfect how they interpret assessment instruments. In other words, the more familiarity with analytical instruments and opportunities for practice teachers have, the less hesitation they are anticipated to show. Hence, it is possible to assume that, as teachers have the chance to apply and implement the rubric with several examples of students' performance, this feeling will decrease.

The second issue one of the participants mentioned is related to using an analytic approach to assess L2 pronunciation. P12 explains: "I believe I still need more practice using the rubric because it is difficult for me to assess the criteria individually instead of providing a general assessment [of the learner's speech]"³⁰. It seems that this participant believes it is more intuitive to provide a global appraisal of learners' pronunciation compared to assessing it using a set of criteria. In his study, Harding (2017) found that teachers preferred a holistic approach

²⁹ Adapted comment in Portuguese: "Às vezes a gente fica em cima do muro entre dois níveis. A rubrica vem para facilitar, mas você precisa decidir. Nesse momento, vale um pouco o julgamento pessoal".

³⁰ Adapted comment in Portuguese: "Acredito que ainda preciso praticar a avaliação utilizando a rubrica, pois tenho dificuldades em avaliar os critérios individualmente em vez de uma avaliação geral da fala".

to assess pronunciation, possibly because they grant faster analyses and are cognitively less demanding for users dealing with time-restricted decisions (Knoch, 2017).

However, in classroom contexts, holistic assessment can limit the potential of formative assessment, since they account for students' performance in a general manner and does not provide lots of information about specific learning aspects that they still need to develop (Bachman; Palmer, 1996). In this sense, Knoch (2017) proposes that one possible ideal assessment scenario for pronunciation would involve a complementary use of holistic and analytic measures, in which main conclusions are made during the performance and more thorough detailed impressions are considered later. Still, how this would translate into a reliable pedagogical assessment instrument, as well as how it would benefit teachers and learners, has yet to be investigated.

Another issue reported has to do with the legitimacy of pronunciation assessment in light of the Brazilian curriculum for English teaching. In this respect, P6 questioned: "I would like to know if we [teachers] are still allowed to assess pronunciation. Recent documents and guidelines have left us uncertain about correcting students during oral activities³¹". National documents, such as BNCC (Brasil, 2019, p. 243), understand pronunciation as an essential component of oral production practices and negotiation for meaning during L2 interactions. Nonetheless, the hesitation this teacher demonstrates towards it suggests that it would be necessary that more concrete orientations should be provided to teachers on how to address pronunciation in everyday classroom assessment.

In addition to that, it is equally possible that her concern stems from the interpretation that pronunciation assessment does not have pedagogical value in the classroom due to the contemporary emphasis BNCC (Brasil, 2018) places on English as a *Lingua Franca*. It is important to note that pronunciation assessment, within an intelligibility perspective, does not aim at correcting learners' production considering deviations from a native norm, for example. On the contrary, when teachers correctly implement it to enhance students' L2 oral development, it can have a significant role in identifying and addressing individual challenges, especially because L2 pronunciation difficulties can vary even when students share the same L1 (Derwing; Munro, 2022).

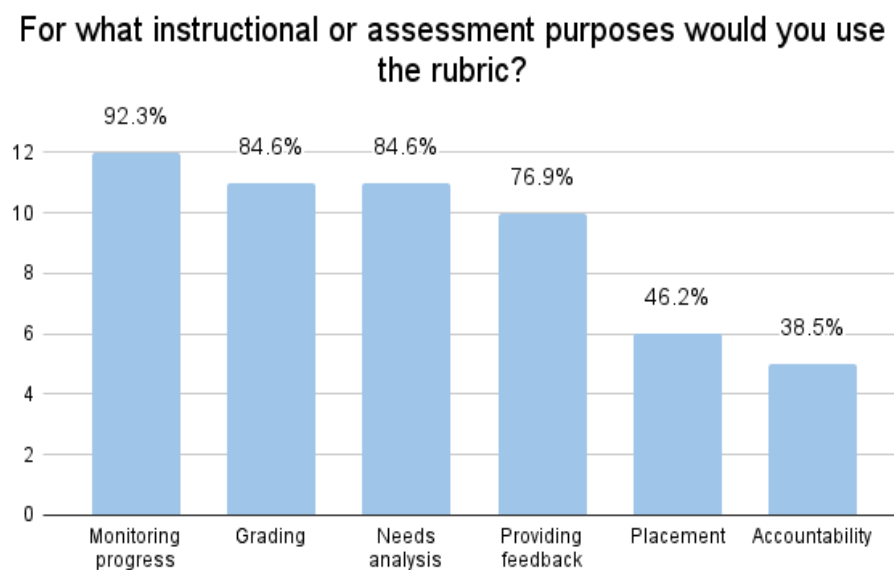
³¹ Original comment in Portuguese: "Eu gostaria de saber se ainda podemos fazer avaliações de pronúncia. Os documentos e orientações mais recentes têm nos deixando inseguros quanto a correções nas atividades de oralidade".

4.2.3.3 Classroom applicability

In a question investigating whether teachers would consider adopting the rubric in their pedagogical practice, seven participants (53.8%) stated they would “certainly” use the rubric in the future, while six participants (46.1%) reported they would use it as long as they had time available and students could gain from its introduction in the assessment process. This result suggests that most teachers have a positive view regarding a possible implementation of the rubric in Brazilian classrooms. However, once more, it shows that time constraints are one important element compelling or preventing teachers from assessing with analytic rubrics, an issue that has emerged several times throughout the discussion of previous results.

In another question that requested teachers to state the purposes for which they would use the rubric, they selected different pedagogical and assessment uses. As shown in Figure 10, the most frequently chosen options were monitoring students’ progress (92.3%), grading (84.6%), needs analysis (84.6%), and feedback provision (76.9%). From teachers’ intended rubric uses, it is possible to conclude that they believe rubrics can support learning, formative assessment and transparency in communicating assessment results (e.g., for the school’s principal or students’ families). The two least mentioned purposes were using the rubric to assist decisions in placement moments (46.2%) and accountability (38.5%).

Figure 10 – Brazilian teachers’ intended uses of the pronunciation assessment rubric



Source: the author (2025).

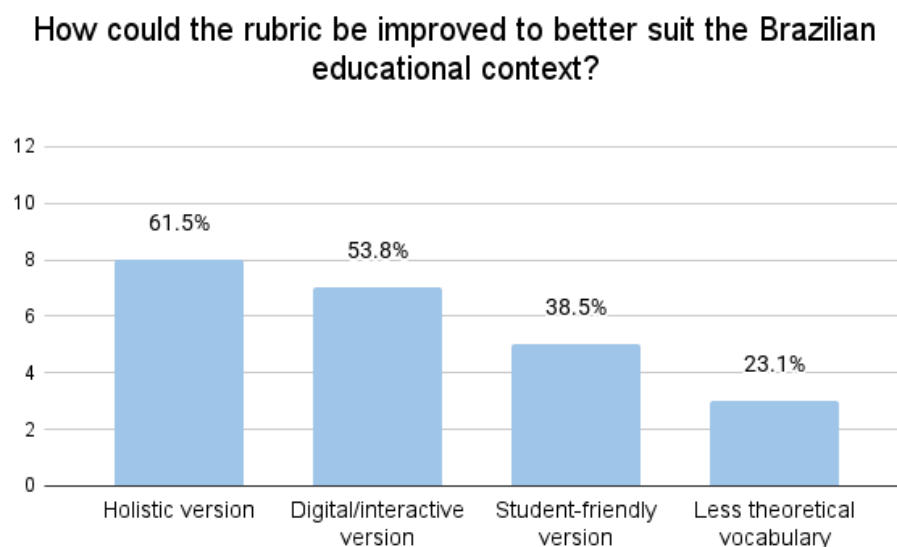
Notes: $N = 13$; Multiple-choice question.

With respect to possible rubric uses, Brookhart (2014) highlight that they are especially valuable and practical for teachers, because they serve as a bridge between formative and summative assessment. In other words, they give teachers the possibility to count on the same criteria and performance descriptors to promote learning-oriented assessment practices and, at the same time, assign valid numerical grades, which are mandatory for external reports. The participants in the present study, therefore, seem to share this view, perceiving the rubric as useful to fulfill pedagogical and accountability functions.

4.2.3.4 Suggestions to improve the rubric and the supporting guide for the Brazilian context

In the second-to-last question participants answered, they were invited to indicate how the rubric could be improved for the Brazilian educational context in the future. They selected all the options available, which point to many directions in which the instrument can be revised to better meet teachers' needs. Figure 11 visually presents that they are especially interested in a holistic version of the rubric (61.5%) and in a digital or more interactive version that automatically signals the performance level learners were assigned (53.8%). Other less mentioned suggestions include a student-friendly version of the rubric for self- and peer-assessment (38.5%) and simplifying the theoretical vocabulary in the instrument (23.1%).

Figure 11 – Teachers' suggestions on how to improve the pronunciation rubric for the Brazilian context



Source: the author (2025)
Notes: $N = 13$; Multiple-choice question.

Furthermore, some participants left additional comments in the “others” option of the same question. One teacher suggested developing and including ready-to-use tasks: “Perhaps the inclusion of pre-established tasks across different [development] levels, so that teachers always assess learners using similar tasks³²” (P2). His proposal is extremely relevant for future studies, since different types of tasks involve levels of linguistic complexity that are distinct. For instance, learners may show different degrees of pronunciation control depending on whether the task requires planned or spontaneous production (Derwing *et al.*, 2004). Accordingly, task characteristics might influence learners’ performance and how proficient they are perceived to be.

A second suggestion concerns including “a version with numerical grades (e.g., ‘avançado’ = 10; ‘progredindo’ = 9, etc.)”. P10 argues that, with this addition, it would “easier and more practical to calculate students’ average grade³³”. Offering guidance for teachers on this aspect is necessary, once simply converting students’ achievement as per rubric performance levels into numerical scores does not always grant an accurate representation of the development level they are (Brookhart, 2014). For example, whereas the grading system from most schools in Brazil represents learners’ progress with scores from 0 to 10, rubrics usually only have from 4 to 6 levels. Hence, it seems that the supporting guide could be revised to include a section to assist teachers during this practice.

The last question of the study investigated other resources that teachers would want to see included in the supporting guide that was developed to accompany the rubric. Most of them (46.2%) recommended including more illustrated examples of articulatory issues and speech samples to exemplify the rubric descriptors. Therefore, future revisions of the guide could focus on providing more multimodal examples to aid rubric users in comprehending some phonological features that they must understand to assess pronunciation. On the other hand, in spite of 30.8% of participants stating they were satisfied with the guide in its current form, three teachers (23.1%) indicated that it needs to be more concise for daily pedagogical practice.

To conclude, in regards to the supporting guide, another suggestion came from P12. He says: “[...] perhaps you [the researcher] could include links to external materials that explain

³² Original comment in Portuguese: “Talvez incluir tarefas direcionadas para diferentes níveis pré-estabelecidas. Assim o professor avalia sempre com tarefas semelhantes”.

³³ Original comment in Portuguese: “Uma versão com nota (ex.: avançado = 10; progredindo = 9, etc.), assim seria mais fácil e prático para calcular a média de cada estudante”.

the criteria more extensively in case teachers experience specific difficulties³⁴”. According to this comment, including supplementary resources for teachers’ self-study and professional development could be effective to prepare less confident and knowledgeable teachers to have a full comprehension of the pronunciation dimensions addressed in the rubric. As the supporting guide introduces the rubric criteria but does not discuss all of them in depth, more attention should be given to suprasegmental and prosodic aspects, considering they were not mentioned in any of the sections.

4.3 THE REVISED VERSION OF THE L2 PRONUNCIATION ASSESSMENT RUBRIC

After discussing Brazilian teachers’ considerations on rubric criteria, performance levels, and usability, this section will present the revised version of the L2 pronunciation analytic rubric developed to be used in classroom formative assessments. Nevertheless, it must be noted that “rubrics are always works in progress” (Arter; Chappuis, 2006, p. 80). Therefore, this version is treated as only a revision and not a final version. It is very possible that the rubric developed in this study still needs to be investigated through distinct perspectives and conditions before it can be used in Brazilian classrooms (for a complete discussion, see section 5.2). And even then, constant improvement will be needed to ensure that it reflects teachers’ realities and the learning outcomes set for students.

As participants’ responses indicated that they perceived the criteria as complete enough to represent what constitutes an intelligible L2 pronunciation, no improvements were made in this aspect. In consequence of that, most of the revisions focused on improving the rubric’s performance levels and descriptors, which were the elements that teachers most identified as requiring significant improvement. Firstly, two additional performance levels were included in the rubric, expanding it from four to six (in addition to the “unable to assess”) levels. Specifically, one level was inserted between the levels 1 and 2 and another between the levels 2 and 3. On this account, it is hoped that the rubric will more accurately reflect the progression of learners’ pronunciation in the L2, especially for beginner and intermediate cases.

In addition to that, with the expansion of the rubric scale, the names of the levels had to be reconsidered as well. Although they were originally assigned written names only to avoid misinterpretation with numerical grades (Brookhart, 2014), teachers reported that the words

³⁴ Original comment in Portuguese: “Talvez você poderia incluir links para materiais externos que mais extensamente expliquem os critérios caso haja alguma dificuldade específica (mas isso seria algo adicional mesmo, pois já está ótimo)”.

chosen created an unintentional leap in the progression from one level to the next, and they did not always reflect learners' ability as described in that level. Thus, with the need for additional performance levels, certain labels were modified to correspond to the finer distinctions that are now part of the rubric. For example, in the revised version, the word *básico* is used to represent students with pronunciation proficiency superior to that of *começando*.

Another element that was mentioned a lot of times was the presence of examples in the descriptors. They seem to have supported teachers' final assessment decisions, as participants reported that explicit examples facilitated interpreting learners' pronunciation development. In response, in performance levels and criteria in which they were not present previously, examples were included whenever possible. While the supporting guide already addresses recurrent L2 pronunciation issues that could appear in Brazilian learners of English production, having concrete illustrations within the rubric itself seems to have offered further guidance to teachers.

Finally, another important change in the revised version of the rubric involves its design and layout. In the first version, the criteria were displayed on the left side and the performance levels presented across the top. The new version organizes each criterion in a separate section. Additionally, there are only two vertical columns, one for the performance levels and one for the criteria. In accordance with Isaacs, Trofimovich and Foote (2018)'s recommendation, dotted lines splitting the criteria were included to emphasize that teachers can adapt the rubric at their discretion. All the changes are highlighted in blue to improve readability. The revised version of the rubric is presented below, in Table 12.

Table 12 – The revised version of the L2 pronunciation assessment rubric

Nível	Vogais
6 Proficiente	As vogais são articuladas com clareza, sem deixar dúvidas quanto à palavra pronunciada. Há consistência na produção de todos os sons vocálicos difíceis para aprendizes brasileiros (por exemplo, 'sheep' e 'ship' são pronunciadas de maneira claramente distintas).
5 Avançado	As vogais são geralmente articuladas com clareza, mas apresentam substituições ou acréscimos ocasionais que não afetam a compreensão. Há pequenas variações na produção de sons vocálicos difíceis, mas a pronúncia permanece precisa (por exemplo, 'sit' e 'seat' são geralmente pronunciadas de maneira distinta).

4 Progredindo	As vogais são compreensíveis na maior parte do tempo, mas há substituições ocasionais que podem dificultar o entendimento de algumas palavras. A produção de sons vocálicos difíceis é inconsistente, com momentos de acerto alternados com desvios (por exemplo, ‘pull’ e ‘pool’).
3 Aprendendo	As vogais são compreensíveis, mas há substituições ou acréscimos frequentes que dificultam o entendimento de várias palavras. Há pouca estabilidade na produção de sons vocálicos difíceis, o que pode tornar a fala menos clara (‘cut’ > ‘cat’).
2 Básico	As vogais são geralmente compreensíveis, mas há pouca consistência no contraste entre alguns sons vocálicos, o que pode gerar confusão. Sons vocálicos difíceis são frequentemente substituídos e a compreensão acontece por conta do contexto (‘lock’ > ‘look’).
1 Começando	As vogais são compreensíveis apenas em alguns momentos, mas não há distinção clara entre sons vocálicos semelhantes, o que dificulta a diferenciação de palavras e prejudica significativamente a compreensão.
0 Não demonstrado	Ainda não é possível avaliar (ausência de fala, respostas muito curtas ou produção desconexa do que foi solicitado).
Nível	Consoantes
6 Proficiente	As consoantes sonoras são articuladas com clareza, permitindo fácil reconhecimento da palavra pronunciada. As consoantes surdas apresentam aspiração apropriada quando necessário (/p/, /t/ e /k/). Nenhum som é substituído por outro no lugar das consoantes esperadas.
5 Avançado	A maioria das consoantes sonoras são articuladas com clareza, com poucas substituições que não comprometem a compreensão. A aspiração de consoantes surdas ocorre de forma consistente na maioria das vezes. Erros de substituição (‘tink’ ou ‘fink’ > ‘think’) acontecem com sons difíceis para brasileiros, mas o contexto apoia a compreensão.
4 Progredindo	As consoantes sonoras são geralmente compreensíveis, mas com articulação irregular em palavras novas/menos familiares. A aspiração em consoantes surdas aparece ocasionalmente. Substituições ocorrem com frequência moderada, especialmente em consoantes de maior dificuldade

	para brasileiros, podendo gerar confusão quando o contexto não é suficiente ('sheep' > 'cheap').
3 Aprendendo	Há dificuldade na articulação clara de consoantes sonoras, ainda que elas sejam compreensíveis em palavras familiares. A aspiração em consoantes surdas ocorre de forma inconsistente. Substituições são frequentes em pares difíceis para brasileiros ('hose' > 'rose'), prejudicando a compreensão em palavras imprevisíveis pelo contexto.
2 Básico	A articulação das consoantes sonoras é pouco clara na maior parte do tempo, dificultando a identificação de palavras. A aspiração é raramente produzida. Substituições são comuns e geralmente resultam em palavras que não correspondem ao que se pretendia dizer ('at' > 'hat').
1 Começando	As consoantes sonoras são raramente articuladas com clareza, sendo difícil entender qual palavra está sendo pronunciada. A aspiração não é produzida. Substituições são muito comuns e comprometem significativamente a inteligibilidade da fala ('noce' > 'nose').
0 Não demonstrado	Ainda não é possível avaliar (ausência de fala, respostas muito curtas ou produção desconexa do que foi solicitado).
Nível	Prosódia (ritmo, entoação e acentuação)
6 Proficiente	Demonstra controle claro de padrões da fala. O acento tônico das palavras (word stress) está corretamente colocado, facilitando a compreensão (PHOtograph/photoGRAPHic). O uso do acento frasal (sentence stress) destaca com precisão as palavras-chave da frase. O ritmo da fala é natural e há o uso eficiente de sílabas fracas e fortes (I WANT to GO to the MARket). A entonação sempre se adequa ao propósito comunicativo, indicando perguntas, surpresas, afirmações, etc.
5 Avançado	Demonstra um controle geral dos padrões da fala. O acento tônico costuma estar bem posicionado, com algumas variações que não comprometem a compreensão (HOTel > hoTEL). Palavras-chave geralmente recebem destaque na frase, ainda que às vezes o ritmo soe menos natural ou com sílabas igualmente fortes (I WANT TO GO to the MARket). A entonação é usada de forma consistente para os propósitos comunicativos, mas com pouca variação.
4 Progredindo	Demonstra alguma regularidade dos padrões da fala. O acento tônico está presente em palavras familiares, mas varia em palavras novas ou longas,

	podendo causar leve dificuldade de compreensão. Palavras-chave se destacam, ainda que de maneira irregular. O ritmo alterna entre sílabas fortes e fracas, mas com irregularidades (I want TO go TO the mar KET). A entonação indica propósitos comunicativos simples (por exemplo, pergunta).
3 Aprendendo	Há tentativas de aplicar os padrões da fala. O acento tônico das palavras pode aparecer em posições inesperadas, o que pode dificultar a compreensão em alguns momentos. Palavras-chave raramente se destacam. O ritmo tende a ser regular, sem contraste claro entre sílabas fortes e fracas (I WANT to go to the market). A entonação aparece de forma limitada, mas com potencial para desenvolvimento.
2 Básico	Os padrões da fala aparecem de forma limitada. O acento tônico frequentemente é colocado em posições inesperadas, dificultando a compreensão mesmo em palavras comuns. O destaque de palavras-chave acontece esporadicamente. O ritmo tende a ser plano ou irregular, com pouco contraste entre sílabas fortes e fracas (i want to GO to the MARK et). A entonação é usada de forma restrita, dificultando identificar intenção comunicativa.
1 Começando	Os padrões da fala são reconhecidos poucas vezes. O acento tônico das palavras muitas vezes não segue o esperado, afetando a compreensão. As frases tendem a soar monótonas ou com ritmo irregular (i want to go to the market). A entonação é pouco explorada, dificultando perceber intenções (como momentos de pergunta ou ênfase).
0 Não demonstrado	Ainda não é possível avaliar (ausência de fala, respostas muito curtas ou produção desconexa do que foi solicitado).
Nível	Fluência e habilidade de autocorreção
6 Proficiente	A fala flui de forma contínua, com um número elevado de palavras considerando o tempo de produção. As pausas ocorrem em pontos apropriados (como entre ideias ou frases) e não afetam a compreensão. O ritmo é mantido mesmo quando é necessário corrigir rapidamente algum engano. Autocorreções são feitas sem interromper a comunicação
5 Avançado	A fala flui geralmente de forma contínua, com um número consistente de palavras considerando o tempo de produção. Algumas pausas ou aumento da velocidade ocorrem em momentos inesperados, mas não afetam a

	comunicação. Há tentativas de correção de enganos, ainda que esse processo possa levar mais tempo e impactar levemente o fluxo da fala.
4 Progredindo	A fala flui em alguns momentos, mas tem pausas que às vezes interrompem o ritmo da comunicação. O número de palavras é suficiente para transmitir ideias básicas, ainda que possa variar conforme a dificuldade do assunto. A autocorreção ocorre, mas geralmente impacta o fluxo da fala.
3 Aprendendo	A fala tem pausas frequentes que podem interromper o fluxo da comunicação, deixando-o mais lento ou confuso. O número de palavras é limitado para o tempo. Necessita de tempo para organizar o que quer dizer e para se autocorrigir em momentos necessários. OU A fala é muito rápida e é difícil acompanhar o fluxo da comunicação. O número de palavras é elevado para o tempo, mas a pronúncia das mesmas é geralmente equivocada. Necessita de tempo para se autocorrigir em momentos necessários.
2 Básico	A fala é marcada por pausas longas ou muito frequentes, dificultando a continuidade das ideias. Poucas palavras são produzidas dentro do tempo disponível. A autocorreção ocorre raramente ou apenas após longas hesitações, impactando de forma evidente a compreensão.
1 Começando	A fala é fragmentada, com pausas longas ou frequentes que dificultam o fluxo da comunicação. Poucas palavras são produzidas e o esforço para formar frases é evidente. A autocorreção é rara ou ocorre de forma ineficaz.
0 Não demonstrado	Ainda não é possível avaliar (ausência de fala, respostas muito curtas ou produção desconexa do que foi solicitado).
Nível	Formação silábica
6 Proficiente	Não há adição ou omissão de sons na formação silábica, mantendo o mesmo número de sílabas esperado. Encontros consonantais complexos (world) para brasileiros e consoantes em posição final de palavra (heart/spam/well) são produzidas adequadamente.

5 Avançado	Poucas ocorrências de adição ou omissão de sons na formação de sílabas. Encontros consonantais e consoantes finais são produzidas corretamente, ainda que ocasionalmente exista simplificação dos sons ou aumento do número de sílabas esperado (‘istop’ > ‘stop’ / ‘milki’ > ‘milk’).
4 Progredindo	Ocorrem adições ou omissões de sons com alguma frequência, mas não em todas as palavras. Encontros consonantais e consoantes finais mostram produção parcialmente correta, com simplificações previsíveis (‘tes’ > ‘test’). A compreensão geralmente é possível com apoio do contexto.
3 Aprendendo	Ocorrências moderadas de adição ou omissão de sons na formação de sílabas, aumentando ou diminuindo o número esperado. As mudanças nos encontros consonantais e em consoantes finais podem interferir na pronúncia (‘istriti’ > ‘street’ / ‘fin’ > ‘fine’), mas algumas palavras são compreendidas com ajuda do contexto.
2 Básico	Adições e omissões de sons acontecem com muita frequência, alterando de maneira significativa o número de sílabas esperadas. Encontros consonantais raramente são produzidos. Consoantes finais quase nunca aparecem, o que dificulta bastante a compreensão mesmo com a ajuda do contexto (‘as’ > ‘ask’).
1 Começando	O número de sílabas é sempre muito maior ou menor que o esperado. Encontros consonantais são frequentemente simplificados e consoantes finais não são produzidas, comprometendo significativamente a compreensão mesmo com a ajuda do contexto (‘lukéd’ > ‘looked’).
0 Não demonstrado	Ainda não é possível avaliar (ausência de fala, respostas muito curtas ou produção desconexa do que foi solicitado).
Nível	Compreensibilidade
6 Proficiente	A fala é extremamente fácil de compreender.
5 Avançado	A fala é fácil de compreender.
4 Progredindo	A fala geralmente é fácil de compreender, mas algumas palavras ou trechos específicos exigem mais esforço do ouvinte.

3 Aprendendo	A fala exige um pouco de esforço do ouvinte para ser compreendida.
2 Básico	A fala exige um esforço considéravel do ouvinte para ser compreendida.
1 Começando	A fala exige muito esforço do ouvinte para ser compreendida.
0 Não demonstrado	Ainda não é possível avaliar (ausência de fala, respostas muito curtas ou produção desconexa do que foi solicitado).

Source: the author (2025).

As a result of teachers' final recommendations to enhance the supporting guide, it (see Appendix E) now includes a section outlining how the rubric results can be adapted to the grades usually required by the Brazilian educational system. This addition is grounded on Brookhart's (2014) suggestions on alternative ways to translate the developmental stages of rubrics into a numerical score. Furthermore, resources for self-study were incorporated in a new section to support teachers who feel they need a deeper understanding of certain pronunciation constructs, especially those they reported did not receive much attention in their pronunciation assessment practices (e.g., prosody, syllable structure, etc.).

Despite that, there are many revisions that can still be conducted in future versions of the rubric, considering the in-service Brazilian English teachers' insights reported in this study. Accordingly, the next and final chapter presents the conclusions drawn from the findings, highlights the limitations of the study, and indicates possibilities for future research.

5 CONCLUSIONS

Considering the pedagogical potential attributed to rubrics in indicating with transparency what quality learning is (Brookhart, 2014) and the complexity of effectively and fairly assessing pronunciation in the L2 classroom (Isbell; Sakai, 2022), this study aimed to develop and refine an analytic rubric for L2 pronunciation formative assessment that was sensitive and responsive to Brazilian teachers' pedagogical needs and contextual constraints. The rubric was developed to support teachers in having more consistent perceptions of Brazilian students' oral development and in identifying pronunciation aspects to address in instruction. In order to achieve this, thirteen in-service Brazilian teachers participated in the data collection and used the rubric to assess speech samples from Brazilian university students with English as an L2. Using a mixed-methods approach, the rubric was analyzed through Brookhart (2014) desired characteristics to write high-quality criteria and performance levels and Arter and McTighe (2001) notion of usability.

The first research question guiding the study was "How do Brazilian teachers of English evaluate an L2 pronunciation assessment rubric in relation to its criteria, performance levels, and usability?". Regarding rubric criteria, the results indicate that teachers perceived them as complete and corresponding to pronunciation aspects important to a view oriented towards intelligibility in L2 speech production. No changes were made to this aspect of the rubric. Nevertheless, the results have two important implications. First, teachers mentioned they would need to provide extensive scaffolding for students to be able to understand the criteria. Second, they would need to adapt the rubric criteria (i.e., use only some of them) in order to implement the rubric in larger class sizes.

In relation to rubric performance levels, teachers indicated they were descriptive, clear, and included the full spectrum of pronunciation development. Yet, some participants reported that some levels needed to be improved to better reflect the pronunciation performance they intended to represent. In the descriptors, the inclusion of examples was seen as a positive design feature of the rubric, as they supported teachers' decisions during assessment. Another important finding for rubric developers of oral production in Brazil indicates that the length of the descriptors should be reduced to a more realistic and easier rubric use. Additionally, once more, teachers highlighted the student's perspective on rubric use, noting that the descriptors might be too complex for learners to understand and apply independently.

With respect to usability, the results indicate that teachers believed the rubric facilitated the assessment of learners' L2 pronunciation. Particularly, the presence of criteria

and rich, detailed performance descriptors was mentioned as something positive that enhanced teachers' assessment experience. In addition, the rubric seems to have drawn teachers' attention to features they would not normally consider in speech production (for instance, prosody was perceived to be important for message and meaning conveyance after using the rubric). This provides evidence that engaging with rubric-mediated assessment can increase teachers' awareness of what constitutes an effective and intelligible pronunciation. One last finding regarding usability indicates that teachers are willing to use the rubric for distinct assessment purposes in classroom contexts. Notwithstanding, time limitations may be a factor that most of them would consider before implementing it.

The second research question guiding this study was "How can the experiences and insights emerging from teachers' evaluation inform a revised version of the L2 pronunciation assessment rubric?". The findings point to many directions for refinement that foreground teachers' pedagogical realities and professional and contextual needs, which reinforces the importance of including rubric end-users in its development processes. Participants emphasized that some adaptations in the format and language of the rubric would be necessary to enhance usability for classroom use. Some adaptations, such as expanding the rubric from four to six to consider comments suggesting there was not a fine distinction between the lowest and middle levels in the scale, were already conducted in this thesis. Moreover, resources for self-study and a proposal to translate criterion-referenced results into numerical grades were included in the supporting guide.

Nonetheless, many other suggestions could not be incorporated into the rubric or the supporting guide either because they were outside the scope of the present study or due to time limitations. Yet, as rubric development should be seen as a continuous process (Galaczi *et al.*, 2011; Brookhart, 2014; Harding, 2017), they now become an invitation for other researchers interested in rubric development and classroom pronunciation assessment. There are several other revisions that can be carried out in the rubric alone. For instance, one of them would be developing a holistic version of the rubric to facilitate and encourage adopting rubric-mediated assessment even in contexts where its use might be challenging. Other revisions could offer complementary analytic and holistic approaches in order to provide teachers the freedom to decide on the format that offers more meaningful information to enhance students' pronunciation.

Ultimately, the study recognizes the difficulty of assessing learners' L2 pronunciation in the classroom, particularly in the Brazilian context, where teachers often have a limited professional preparation to address phonological components and are not provided with

sufficient time or encouragement to include pronunciation as an integral part of language classes and L2 proficiency development. At the same time, it offers empirical evidence on how to develop an assessment instrument that incorporates teachers' contextual knowledge and perceptions during its design process as a possibility to circumvent these challenges. One of the main contributions of the study is demonstrating that in-service Brazilian English teachers are inclined to assess L2 pronunciation formatively in the classroom with rubrics as long as they receive adequate support and the instrument presents sufficient flexibility for distinct pedagogical realities.

5.1 PEDAGOGICAL IMPLICATIONS

The findings of the study led to significant pedagogical implications for the fields of rubric development and teacher education in L2 pronunciation assessment. In regard to the former, the results showed that teachers' responses regarding rubric usability varied according to the pedagogical context they are part of, as each context presents distinct specificities. This suggests that rubric design should try to account for possible adaptations in order to meet different assessment purposes and classroom realities. Given the time and effort necessary to create and validate analytic general rubrics (Arter; Chappuis, 2006), teachers should therefore have access to multiple rubrics from which to choose and adapt. This flexibility and wide selection of options could encourage them to integrate pronunciation more consistently in the assessment of students' L2 oral production.

Additionally, rubric developers should value rubric implementation strategies that view students' participation in the assessment process as equally important to effective rubric use. It is expected that, especially with rubrics to assess L2 pronunciation, often seen as a construct challenging to be comprehended in practice, younger or beginner learners may not be ready to interpret them effectively for peer- and self-assessment without teacher mediation. Therefore, rubric developers should include explicit instructions on how teachers can introduce these instruments to students, model how they are used, and enhance student engagement with the rubric. As Andrade (2005) argues, the way teachers implement a rubric can be as important as how well-designed they are to, in fact, improve student learning and promote greater achievement in pronunciation development.

When it comes to the latter, the findings indicate that L2 teacher education programs should have a greater emphasis on rubric development and pronunciation assessment. Participants' limited experience with rubric use shows that, the importance of rubrics during

formative assessment practices and how they can assist teaching and learning, especially for pronunciation, are not yet totally incorporated in teacher education curricula. Thus, more professional development opportunities focused on rubrics for pre-service and in-service teachers are necessary to support them in developing a clear perception of rubrics' characteristics and purposes, which would empower teachers to be informed of when and how they best support instruction and the development of certain L2 abilities. Additionally, teachers should have opportunities to critically analyze existing rubrics, adapt them to different learning groups, and reflect on how complex constructs were enacted through assessment criteria in order to enhance more student-centered assessments.

Accordingly, interaction with the rubric during the assessment process seemed to have raised teachers' awareness of what entails pronunciation quality in the L2 and invited them to reconsider traditional approaches to pronunciation assessment, changing their focus from segmental accuracy to other important dimensions, such as prosody. In this regard, as teachers reported that rubric led them to pay attention to features they would normally disregard and to adopt more criterion-referenced judgments, it can be argued that engaging with rubrics for assessment positions them as potential "[...] catalyst[s] for teacher reflection" (Kennedy; Shield, 2022, p. 142), which may contribute to increase their pedagogical knowledge and practice. This finding highlights the role of rubrics to function as resources for professional development, as they may have made the nature of the pronunciation construct more visible to teachers.

Finally, the findings have implications for curricular guidance and education policy as well. As teachers reported different assessment practices and perceptions of the importance of pronunciation assessment in the classroom, it is necessary that national documents that regulate English education in Brazil, such as BNCC (Brasil, 2018), orientate teachers more clearly about how to assess students' pronunciation and which formative practices can be adopted to do it. It is possible that, due to detailed and official orientations, teachers would feel more confident when integrating pronunciation into oral production assessments in their classroom. Additionally, this movement could reaffirm the role of pronunciation as an important component in L2 learning.

5.2 LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

Despite its relevant contributions to the field of rubric development and L2 pronunciation assessment, this study presents several limitations that should be acknowledged and signal opportunities for future research. First, it is important to mention that not all the data

collected were analyzed in this thesis³⁵. In addition, as the study had a limited number of participants, the results cannot be generalized to all Brazilian English teachers. And, still regarding participants, it is prudent to assume that the recruitment method that was used possibly attracted teachers who already had some kind of interest in pronunciation assessment or rubrics. As a consequence, future investigations would benefit from a larger and more diverse group of participants, as well as attempt to include teachers with distinct levels of interest and familiarity with pronunciation assessment.

To continue, it is important to discuss some limitations regarding the methodological choices made. To be mindful of Brazilian teachers' time and voluntary dedication, the study analyzed participants' perception primarily with surveys, considering they were instruments that could be completed individually at any moment of preference. Nonetheless, research on rubric development usually collects data through focus groups and uses a repeated rubric refinement approach (Isaacs; Trofimovich, Foote; 2018; Harding, 2017; Galaczi *et al.*, 2011). In such studies, the focus group sessions allow participants to use and analyze successive versions of the instrument at hand. Because of that, new findings are reported as changes are implemented. Using a similar study design in Brazil to understand teachers' perspectives throughout different moments of rubric development could reveal issues in rubric design and usability that come to the surface only after participants have extensive contact with the instrument.

Another limitation was the nature of the assessment task employed in the study, as well as the profile of learners analyzed through the speech samples. Teachers assessed university students' pronunciation only on a task type, namely, a non-planned reading aloud activity. This decision was important to control for participants' experience with the rubric during the data collection, but it does not represent the wide range of tasks to elicit L2 oral speech that teachers often use in classroom assessments or, even, the expected performance of learners from basic education. Consequently, it is important to investigate how teachers perceive the rubric after assessing different task types and student groups, as well as explore whether they exhibit the same level of consistency in assessment results seen in this study. This could provide further evidence of the rubric's pedagogical completeness and versatility for future implementation in Brazilian classrooms.

In addition, the data collection design used in this study does not entirely reflect the conditions under which teachers usually assess oral production in real classrooms, especially

³⁵ Data that was not relevant to answer the research questions will be analyzed in future studies.

when dealing with multiple students' performances in sequence, something that often occurs during oral presentations, for example. In light of how demanding analytic instruments are for users' mental effort (Knoch, 2017), future research could explore teachers' cognition and practices using the rubric in authentic moments during real-time, interactive assessment. Classroom observation studies conducted in, particularly public schools from Brazil, could deepen our understanding of the rubric's usability and how to negotiate descriptive information about learners' oral development for formative purposes and realistic rubric implementation.

In conclusion, as mentioned before, considering teachers' concerns related to time constraints, future research could focus on developing a holistic version of the rubric grounded in the analytic criteria and performance levels descriptors proposed in this study. After its development, teachers could be invited to use the two versions to assess the same speech samples and compare the instruments in terms of reliability and practicality for pedagogical implementation. Most importantly, a final recommendation regards rubric impact on student learning. More longitudinal studies are needed to investigate the effects of rubrics on learners' L2 pronunciation development. While empirical evidence suggests that implementing rubrics in classroom assessment practices can positively foster students' awareness and self-assessment in writing components (e.g., Andrade; Du; Mycek, 2010; Bradford *et al.*, 2016; Kim, 2019), research exploring how and to what extent they are effective in improving L2 oral production is still limited.

REFERENCES

- ANDRADE, H. L. Teaching with rubrics: The good, the bad, and the ugly. **College teaching**, v. 53, n. 1, p. 27-31, 2005.
- ANDRADE, H. L. Summing up and moving forward: Key challenges and future directions for research and development in formative assessment. *In*: ANDRADE, H. L.; CIZEK, G. J. (Ed.). **Handbook of Formative Assessment**. Routledge, 2010. p. 344-351.
- ANDRADE, H. L. What Is Next for Rubrics?. *In*: GOLSAVES, C.; PEARSON, J. (Ed.). **Improving Learning Through Assessment Rubrics: Student Awareness of What and How They Learn**. IGI Global, 2023. p. 314-326.
- ANDRADE, H. L.; DU, Y.; MYCEK, K. Rubric-referenced self-assessment and middle school students' writing. **Assessment in Education: Principles, Policy & Practice**, v. 17, n. 2, p. 199-214, 2010.
- ARTER, J. A.; CHAPPUIS, J. **Creating and Recognizing Quality Rubrics**. Boston: Pearson, 2006.
- ARTER, J.; MCTIGHE, J. **Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance**. Corwin Press, 2001.
- BACHMAN, L. F. Ongoing Challenges in Language Assessment. *In*: _____. **The Writings of Lyle F. Bachman**. Routledge, 2013. p. 438-458.
- BACHMAN, L. F.; PALMER, A. **Language testing in practice**. Oxford: Oxford University Press, 1996.
- BALLARD, L.; WINKE, P. Students' attitudes towards English teachers' accents: The interplay of accent familiarity, comprehensibility, intelligibility, perceived native speaker status, and acceptability as a teacher. *In*: ISAACS, T.; TROFIMOVICH, P. (Ed.). **Second language pronunciation assessment: Interdisciplinary perspectives**. Multilingual Matters, 2017. p. 121-140.
- BANEGAS, D. L.; CONSOLI, S. Action research in language education. *In*: MCNAMARA, T.; PLONSKY, L. **The Routledge handbook of research methods in applied linguistics**. Routledge, 2020. P. 176-187.
- BLACK, P.; WILIAM, D. Assessment and classroom learning. **Assessment in Education: principles, policy & practice**, v. 5, n. 1, p. 7-74, 1998.
- BRADFORD, K. L. *et al.* Rubrics as a tool in writing instruction: Effects on the opinion essays of first and second graders. **Early Childhood Education Journal**, v. 44, n. 5, p. 463-472, 2016.
- BRASIL. Ministério da Educação. **Base Nacional Comum Curricular**. Brasília, 2018.
- BRINDLEY, G. Language assessment and professional development. *In*: ELDER, A. *et al.* **Experimenting with uncertainty: Essays in honour of Alan Davies**. Cambridge University Press, 2001. p. 126-135.

BROWN, A. Syllable structure. *In*: REED, M.; LEVIS, J. M. (Ed.). **The Handbook of English Pronunciation**. John Wiley & Sons, 2015.

BROOKHART, S. M. **How to create and use rubrics for formative assessment and grading**. Ascd, 2014.

BROOKHART, S. M. Using Rubrics in Basic Education: A review and Recommendations. **Estudos em Avaliação Educacional**, v. 35, p. 1-22, 2024. DOI: http://educa.fcc.org.br/scielo.php?pid=S0103-68312024000100202&script=sci_arttext&tlng=en. Available at: http://educa.fcc.org.br/scielo.php?script=sci_arttext&pid=S0103-68312024000100202&lng=pt&nrm=iso. Access on 5. May 2025.

BROOKHART, S. M.; CHEN, F. The quality and effectiveness of descriptive rubrics. **Educational Review**, v. 67, n. 3, p. 343-368, 2015.

BØHN, H.; HANSEN, T. Assessing pronunciation in an EFL context: Teachers' orientations towards nativeness and intelligibility. **Language Assessment Quarterly**, v. 14, n. 1, p. 54-68, 2017.

BUSS, L. Beliefs and practices of Brazilian EFL teachers regarding pronunciation. **Language Teaching Research**, v. 20, n. 5, p. 619-637, 2016.

CAMARGO, C. G. **Crenças e práticas de professores formadores sobre o ensino da compreensão oral em língua inglesa**. 2020. 240 f. Tese (Doutorado em Letras) — Universidade Federal do Paraná, Curitiba, 2020.

CELCE-MURCIA, M. *et al.* **Teaching pronunciation: A reference for teachers of English**. 2^a ed. New York: Cambridge University Press, 2010.

CHAPPUIS, J. *et al.* **Classroom assessment for student learning: Doing it right - using it well**. New York: Pearson, 2012.

CHINDA, B. Considerations in Performance-Based Language Assessment: Rating Scales and Rater Training. **PASAA**, v. 46, n. 1, p. 141-160, 2013.

CHONG, S. W.; ISAACS, T. An Ecological Perspective on Classroom-Based Assessment. **Tesol Quarterly**, v. 57, n. 4, p. 1558-1570, 2023.

COOK, V. The goals of ELT: Reproducing native-speakers or promoting multicompetence among second language users?. *In*: CUMMINS, J.; DAVISON, C. (Ed.). **International handbook of English language teaching**. Boston, MA: Springer US, 2007. p. 237-248.

COUNCIL OF EUROPE. **Common European Framework of Reference for Languages: learning, teaching, assessment – companion volume with new descriptors**. Strasbourg, 2018. Available at: <https://rm.coe.int/cefr-companionvolume-with-new-descriptors-2018/1680787989>. Access on 25 Sep. 2024.

COSTA, B. C. A. **Pronunciation teaching is not a one-size-fits-all endeavor: EFL teachers' beliefs and classroom practices**. 2016. 112 f. Dissertação (Mestrado em Letras) — Universidade Federal de Santa Catarina, Florianópolis, 2016.

- ÇELİK, F. *et al.* Modelling the complex dynamic effect of L2 speaking anxiety in metaverse on students' academic enjoyment, pronunciation anxiety, speaking performance, and task load. **British Educational Research Journal**, 2025. DOI: <https://doi.org/10.1002/berj.70062>.
- DAMACENA, M.; QUEVEDO-CAMARGO, G. Avaliação e formação de professores de línguas: uma discussão sobre o currículo e as percepções dos formandos. **Olhares & Trilhas**, v. 23, n. 3, p. 1054-1073, 2021.
- DARCY, I.; EWERT, D.; LIDSTER, R. Bringing pronunciation instruction back into the classroom: an ESL teachers' pronunciation toolbox. *In*: LEVIS, J.; LEVELLE, K. (org.). **Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference**. Ames: Iowa State University, 2012. p. 93-108. Available at: <https://hdl.handle.net/2022/22866>. Access on: 28 Jan. 2026.
- DAVIES, A. Textbook trends in teaching language testing. **Language Testing**, v. 25, n. 3, p. 327-347, 2008. DOI: <https://doi.org/10.1177/0265532208090156>.
- DAWSON, P. Assessment rubrics: towards clearer and more replicable design, research and practice. **Assessment & Evaluation in Higher Education**, v. 42, n. 3, p. 347-360, 2017.
- DERWING, T. M.; MUNRO, M. J. Pronunciation learning and teaching. *In*: DERWING, T. M.; MUNRO, M. J.; THOMSON, R. I. **The Routledge handbook of second language acquisition and speaking**. Routledge, 2022. p. 147-159.
- DERWING, T. M.; MUNRO, M. J. Comprehensibility as a factor in listener interaction preferences: Implications for the workplace. **Canadian Modern Language Review**, v. 66, n. 2, p. 181-202, 2009.
- DERWING, T. M.; MUNRO, M. J. Accent, intelligibility, and comprehensibility: Evidence from four L1s. **Studies in second language acquisition**, v. 19, n. 1, p. 1-16, 1997.
- DORNYEI, Z. **Research methods in applied linguistics**. UK: Oxford University Press, 2007.
- EARL, L. M. **Assessment as learning: Using classroom assessment to maximize student learning**. Corwin press, 2003.
- FREITAS, L. C. *et al.* **Avaliação educacional: caminhando pela contramão**. Editora Vozes Limitada, 2009.
- FULCHER, G. Assessment literacy for the language classroom. **Language Assessment Quarterly**, v. 9, n. 2, p. 113-132, 2012.
- FULCHER, G. Assessing second language speaking. **Language Teaching**, v. 48, n. 2, p. 198-216, 2015.
- FOX, J.; ABDULHAMID, N.; TURNER, C. E. Classroom-based assessment. *In*: FULCHER, G.; HARDING, L. (Ed.). **The Routledge Handbook of Language Testing**. New York, NY: Routledge, 2012. p. 119-135.
- GALACZI, E. D. *et al.* Developing assessment scales for large-scale speaking tests: A multiple-method approach. **Assessment in Education: Principles, Policy & Practice**, v. 18, n. 3, p. 217-237, 2011.

GALLEGO-ARRUFAT, M. J.; DANDIS, M. Rubrics in a secondary mathematics class. **International Electronic Journal of Mathematics Education**, v. 9, n. 1-2, p. 75-84. Available at: <https://www.iejme.com/download/rubrics-in-a-secondary-mathematics-class.pdf>. Access on 27 Jan. 2026.

GIRALDO, F. Language assessment literacy and teachers' professional development: A review of the literature. **Profile: Issues in Teachers' Professional Development**, v. 23, n. 2, p. 265-279, 2021. <https://doi.org/10.15446/profile.v23n2.90533>

GU, P. Y.; YU, G. Researching classroom-based assessment for formative purposes. **Chinese Journal of Applied Linguistics**, v. 43, n. 2, p. 150-168, 2020.

HAMP-LYONS, L. The impact of testing practices on teaching: Ideologies and alternatives. In: CUMMINS, J.; DAVISON, C. (Ed.). **International Handbook of English Language Teaching**. Boston, MA: Springer US, 2007. p. 487-504.

HARDING, L. What do raters need in a pronunciation scale? The user's view. In: ISAACS, T.; TROFIMOVICH, P. (Ed.). **Second language pronunciation assessment: Interdisciplinary perspectives**. Multilingual Matters, 2017, p. 12-34.

HAUS, C. **Ensino de pronúncia sob a perspectiva do inglês como língua franca: crenças e práticas de professores de inglês do CELIN-UFPR**. 2018. 171 f. Dissertação (Mestrado em Letras) – Universidade Federal do Paraná, Curitiba, 2018.

INBAR-LOURIE, O. Constructing a language assessment knowledge base: A focus on language assessment courses. **Language testing**, v. 25, n. 3, p. 385-402, 2008.

INBAR-LOURIE, O. Language Assessment Literacy: Where to from Here?. **Language Testing**, v. 30, p. 505-519, 2013.

JEONG, H. Rubrics in the classroom: do teachers really follow them?. **Language Testing in Asia**, v. 5, n. 1, p. 6, 2015. DOI: <https://doi.org/10.1186/s40468-015-0013-5>.

JONES, J.; ISAACS, T. Assessing second language pronunciation. In: MOHEBBI, H.; COOMBE, C. **Research Questions in Language Education and Applied Linguistics**. Springer, 2021. p. 305-310.

ISAACS, T.; TROFIMOVICH, P. Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. **Studies in Second Language Acquisition**, v. 34, n. 3, p. 475-505, 2012.

ISAACS, T. Shifting sands in second language pronunciation teaching and assessment research and practice. **Language Assessment Quarterly**, v. 15, n. 3, p. 273-293, 2018.

ISAACS, T.; THOMSON, R. I. Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. **Language Assessment Quarterly**, v. 10, n. 2, p. 135-159, 2013.

ISAACS, T.; TROFIMOVICH, P. Key Themes, constructs and interdisciplinary perspectives in second language pronunciation assessment. In: _____. (Ed.). **Second language pronunciation assessment: Interdisciplinary perspectives**. Multilingual Matters, 2017. p. 3-8.

- ISAACS, T.; TROFIMOVICH, P.; FOOTE, J. A. Developing a user-oriented second language comprehensibility scale for English-medium universities. **Language Testing**, v. 35, n. 2, p. 193-216, 2018.
- ISBELL, D. R. Can the test support student learning? Validating the use of a second language pronunciation diagnostic. **Language Assessment Quarterly**, v. 18, n. 4, p. 331-356, 2021.
- ISBELL, D. R.; SAKAI, M. Pronunciation assessment in classroom contexts. *In*: LEVIS, J. M.; DERWING, T. M.; SONSAAT-HEGELHEIMER, S. (Ed.) **Second language pronunciation: Bridging the gap between research and teaching**. Wiley Blackwell, 2022. p. 194-214.
- KANG, O.; KERMAD, A. Assessment in second language pronunciation. *In*: KANG, O.; THOMSON, R.; MURPHY, J. **The Routledge Handbook of Contemporary English Pronunciation**. New York: Routledge, 2018. p. 511-526.
- KIM, J. Effects of Rubric-Referenced Self-Assessment Training on Korean High School Students' English Writing. **English Teaching**, v. 74, n. 3, p. 79-111, 2019.
- KENNEDY, E.; SHIEL, G. Writing assessment for communities of writers: Rubric validation to support formative assessment of writing in Pre-K to grade 2. **Assessment in Education: Principles, Policy & Practice**, v. 29, n. 2, p. 127-149, 2022.
- KNOCH, U. What Can Pronunciation Researchers Learn From Research into Second Language Writing?. *In*: ISAACS, T.; TROFIMOVICH, P. (Ed.) **Second language pronunciation assessment: Interdisciplinary perspectives**. Multilingual Matters, 2017. p. 54-67.
- KUIKEN, F.; VEDDER, I. Scoring approaches: scales/rubrics. *In*: WINKE, P.; BRUNFAUT, T. (Ed.). **The Routledge Handbook of Second Language Acquisition and Language Testing**. New York, NY: Routledge, 2021. p. 125-134.
- LAN, C.; FAN, S. Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. **Studies in Educational Evaluation**, v. 61, p. 112-122, 2019.
- LEUNG, C. Classroom-based assessment issues for language teacher education. **The Concise Companion to Language Assessment**. Malden, MA: John Wiley & Sons, 2013. p. 1510-1519.
- LEVIS, J. M. Changing contexts and shifting paradigms in pronunciation teaching. **TESOL quarterly**, v. 39, n. 3, p. 369-377, 2005.
- LEVIS, J. M. **Intelligibility, oral communication, and the teaching of pronunciation**. Cambridge University Press, 2018.
- LEVIS, J. M. Teaching pronunciation: Truths and lies. *In*: BARDEL, C. *et al.* (Eds.) **Exploring language education: Global and local perspectives**. Stockholm University Press, 2022. DOI: <https://doi.org/10.16993/bbz.c>.
- LUCKESI, C. C. **Avaliação da aprendizagem escolar: estudos e proposições**. 22^a ed. São Paulo: Cortez, 2011. p. 39-72.
- MACKEY, A.; GASS, S. M. (Ed.). **Second language research: Methodology and design**. NJ: Lawrence Erlbaum Associates, 2005.

MCNAMARA, T. F. **Measuring second language performance**. Addison Wesley Longman, 1996.

MARTINS, T. da. S.; SILVEIRA, R. Pronunciation Assessment: Relating the CEFR's Phonological Control Scale to Intelligibility and Comprehensibility Measures. **Revista de Estudos da Linguagem**, [S. l.], v. 33, n. 1, p. 195–219, 2025. DOI: 10.17851/2237-2083.33.1.195-219. Available at: <https://periodicos.ufmg.br/index.php/relin/article/view/59603>. Access on: 6 Feb. 2026.

MCTIGHE, J.; BROOKHART, S.; GUSKEY, T. The value of descriptive, multi-level rubrics. **Educational Leadership**, v. 81, n. 6, p. 1-8, 2024.

MENDOZA, L. E. Using Rubrics for Language Assessment. In: GOLSAVES, C.; PEARSON, J. (Ed.). **Improving Learning Through Assessment Rubrics: Student Awareness of What and How They Learn**, 2023. p. 1-15.

MUNRO, M. J.; DERWING, T. M. Intelligibility in research and practice: Teaching priorities. In: _____. **The Handbook of English Pronunciation**, 2015. p. 375-396.

MUNRO, M. J.; DERWING, T. M. Collecting data in L2 pronunciation research. In: KANG, O.; STAPLES, S.; YAW, K.; HIRSCHI, K. (org.). **Proceedings of the 11th Pronunciation in Second Language Learning and Teaching conference**. Northern Arizona University, September 2019. Ames, IA: Iowa State University, 2020. p. 8-18.

MORAES, I. T.; FURTOSO, V. B.; GIRALDO, F. Developing Pre-Service Teachers' Language Assessment Literacy: An Analysis of Classroom Tasks. **Ilha do Desterro**, v. 77, p. 1-20, 2024.

NAGLE, C.; SACHS, R.; ZÁRATE-SÁNDEZ, G. Spanish teachers' beliefs on the usefulness of pronunciation knowledge, skills, and activities and their confidence in implementing them. **Language Teaching Research**, v. 27, n. 3, p. 491-517, 2023.

PANADERO, E.; ANDRADE, H.; BROOKHART, S. Fusing self-regulated learning and formative assessment: A roadmap of where we are, how we got here, and where we are going. **The Australian Educational Researcher**, v. 45, n. 1, p. 13-31, 2018.

PANADERO, E.; JONSSON, A. The use of scoring rubrics for formative assessment purposes revisited: A review. **Educational research review**, v. 9, p. 129-144, 2013.

QUEVEDO-CAMARGO, G. Formação de professores de línguas adicionais e letramento em avaliação: breve panorama e desafios para os cursos de licenciatura em LEM no Brasil. **Calidoscópio**, v. 18, n. 2, p. 435-459, 2020. DOI: <https://doi.org/10.4013/cld.2020.182.10>. Access on 20 Set. 2024.

QUEVEDO-CAMARGO, G.; GARCIA, L. H. Avaliação da oralidade em língua inglesa: a atribuição de notas como expressão da subjetividade do professor. **Signum: Estudos da Linguagem**, v. 20, n. 3, p. 93-117, 2017.

QUEVEDO-CAMARGO, G.; SCARAMUCCI, M. R. V. O conceito de letramento em avaliação de línguas: origem de relevância para o contexto brasileiro. **Linguagem: estudos e pesquisas**, v. 22, n. 1, p. 221-245, 2018.

- QUEVEDO-CAMARGO, G.; SOUSA, A. A. C. de P. Letramento em avaliação para professores de inglês e a perspectiva glocal: um microestudo. **Revista do GEL**, v. 19, n. 1, p. 223-245, 2022.
- ROEVER, C.; MCNAMARA, T. Language testing: The social dimension. **International Journal of Applied Linguistics**, v. 16, n. 2, p. 242-258, 2006.
- SAITO, K.; PLONSKY, L. Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. **Language Learning**, v. 69, n. 3, p. 652-708, 2019.
- SAITO, K. Effects of corrective feedback on second language pronunciation development. *In*: NASSAJI, H.; KARTCHAVA, E. (Eds.). **The Cambridge Handbook of Corrective Feedback in Second Language Learning and Teaching**. Cambridge University Press: Cambridge, UK. p. 407-428.
- SCARAMUCCI, M. V. R. Avaliação: mecanismo propulsor de mudanças no ensino/aprendizagem de língua estrangeira. **Revista Contexturas**, v. 4, p. 75-81, 1999.
- SCARAMUCCI, M. V. R. Letramento em avaliação (em contexto de línguas): contribuições para a linguística aplicada, educação e sociedade. *In*: JORDÃO, C. M. (Org.). **A linguística aplicada no Brasil: rumos e passagens**. Campinas, SP: Pontes Editores, 2016. p. 141-165.
- SCARAMUCCI, M. V. R. O professor avaliador: sobre a importância da avaliação na formação do professor de língua estrangeira. *In*: ROTTAVA, L.; SANTOS, S. (Orgs.). **Ensino-aprendizagem de línguas: língua estrangeira**. Ijuí: Editora da UNIJUÍ, 2006. p. 49-64.
- SCARINO, A. Developing assessment literacy of teachers of languages: A conceptual and interpretive challenge. **Papers in Language Testing and assessment**, v. 6, n. 1, p. 18-40, 2017.
- SILVEIRA, R.; MARTINS, T. da S. Assessing second language oral proficiency development with holistic and analytic scales. **Ilha do Desterro**, v. 73, p. 227-250, 2020.
- SPADA, N. Communicative Language Teaching: Current Status and Future Prospects. CUMMINS, J.; DAVISON, C. (Ed.). *In*: **International handbook of English language teaching**. Boston, MA: Springer US, 2007. p. 271-288.
- STIGGINS, R. Assessment literacy. **Phi Delta Kappan**, v. 72, n. 7, p. 534-539, 1991.
- SUZUKIDA, Y.; SAITO, K. What is second language pronunciation proficiency? An empirical study. *System*, v. 106, p. 102754, 2022.
- TAYLOR, L. Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. **Language Testing**, v. 30, n. 3, p. 403-412, 2013. DOI: <https://doi.org/10.1177/0265532213480338>
- TURNER, C. The need for impact studies of L2 performance testing and rating: Identifying areas of potential consequences at all levels of the testing cycle. *In*: ELDER, A. *et al.* **Experimenting with uncertainty: Essays in honour of Alan Davies**. Cambridge University Press, 2001. p. 138-149.

TURNER, C. E.; PURPURA, J. E. Learning-oriented assessment second and foreign classrooms. *In*: TSAGARI, D.; BANERJEE, J. (Ed.). **Handbook of second language assessment**. De Gruyter, 2016. DOI: <https://doi.org/10.1515/9781614513827>.

THOMSON, R. I. Fluency. *In*: REED, M.; LEVIS, J. M. (Ed.). **The Handbook of English Pronunciation**. John Wiley & Sons, 2015.

VOGT, K.; TSAGARI, D.; SPANOUDIS, G. What do teachers think they want? A comparative study of in-service language teachers' beliefs on LAL training needs. **Language Assessment Quarterly**, v. 17, n. 4, p. 386-409, 2020.

WEINBERGER, S. **Speech Accent Archive**. Washington, D.C.: George Mason University, [s.d.]. Available at: <https://accent.gmu.edu/>. Access on: 21 Dec. 2025.

WILLIAM, D. Assessment: The bridge between teaching and learning. **Voices from the Middle**, v. 21, n. 2, p. 15-20, 2013.

ZHONG, W. Pronunciation rating scale in second language pronunciation assessment: A Review. **Journal of Language Teaching and Research**, v. 10, n. 1, p. 141, 2019.

APPENDIX A - INFORMED CONSENT FORM



UNIVERSIDADE FEDERAL DE SANTA CATARINA CENTRO DE COMUNICAÇÃO E EXPRESSÃO PROGRAMA DE PÓS-GRADUAÇÃO EM INGLÊS: ESTUDOS LINGÜÍSTICOS E LITERÁRIOS

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Prezado(a) participante,

Eu me chamo Manuella Hinchinger Pereira e sou mestranda no Programa de Pós-Graduação em Inglês (PPGI) da Universidade Federal de Santa Catarina (UFSC) sob orientação da Profa. Dra. Rosane Silveira, onde investigamos a área de ensino da pronúncia em segunda língua.

Estou convidando você para participar da minha pesquisa intitulada: “Desenvolvendo uma rubrica analítica preliminar para avaliar a pronúncia em segunda língua na sala de aula de inglês: uma investigação das preferências e necessidades de professores brasileiros em serviço”.

Por que estou realizando essa pesquisa?

Na sala de aula de inglês, a avaliação da pronúncia dos estudantes muitas vezes é negligenciada. Avaliar a pronúncia é importante porque permite aos professores identificar quais aspectos da fala já foram plenamente adquiridos e quais precisam de mais atenção no currículo, prevendo dificuldades que os estudantes possam enfrentar durante momentos de comunicação oral. O processo de avaliação da pronúncia, por ser detalhado e complexo, exige tempo, conhecimento específico e o uso de critérios bem definidos, além de instrumentos de avaliação confiáveis. Esses desafios podem explicar porque a avaliação da pronúncia nem sempre é feita de maneira eficaz e justa. Por isso, o objetivo desta pesquisa é criar um instrumento avaliativo, chamado rubrica, que auxilie os professores brasileiros de inglês a realizarem a avaliação formativa da pronúncia de seus estudantes, levando em consideração suas necessidades e preferências, de modo a tornar a rubrica desenvolvida mais eficiente, adequada e útil para o contexto educacional do Brasil.

Quem pode participar desta pesquisa?

O público-alvo desta pesquisa são professores brasileiros de língua inglesa **em serviço com graduação (licenciatura/bacharel) concluída em Letras - Inglês ou Letras - Inglês/Português.**

O que vai acontecer?

Se você aceitar participar da pesquisa, será solicitado a responder dois questionários no Google Forms e estudar a rubrica de avaliação da pronúncia desenvolvida nesta pesquisa, bem como o seu manual de uso. A participação em todas as etapas da pesquisa requer em torno de **2 horas e 15 minutos**. Detalhes sobre cada instrumento que será utilizado na coleta de dados e o que acontecerá em cada etapa da pesquisa estão descritos a seguir:

1. Questionário de experiência profissional e acadêmica:

Neste primeiro questionário, você responderá a perguntas sobre aspectos pessoais (como sua idade e escolaridade) e profissionais (como sua formação acadêmica, experiência com o ensino de língua inglesa e contexto pedagógico, bem como treinamento em avaliação de línguas). Além disso, você responderá a perguntas sobre suas práticas de avaliação em sala de aula com relação à pronúncia dos seus estudantes. O tempo estimado para completar este questionário é de 25 a 35 minutos.

2. Estudo da rubrica e do manual de uso:

Após completar o primeiro questionário, você receberá, através do Moodle Grupos, à rubrica desenvolvida pela pesquisadora e a um manual de uso da mesma. Esse manual esclarece as principais dúvidas que podem surgir ao utilizar a rubrica e ao avaliar a pronúncia de estudantes brasileiros de inglês. É importante que você estude este material antes de responder ao questionário dois. O tempo estimado para leitura e compreensão da rubrica e do manual é de 40 a 50 minutos.

3. Questionário de avaliação da rubrica:

O segundo questionário será uma semana após você completar o estudo da rubrica e leitura do manual de uso. Neste momento, você ouvirá três amostras de falas e deverá avaliá-las utilizando os critérios presentes na rubrica de avaliação da pronúncia. Após isso, você irá para a próxima seção, onde responderá perguntas sobre o instrumento avaliativo que acabou de utilizar com o intuito de dar sugestões sobre como melhorá-la. O tempo estimado para completar este questionário é de 30 a 40 minutos.

Todas as etapas serão realizadas inteiramente de forma **online/remota**. Você receberá o link para participar de cada etapa por e-mail ou WhatsApp com instruções da pesquisadora sobre como proceder.

Caso você não se sinta confortável ou prefira não responder a qualquer pergunta de qualquer um dos questionários, poderá deixá-las em branco sem nenhum prejuízo para você.

Haverá algum risco na realização das atividades dessa pesquisa?

Ainda que as atividades solicitadas de você para este estudo sejam parecidas com as atividades que você provavelmente vivencia em sua profissão como professor(a), toda pesquisa pode apresentar algum risco. Por isso, você pode se sentir ansioso ao ter que responder perguntas de cunho pessoal ou realizar avaliações da pronúncia de um falante. Além disso, você pode sentir cansaço físico ao preencher os questionários em ambiente virtual e experienciar irritabilidade durante possíveis falhas advindas do uso de equipamentos tecnológicos. Para mediar esses possíveis desconfortos e oferecer as melhores condições para a sua participação na pesquisa, estarei à disposição durante todo o processo, respondendo suas dúvidas com o máximo de clareza e auxiliando você a resolver qualquer problema relacionado ao uso de questionários e videoconferência online.

Haverá algum benefício?

Ao participar deste estudo, você estará contribuindo para o desenvolvimento da ciência no Brasil em duas instâncias, sendo elas i. a realização da pesquisa poderá ajudar a expandir a área de avaliação da pronúncia e letramento em avaliação de línguas e ii. os resultados poderão ser úteis para o tornar os processos avaliativos da fala em segunda língua, realizados em salas de aula regulares, mais justos. Além disso, como professor(a), você terá a oportunidade de aprofundar suas reflexões sobre suas práticas e crenças em relação à avaliação da pronúncia dos seus estudantes, tornando-se mais consciente e crítico a respeito das mesmas, o que pode impactar positivamente seu desenvolvimento profissional docente no que diz respeito a importância da avaliação no processo de ensino-aprendizagem.

A minha identidade será revelada?

Não, os dados coletados serão mantidos em sigilo, ou seja, seu nome não será mencionado na pesquisa. Os participantes serão identificados apenas por números. Por exemplo, participante 1. Apesar de toda pesquisa envolver um risco de quebra de sigilo, estou comprometida em adotar todas as medidas necessárias para proteger o anonimato dos seus dados. Além disso, irei armazenar suas informações em contas seguras, protegidas por senhas fortes. Ao fim da coleta de dados, todos os dados coletados serão transferidos para dispositivos eletrônicos locais e excluídos de plataformas virtuais, ambientes compartilhados e da nuvem.

Haverá acompanhamento de alguém?

Toda a coleta de dados acontecerá de forma remota/online. Assim, você poderá responder aos questionários no horário e local de sua preferência. No entanto, estou comprometida em oferecer a assistência necessária em todas as etapas de realização da pesquisa. Você poderá solicitar uma videoconferência individual em um horário de sua escolha para esclarecer suas dúvidas com a pesquisadora e receber ajuda com qualquer questão que precise.

Haverá alguma despesa?

Como a pesquisa será realizada de forma remota/online, não estão previstas despesas com aquisição de equipamentos, deslocamento ou alimentação. Além disso, como prevê a legislação brasileira, você não receberá nenhuma compensação financeira pela sua participação. No entanto, caso a pesquisa lhe cause danos materiais ou gastos que sejam devidamente comprovados, você tem direito a ressarcimento ou indenização oriundos da sua participação.

É possível desistir de participar da pesquisa ou cancelar essa autorização?

Sim. É possível encerrar ou cancelar a sua participação na pesquisa a qualquer momento e isso não trará nenhum prejuízo para você. Se você desistir da pesquisa, os seus dados serão permanentemente excluídos e não serão usados nos resultados da pesquisa. Para encerrar ou cancelar a sua participação, envie uma mensagem para o meu e-mail: manuellaheinchingr@hotmail.com. Você não precisa explicar porque está solicitando o encerramento ou cancelamento da sua participação.

Como faço o contato para esclarecer dúvidas?

Durante a pesquisa, você receberá acompanhamento constante. Porém, caso haja alguma dúvida ou problema, abaixo estão o meu contato e endereço, bem como o da minha orientadora:

Pesquisadora: Manuella Hinchinger Pereira

WhatsApp: (número de telefone)

E-mail: manuella_hinchinger@hotmail.com

Orientadora: Profa. Dra. Rosane Silveira

E-mail: rosanasilveira@hotmail.com

Endereço do Programa de Pós-Graduação em Inglês (PPGI): Centro de Comunicação e Expressão (CCE) “B”, sala 313, Campus Universitário, Trindade, Florianópolis, CEP 88.040-900.

Caso você queira entrar em contato com o Comitê de Ética em Pesquisas com Seres Humanos da UFSC, use uma dessas formas de contato:

Presencialmente no Prédio Reitoria II, 7º andar, sala 701, localizado na Rua Desembargador Vitor Lima, n.º 222, Trindade, Florianópolis, CEP 88.040-400.

Telefone para contato: 3721-6094

E-mail: cep.propesq@contato.ufsc.br

O Comitê de Ética em Pesquisas com Seres Humanos da UFSC (art. 17º., inc. IX da res. 510/16) é o órgão responsável por garantir a sua segurança como participante. Ele é “interdisciplinar, deliberativo, consultivo e educativo, vinculado à Universidade Federal de Santa Catarina, mas independente na tomada de decisões, criado para defender os interesses dos participantes da pesquisa em sua integridade e dignidade e para contribuir no desenvolvimento da pesquisa dentro de padrões éticos”.

A pesquisadora responsável, compromete-se a conduzir a pesquisa segundo o que preconiza as Resoluções 466/12 e 510/16. Essas resoluções dispõem sobre pesquisas envolvendo seres humanos no Brasil na área de Ciências Humanas e Sociais. Além disso, durante o estudo, serão cumpridos os requisitos da Lei Geral de Proteção de Dados (Lei Nº 13.709, de 14 de agosto de 2018) no que diz respeito ao tratamento de dados pessoais e dados pessoais sensíveis necessários para a realização da pesquisa. Por fim, visto que este estudo envolve etapas remotas/virtuais, seus procedimentos também seguirão as orientações estabelecidas no Ofício Circular Nº 2/2021/CONEP/SECNS/MS.

Ao concordar com este Termo de Consentimento Livre e Esclarecido, você está aceitando participar da pesquisa. Muito obrigada pela sua atenção e autorização!

☐ Eu li e concordo em participar da pesquisa.

☐ Não concordo em participar da pesquisa.

Assinatura da pesquisadora responsável:

Florianópolis (SC)

_____ de _____ de 2025

APPENDIX B - BACKGROUND QUESTIONNAIRE

Seção I

1. Após ler o Termo de Consentimento Livre e Esclarecido, você concorda em participar da pesquisa?

☐ Eu li e concordo em participar da pesquisa. Desejo prosseguir com a coleta de dados.

2. Qual o seu nome?

3. Indique seu e-mail e/ou WhatsApp para a pesquisadora falar com você

Seção II

Formação acadêmico-profissional

Gostaríamos de conhecer um pouco mais sobre você e sua trajetória acadêmico-profissional. Por favor, responda às perguntas a seguir considerando suas experiências prévias de formação e atuação docente.

1. Quantos anos você tem?

2. Qual a sua formação? *Marcar apenas uma oval.*

☐ Licenciatura em Letras - Inglês

☐ Bacharel em Letras - Inglês

☐ Licenciatura em Letras - Inglês/Português

☐ Bacharel em Letras - Inglês/Português

Outro: _____

3. Qual é o nível mais alto de escolaridade que você possui na grande área de Letras além da graduação? *Marcar apenas uma oval.*

☐ Especialização

☐ Mestrado

☐ Doutorado

☐ Pós-doutorado

Outro: _____

4. Há quantos anos você atua como professor de inglês? *Marcar apenas uma oval.*

☐ Menos de 1 ano

☐ De 1 a 5 anos

☐ De 6 a 10 anos

☐ Mais de 10 anos

☐ Mais de 15 anos

Outro: _____

5. Em que tipo de instituição de ensino você leciona atualmente? *Marque todas que se aplicam.*

☐ Escola pública (ensino fundamental ou médio)

☐ Escola particular (ensino fundamental ou médio)

☐ Instituto federal

☐ Curso de idiomas

☐ Aulas particulares

Outro: _____

6. Qual é a faixa etária dos seus estudantes? *Marque todas que se aplicam.*

☐ Crianças (até 10 anos)

☐ Pré-adolescentes (11 a 13 anos)

☐ Adolescentes (14 a 17 anos)

☐ Jovens adultos (18 a 25 anos)

☐ Adultos (26 a 59 anos)

☐ Idosos (60 anos ou mais)

7. Geralmente, os seus estudantes pertencem à qual nível de proficiência? *Marque todas que se aplicam.*

☐ Iniciante (A1-A2)

☐ Intermediário (B1-B2)

☐ Avançado (C1-C2)

Outro: _____

8. Depois da graduação, você já participou de algum curso ou programa focado em fonética e fonologia/ensino da pronúncia? *Marcar apenas uma oval.*

☐ Sim

☐ Não

Outro: _____

APPENDIX C - ASSESSMENT SURVEY

Seção I

Formação na área de avaliação de línguas

Durante a sua formação docente, seja na graduação ou depois dela, você teve algum contato com teorias e/ou práticas de avaliação em contexto de ensino de segunda língua? Por favor, indique o nível de estudo que você teve em sua formação docente nas áreas a seguir.

1. Avaliação voltada para a sala de aula em contextos de segunda língua. *Marcar apenas uma oval por linha.*

Avaliação de línguas voltada para a sala de aula	Nenhum contato	Breve contato	Treinamento avançado
Desenvolver atividades avaliativas e provas para a sala de aula.			
Utilizar atividades avaliativas e provas prontas de livros didáticos.			
Fornecer feedback aos estudantes a partir dos resultados de atividades avaliativas e provas.			
Incorporar autoavaliações e avaliações entre estudantes no processo de ensino.			
Realizar avaliação informais (de maneira espontânea, sem provas ou notas) e contínuas (visando acompanhar o progresso dos estudantes).			
Criar e utilizar rubricas para avaliar os estudantes.			

2. Finalidade da avaliação e de utilizar provas em contextos de segunda língua. *Marcar apenas uma oval por linha.*

Finalidade da avaliação e de utilizar provas	Nenhum contato	Breve contato	Treinamento avançado
Atribuir notas.			
Diagnosticar pontos que necessitam de mais atenção no ensino/aprendizagem.			
Nivelar os estudantes por turmas, níveis de proficiência, etc.			

3. Conteúdos e conceitos da avaliação em contextos de segunda língua

Conteúdos e conceitos da avaliação em segunda língua	Nenhum contato	Breve contato	Treinamento avançado
Avaliar a habilidade de produção escrita.			
Avaliar a habilidade de produção oral.			
Avaliar a habilidade de compreensão escrita.			
Avaliar a habilidade de compreensão oral.			
Avaliar a pronúncia.			
Avaliar a gramática/vocabulário.			
Determinar se a avaliação gera resultados parecidos quando é realizada mais de uma vez.			
Determinar se o que está sendo avaliado corresponde aos objetivos da avaliação e se os resultados fazem sentido para orientar o feedback e as decisões pedagógicas.			
Usar métodos estatísticos para estudar a qualidade de uma prova/atividade avaliativa.			

4. Se desejar, aproveite este espaço para descrever com mais detalhes sua experiência caso tenha selecionado "breve contato" ou "treinamento avançado" em algum dos itens acima.

Seção II

Práticas de sala de aula de avaliação da pronúncia

Enquanto professor, avaliar a pronúncia dos seus estudantes é algo que faz parte da sua realidade? Por favor, responda às perguntas abaixo levando em consideração suas práticas em sala de aula.

1. Você avalia a pronúncia dos seus estudantes? * *Marcar apenas uma oval.*
☐ Sim
☐ Não
2. Você costuma avaliar a pronúncia de quais estudantes? *Marque todas que se aplicam.*
☐ Iniciantes (A1-A2)
☐ Intermediários (B1-B2)
☐ Avançados (C1-C2)

☐ Apenas daqueles que solicitam uma avaliação mais detalhada de sua pronúncia

Outro: _____

3. Para avaliar a pronúncia dos seus estudantes, quais critérios você utiliza? *Marque todas que se aplicam.*

☐ Utilizo critérios definidos previamente por mim

☐ Utilizo critérios definidos previamente por mim em conjunto com outros professores da escola

☐ Utilizo critérios fornecidos pelo livro didático e/ou pela escola

☐ Não utilizo critérios definidos previamente, apenas sigo minha intuição no momento da avaliação

Outro: _____

4. Você utiliza algum instrumento para auxiliá-lo na avaliação da pronúncia dos seus estudantes? *Marque todas que se aplicam.*

☐ Não utilizo nenhum instrumento

☐ Rubricas (com critérios e níveis de desempenho)

☐ Checklists (listas para indicar a presença ou ausência de elementos na pronúncia do estudante)

☐ Escalas de proficiência (por exemplo, o CEFR)

Outro: _____

5. Caso utilize algum instrumento para avaliar a pronúncia, você costuma compartilhá-lo ou discuti-lo com os estudantes antes de sua aplicação? *Marque apenas uma oval.*

☐ Sim, sempre.

☐ Sim, às vezes.

☐ Raramente.

☐ Nunca.

Outro: _____

6. Como a pronúncia é considerada na avaliação dos seus estudantes? *Marque todas que se aplicam.*

☐ A pronúncia tem o mesmo peso que outras habilidades (como leitura, escrita, audição e fala) nas avaliações

☐ A pronúncia é um critério importante e representa uma parte significativa da nota de produção oral

☐ A pronúncia é um dos critérios considerados na avaliação da produção oral, mas não tem tanto peso

☐ A pronúncia é avaliada, mas não influencia a nota do estudante

Outro: _____

7. Qual o seu objetivo ao avaliar a pronúncia dos seus estudantes? *Marque todas que se aplicam.*

☐ Aproximar ao máximo a pronúncia dos estudantes à de falantes nativos da língua inglesa

☐ Ajudar os estudantes a produzirem sons difíceis com clareza, priorizando a inteligibilidade

Outro: _____

8. Como você costuma fornecer feedback sobre a pronúncia dos seus estudantes? *Marque todas que se aplicam.*

☐ De maneira planejada (por exemplo, após provas ou atividades avaliativas)

- ☐ De maneira espontânea (por exemplo, durante interações em sala de aula)
- ☐ Apenas em momentos específicos a pedido do estudante
- ☐ Raramente forneço feedback sobre pronúncia
- Outro: _____

9. Utilizando a escala abaixo, por favor indique o quão confiante você se sente avaliando cada um dos elementos da pronúncia, sendo 1 = nada confiante e 10 = muito confiante.

- a. Vogais.
(Nada confiante) 1 – 10 (Muito confiante)
- b. Consoantes.
(Nada confiante) 1 – 10 (Muito confiante)
- c. Ritmo (intervalo entre sons fortes e fracos).
(Nada confiante) 1 – 10 (Muito confiante)
- d. Entoação (indica o propósito comunicativo, por exemplo, pergunta, afirmação, etc.).
(Nada confiante) 1 – 10 (Muito confiante)
- e. Acento tônico (ênfase dada a sílaba de uma palavra, por exemplo, PHOtograph).
(Nada confiante) 1 – 10 (Muito confiante)
- f. Acento frasal (ênfase dada as palavras de uma frase, por exemplo, I WANT to GO to the MARket).
(Nada confiante) 1 – 10 (Muito confiante)
- g. Fluência (uso de pausas e número de palavras ditas por minuto).
(Nada confiante) 1 – 10 (Muito confiante)
- h. Compreensibilidade (nível de esforço necessário para entender o que está sendo dito).
(Nada confiante) 1 – 10 (Muito confiante)

10. Em que medida as informações obtidas por meio da avaliação da pronúncia dos seus estudantes contribuem para o planejamento das suas aulas? *Marque todas que se aplicam.*

- ☐ Costumo sempre usar os resultados da avaliação para adaptar minhas aulas às necessidades de pronúncia observadas
- ☐ Às vezes ajusto o ensino com base nas dificuldades de pronúncia observadas
- ☐ Raramente mudo meu planejamento com base nas dificuldades de pronúncia dos estudantes
- ☐ A avaliação da pronúncia não influencia meu planejamento de ensino
- Outro: _____

11. Quais dificuldades você encontra ao avaliar a pronúncia dos seus estudantes em sala de aula? *Marque todas que se aplicam.*

- ☐ Dificuldade em encontrar tempo durante o semestre
- ☐ Pouco conhecimento teórico e preparo pedagógico sobre avaliação de línguas e fonética e fonologia
- ☐ Falta de confiança na minha própria habilidade enquanto professor-avaliador
- ☐ Medo de constranger o estudante
- ☐ Falta de instrumentos e materiais específicos para avaliação da pronúncia no contexto brasileiro
- ☐ Número elevado de estudantes/turmas
- ☐ Não enfrento dificuldades
- Outro: _____

12. Se desejar, aproveite este espaço para descrever com mais detalhes suas práticas de avaliação da pronúncia ou para comentar alguma resposta dada anteriormente nesta seção.

Obrigada! Em breve a pesquisadora entrará em contato com você para enviar uma cópia do seu Termo de Consentimento Livre e Esclarecido da sua participação no estudo e para explicar a próxima etapa da pesquisa.

APPENDIX D - RUBRIC SURVEY

Olá!

Este questionário visa coletar informações sobre a rubrica de avaliação da pronúncia desenvolvida nesta pesquisa, com o intuito de aprimorá-la para as necessidades e preferências de professores brasileiros. Por isso, pedimos que você avalie a pronúncia de três estudantes falando em inglês com a rubrica e, em seguida, responda a algumas perguntas sobre sua experiência utilizando este instrumento.

O questionário está dividido em duas seções:

1. Avaliação da pronúncia de estudantes de inglês;
2. Avaliação da rubrica e do manual.

A sua participação é muito importante para esta pesquisa. Em caso de dúvidas, sinta-se à vontade para entrar em contato com a pesquisadora.

Muito obrigada!

1. Qual o seu nome?

Seção I

Avaliação da pronúncia de estudantes de inglês

Por favor, ouça atentamente aos áudios a seguir quantas vezes precisar e as avalie utilizando a rubrica. Para consultar a rubrica novamente, clique aqui.

1. Ouça a falante 1 de nome fictício Julia (link para ouvir o áudio).
2. Avalie a falante 1.

Avaliação da falante 1 *

Marcar apenas uma oval por linha.

	Avançado	Progredindo	Começando	Básico	Não demonstrado
Vogais	☐	☐	☐	☐	☐
Consoantes	☐	☐	☐	☐	☐
Prosódia (ritmo, entoação e acentuação)	☐	☐	☐	☐	☐
Fluência e habilidade de auto-correção	☐	☐	☐	☐	☐
Formação silábica	☐	☐	☐	☐	☐
Compreensibilidade	☐	☐	☐	☐	☐

3. Ouça ao falante 2 de nome fictício Guilherme (link para ouvir o áudio).

4. Avalie o falante 2.

Avaliação do falante 2 *

Marcar apenas uma oval por linha.

	Avançado	Progredindo	Começando	Básico	Não demonstrado
Vogais	☐	☐	☐	☐	☐
Consoantes	☐	☐	☐	☐	☐
Prosódia (ritmo, entoação e acentuação)	☐	☐	☐	☐	☐
Fluência e habilidade de auto-correção	☐	☐	☐	☐	☐
Formação silábica	☐	☐	☐	☐	☐
Compreensibilidade	☐	☐	☐	☐	☐

5. Ouça a falante 3 de nome fictício Clara (link para ouvir o áudio).

6. Avalie a falante 3.

Avaliação da falante 3 *

Marcar apenas uma oval por linha.

	Avançado	Progredindo	Começando	Básico	Não demonstrado
Vogais	☐	☐	☐	☐	☐
Consoantes	☐	☐	☐	☐	☐
Prosódia (ritmo, entoação e acentuação)	☐	☐	☐	☐	☐
Fluência e habilidade de auto-correção	☐	☐	☐	☐	☐
Formação silábica	☐	☐	☐	☐	☐
Compreensibilidade	☐	☐	☐	☐	☐

7. Utilizando a escala abaixo, por favor indique o quão confiante você se sentiu avaliando cada um dos elementos da pronúncia em L2 com a rubrica, sendo 1 = nada confiante e 10 = muito confiante.

a. Vogais.

(Nada confiante) 1 – 10 (Muito confiante)

b. Consoantes.

(Nada confiante) 1 – 10 (Muito confiante)

c. Ritmo (intervalo entre sons fortes e fracos).

(Nada confiante) 1 – 10 (Muito confiante)

d. Entoação (indica o propósito comunicativo, por exemplo, pergunta, afirmação, etc.).

(Nada confiante) 1 – 10 (Muito confiante)

e. Acento tônico (ênfase dada a sílaba de uma palavra, por exemplo, PHOtograph).

(Nada confiante) 1 – 10 (Muito confiante)

f. Acento frasal (ênfase dada as palavras de uma frase, por exemplo, I WANT to GO to the MARket).

(Nada confiante) 1 – 10 (Muito confiante)

g. Fluência (uso de pausas e número de palavras ditas por minuto).

(Nada confiante) 1 – 10 (Muito confiante)

h. Compreensibilidade (nível de esforço necessário para entender o que está sendo dito).

(Nada confiante) 1 – 10 (Muito confiante)

Seção II

Avaliação da rubrica

Agora, você irá avaliar a rubrica que acabou de utilizar em três categorias: critérios presentes na rubrica, níveis de desempenho presentes na rubrica e a rubrica em sua totalidade, além do manual que a acompanha. Por favor, responda às perguntas a seguir, levando em consideração sua própria experiência e opinião enquanto professor(a).

Critérios presentes na rubrica: Os critérios da rubrica desenvolvida na pesquisa incluem vogais, consoantes, prosódia (ritmo, entoação e acentuação), fluência e habilidade de auto-correção, formação silábica e compreensibilidade.

1. Os critérios são relevantes para avaliar se a pronúncia do estudante é inteligível.
(Discorto totalmente) 1 – 10 (Concordo totalmente)
2. Os critérios estão de acordo com orientações educacionais importantes presentes em diretrizes curriculares e quadros de proficiência linguística.
(Discorto totalmente) 1 – 10 (Concordo totalmente)
3. Cada critério está claro e fácil de entender.
(Discorto totalmente) 1 – 10 (Concordo totalmente)
4. Os critérios podem ser explicados de forma simples para outros professores e estudantes.
(Discorto totalmente) 1 – 10 (Concordo totalmente)
5. O que cada critério descreve pode ser percebido nas falas avaliadas.
(Discorto totalmente) 1 – 10 (Concordo totalmente)
6. Cada critério avalia um aspecto diferente da pronúncia.
(Discorto totalmente) 1 – 10 (Concordo totalmente)
7. Os critérios, juntos, incluem os principais aspectos de uma pronúncia inteligível em segunda língua.
(Discorto totalmente) 1 – 10 (Concordo totalmente)
8. Você incluiria ou excluiria algum critério da rubrica? Se sim, qual e por quê?

Níveis de desempenho presentes na rubrica: Os níveis de desempenho da rubrica incluem os níveis avançado, progredindo, começando, básico e não demonstrado, bem como as descrições presentes em cada um.

9. As descrições da rubrica são detalhadas e precisas, além de fundamentados em evidências observáveis, assim professores não precisam depender de seu julgamento pessoal.
(Discorto totalmente) 1 – 10 (Concordo totalmente)
10. As descrições são claras, de modo que tanto professores quanto estudantes podem facilmente entender o que se espera de cada nível de pronúncia.
(Discorto totalmente) 1 – 10 (Concordo totalmente)

11. A rubrica cobre desde os níveis mais baixos até os mais altos de pronúncia, descrevendo o desempenho de estudantes em diferentes momentos de desenvolvimento.

(Discorto totalmente) 1 – 10 (Concordo totalmente)

12. Cada nível de desempenho da pronúncia é bem definido e diferente entre si, de forma que o professor pode indicar de maneira inequívoca o nível de desenvolvimento um estudante.

(Discorto totalmente) 1 – 10 (Concordo totalmente)

13. A descrição da pronúncia esperada para o estudante está posicionada em um nível correspondente na rubrica. Assim, o nível que o estudante é avaliado reflete exatamente a habilidade dele(a).

(Discorto totalmente) 1 – 10 (Concordo totalmente)

14. As descrições são adequadas ao tempo que professores geralmente têm para avaliar a pronúncia de seus estudantes, o que facilita o seu uso regular da rubrica em sala de aula.

(Discorto totalmente) 1 – 10 (Concordo totalmente)

15. As descrições são adequadas para as necessidades de professores que não são especialistas em fonética e fonologia e orientam mesmo aqueles sem tanta experiência.

(Discorto totalmente) 1 – 10 (Concordo totalmente)

16. Você considera necessário adicionar ou excluir algum nível de desempenho da rubrica? Se sim, qual nível e por qual motivo?

17. Alguma palavra ou expressão nas descrições dos níveis da rubrica te deixou em dúvida? Se sim, como você acha que ela poderia ser escrita para ficar mais clara?

A rubrica:

18. O uso da rubrica facilitou ou dificultou o processo de tomada de decisões ao avaliar a pronúncia de estudantes brasileiro de inglês? *Marcar apenas uma oval.*

- ☐ Facilitou muito
- ☐ Facilitou em alguns aspectos, mas dificultou em outros
- ☐ Não facilitou, nem dificultou
- ☐ Dificultou em alguns aspectos
- ☐ Dificultou muito
- ☐ Não sei

Outro: _____

19. Houve algum falante que você teve dificuldade de avaliar utilizando a rubrica? *Marque todas que se aplicam.*

- ☐ Sim, falante 1

- ☐ Sim, falante 2
- ☐ Sim, falante 3
- ☐ Não tive dificuldade em avaliar nenhum falante

Outro: _____

20. Caso tenha selecionado 'sim' na pergunta anterior, explique quais fatores exatamente te levaram a ter dificuldades para avaliar esse(s) falante(s).

21. Você utilizaria esta rubrica para avaliar a pronúncia dos seus estudantes em sala de aula? *Marcar apenas uma oval.*

- ☐ Sim, com certeza utilizaria
- ☐ Sim, mas apenas se tiver tempo e achar benéfico para meus estudantes
- ☐ Talvez, ainda estou em dúvida sobre como aplicá-la em sala de aula
- ☐ Provavelmente não utilizaria
- ☐ Não utilizaria

Outro: _____

22. Caso tenha selecionado 'talvez' ou 'não utilizaria' na pergunta anterior, explique quais fatores exatamente te levaram a considerar não utilizar a rubrica com seus estudantes.

23. Se você fosse utilizar a rubrica em sua sala de aula, para quais objetivos específicos ela seria mais útil? *Marque todas que se aplicam.*

- ☐ Atribuir nota a pronúncia dos estudantes de maneira mais justa
- ☐ Planejar ou ajustar as aulas considerando as dificuldades dos estudantes
- ☐ Identificar a habilidade dos estudantes para nivelá-los
- ☐ Acompanhar o progresso dos estudantes ao longo do tempo
- ☐ Compartilhar com os estudantes o que já alcançaram e o que precisam melhorar
- ☐ Comunicar resultados para famílias, equipe pedagógica, etc.
- ☐ Todas as opções.

Outro: _____

24. O que tornaria a rubrica mais eficaz e acessível para a avaliação da pronúncia considerando o contexto brasileiro de ensino e aprendizagem? *Marque todas que se aplicam.*

- ☐ Uma versão digital e interativa
- ☐ Uma versão menor que seja mais rápida de utilizar
- ☐ Uma versão adaptada para estudantes utilizarem em auto avaliações/avaliações entre colegas
- ☐ Uma versão com uma linguagem menos teórica

Outro: _____

25. Quais partes do manual disponibilizado para auxiliá-lo(a) com a rubrica foram mais difíceis de entender e/ou podem ser melhorados? *Marque todas que se aplicam.*

- ☐ A linguagem utilizada é teórica e difícil de entender
 - ☐ O tamanho do manual é muito extenso para estudo e/ou consulta rápida
 - ☐ O glossário poderia ser mais completo
 - ☐ Algumas explicações poderiam ser ilustradas com exemplos em formato de imagem e áudio
 - ☐ O design não facilita a leitura e compreensão
 - ☐ Algumas dúvidas que tenho não foram respondidas
 - ☐ As instruções são muito complexas
 - ☐ O manual está ótimo, não há nada que possa ser melhorado
- Outro: _____

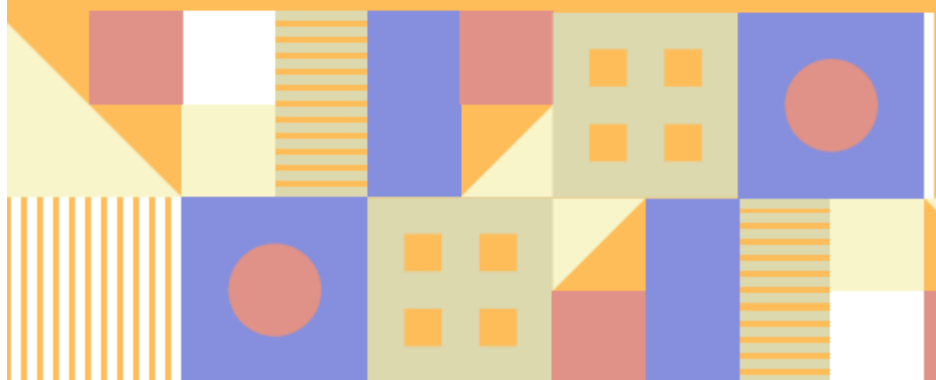
26. Caso queira fazer algum comentário ou sugestão adicional, utilize este espaço.

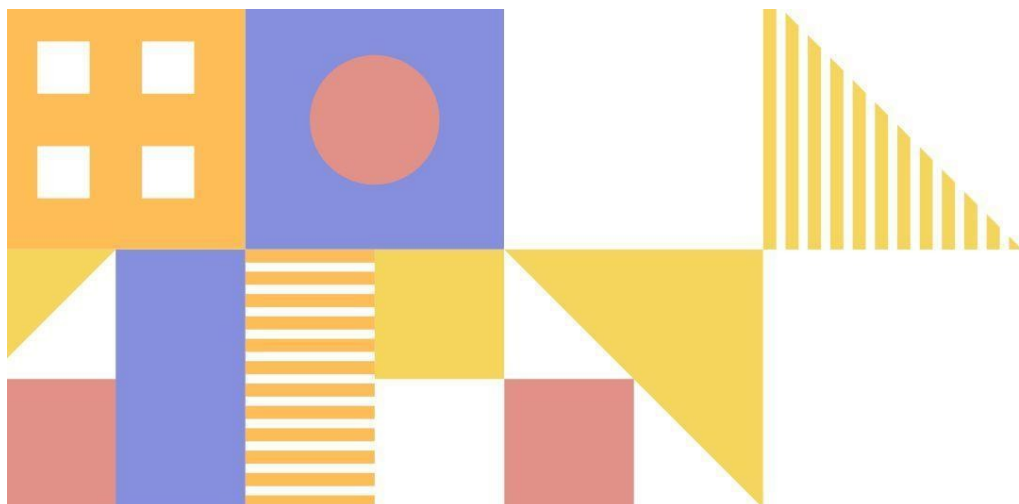
APPENDIX E - SUPPORTING GUIDE

MANUAL

Rubrica de avaliação
da pronúncia em
Segunda Língua para
estudantes brasileiros

Mestranda:
Manuella Hinchings Pereira
Orientadora:
Profa. Dra. Rosane Silveira





Pesquisa de Mestrado

Mestranda:
Manuella Hinchinger Pereira
Orientadora:
Profa. Dra. Rosane Silveira

Certificado de Apresentação
de Apreciação Ética:
87404725.3.0000.0121



Introdução

O que é uma rubrica?

Segundo McTighe, Brookhart e Guskey (2024, p. 02, minha tradução), uma rubrica possui **“1. um conjunto de critérios alinhados aos objetivos de aprendizagem; 2. uma escala de desempenho que normalmente consiste em três ou quatro níveis; e 3. descritores que diferenciam os níveis de desempenho nessa escala”**.

Uma rubrica analítica tem dois componentes principais: **critérios** para avaliar a performance do estudante e **descrições de níveis de desenvolvimento** (Brookhart, 2014). Em uma rubrica analítica, o professor avalia **individualmente** uma variedade de critérios relevantes para a performance do estudante, atribuindo uma classificação a cada um deles de forma independente (Arter; McTighe, 2001).

Dessa maneira, é possível indicar com detalhes que um aprendiz em um nível avançado na produção de consoantes, pode ainda estar desenvolvendo suas habilidades de ritmo e entoação, por exemplo. Isso contribui para uma avaliação mais completa que indica quais caminhos já foram e quais ainda precisam ser trilhados.

Diferente de outros instrumentos avaliativos utilizados em contextos de larga escala (como exames de proficiência, por exemplo), uma rubrica desenvolvida para avaliações formativas da pronúncia conduzidas em sala de aula deve **detalhar a qualidade da produção oral do estudante**, sem necessariamente listar a necessidade de incluir ou excluir algo (“produz cinco sons diferentes...”), a frequência em que algo deve acontecer (“a cada trinta segundos, utiliza um novo ritmo...”) ou utilizar palavras que indiquem julgamento (“ótima pronúncia ao...”).

Assim, rubricas formativas, voltadas a impulsionar o processo de ensino-aprendizagem, não são instrumentos para avaliar se uma determinada atividade foi completada ou não, mas sim para, sobretudo, descreverem evidências do resultado de aprendizagem demonstrado pelos estudantes (Brookhart, 2014).

Por que avaliar a pronúncia?

A pronúncia em L2 inclui níveis fonológicos segmentais e prosódicos da fala. Enquanto o primeiro consiste em sons de **vogais e consoantes**, o segundo abrange fenômenos maiores, como o **acento tônico de palavras, a entoação e o ritmo**. Ela é uma parte fundamental da comunicação oral entre duas pessoas, garantindo que a conversa seja bem sucedida e ocorra sem dificuldades ou constrangimentos (Celce-Murcia *et al.*, 2010).

Em sala de aula, a avaliação da pronúncia permite que professores e estudantes identifiquem quais aspectos fonológicos foram plenamente adquiridos e quais ainda demandam maior atenção, enfatizando aqueles que se mostram um desafio para aprendizes brasileiros e podem comprometer significativamente a **inteligibilidade** de sua fala (Levis, 2018). Além disso, adquirir um **novo sistema fonológico** é um processo complexo, mas que pode ser facilitado pela **intervenção pedagógica** (Isbell; Sakai, 2022).

Por que utilizar rubricas?

Na avaliação de habilidades complexas, como a pronúncia, onde não podemos avaliar os estudantes com apenas “certo” ou “errado”, instrumentos avaliativos podem ser um aliado para garantir que a avaliação seja conduzida de maneira **imparcial, válida e confiável**.

Além disso, rubricas permitem que os critérios de avaliação sejam definidos e descritos com clareza; elas expressam as expectativas dos professores para com os estudantes; direcionam os esforços de ensino; monitoram o progresso de aprendizagem; promovem a autoavaliação e aumentam a consistência e a justiça da avaliação, sendo amplamente utilizadas em avaliações formativas, onde o intuito é o desenvolvimento da aprendizagem (Arter; McTighe, 2001).

Por fim, elas podem assegurar que os estudantes sejam avaliados por **critérios pertinentes** quando pensamos em uma pronúncia inteligível de usuários bem sucedidos de uma L2, não em metas inalcançáveis e irreais como “soar como um nativo”.

Componentes da rubrica

Critérios

Podemos pensar nos critérios como aquilo que estamos avaliando quando pensamos em componentes essenciais que contribuem para uma pronúncia inteligível em L2.

Vogais

Consoantes

Prosódia

Fluência e habilidade de autocorreção

Construção silábica

Compreensibilidade

Níveis de desenvolvimento

Os níveis de desenvolvimento são os diferentes estágios que os estudantes podem apresentar em sua pronúncia em cada um dos critérios acima.

Proficiente

Avançado

Progredindo

Aprendendo

Básico

Começando

Não demonstrado

Como utilizar a rubrica?

Versatilidade

Essa é uma **rubrica analítica geral** de avaliação da pronúncia em L2. Isso significa que ela pode ser utilizada em uma **variedade de contextos** de avaliação e com estudantes em níveis distintos de aprendizagem e proficiência na L2. Por exemplo, você pode utilizá-la para avaliar a pronúncia dos estudantes em uma atividade de leitura em voz alta, apresentação, provas orais, debates, etc. e com estudantes iniciantes, intermediários ou avançados. Dependendo do seu contexto de ensino e do objetivo da avaliação, adaptações podem ser necessárias (para mais informações, leia **Adaptando a rubrica**, na página 6).

Avaliação formativa

Essa rubrica é, fundamentalmente, um instrumento de avaliação formativa. Ou seja, ainda que seja possível utilizá-la para atribuir nota à produção oral de estudantes, seu intuito principal é servir como uma **ponte entre o ensino e o aprendizado** (William, 2013). Dessa maneira, identifica-se o progresso alcançado e entende-se também o caminho para chegar até mais longe. Isso permite que os estudantes compreendam com clareza o que precisam melhorar e que você, enquanto professor(a), direcione o tempo e os recursos disponíveis para enfrentar desafios específicos do seu contexto de atuação.

Envolvimento dos estudantes

Rubricas podem e devem ser compartilhadas com os estudantes durante todo o processo de ensino-aprendizagem, especialmente **antes** do seu uso em uma atividade. Ainda que esse processo possa demorar, é muito importante que a rubrica seja **apresentada e discutida com os estudantes**, garantindo que eles também a interpretem e usem com precisão. Para isso, você pode fazer uma leitura em conjunto da rubrica, responder perguntas que os estudantes tenham sobre seus critérios ou níveis de desenvolvimento, pedir para que eles se autoavaliem ou avaliem colegas com ela, etc. (Brookhart, 2024).

Como utilizar a rubrica?

Antes da avaliação

- Utilize a rubrica com as amostras de falas disponibilizadas de modo a calibrar suas avaliações;
- Discuta com colegas as suas interpretações sobre cada elemento, critério e palavra da rubrica;
- Disponibilize a rubrica para seus estudantes e, preferencialmente, estude a mesma com eles. Certifique-se que eles a entendam.

Durante a avaliação

- Se for possível, grave a fala do seu estudante;
- Anote suas observações e raciocínio ao ouvi-lo(a);
- Se ficar em dúvida, escute a amostra de fala sendo avaliada duas vezes e/ou peça para um colega avaliá-la, comparando as duas avaliações;
- Preencha a rubrica, indicando o nível de desenvolvimento do estudante em cada um dos critérios.

Depois da avaliação

- Disponibilize a rubrica preenchida para os estudantes no momento de feedback;
- Use a rubrica várias vezes ao longo do semestre para que os estudantes consigam acompanhar seu progresso;
- Beneficie-se das informações encontradas para personalizar o ensino, considerando as maiores dificuldades apresentadas.

Adaptando a rubrica

Critérios

O intuito principal desta rubrica é ser um instrumento benéfico e útil a diversos contextos pedagógicos. Por isso, ela pode ser adaptada conforme as necessidades de cada turma, tempo disponível e atividade. Por exemplo, se em um determinado semestre seus estudantes demonstraram mais dificuldades na produção de consoantes, é possível utilizar apenas esse critério da rubrica na avaliação dos mesmos.

Tipo de atividade

Além disso, você também pode focar em diferentes critérios dependendo da atividade que está sendo realizada. Por exemplo, dar uma ênfase maior para o critério de compreensibilidade durante uma atividade onde os estudantes conversam espontaneamente utilizando a L2, enquanto prioriza vogais em uma atividade de leitura em voz alta.

Nível de proficiência

O mesmo vale para os níveis de desempenho da rubrica, você pode adaptá-los de acordo com o nível de proficiência de seus estudantes. Na avaliação de aprendizes iniciantes, para não sobrecarregá-los, você pode utilizar apenas os primeiros níveis de desempenho da rubrica e ir adicionando os demais apenas conforme eles forem progredindo.

Outras adaptações não mencionadas aqui também são possíveis e devem ser consideradas por você com os objetivos de ensino que tem em mente.

Avaliando estudantes brasileiros de inglês como L2

Estudantes brasileiros podem apresentar dificuldade em sua pronúncia por uma série de motivos (Zimmer; Silveira; Alves, 2009). Ainda que muitas dessas dificuldades estejam relacionadas à **complexidade motora** de articular sons que diferem significativamente daqueles presentes em sua língua materna, é importante destacar dois processos de transferência que também influenciam esse momento da aquisição linguística. Além disso, os efeitos desses processos não ocorrem de maneira separada, sendo possível que um impacte o outro de forma cíclica.

O **primeiro** deles é o **processo de transferência fonético-fonológica** entre L1 e L2 devido à habilidade de percepção. Ao ouvir um som que só existe na L2, o aprendiz pode não o reconhecer como um novo som e, conseqüentemente, assimilá-lo a uma categoria fonológica acústica semelhante a um som de sua L1 que já é conhecido. Assim, há uma tendência que eles passem a ser percebidos e produzidos como um som existente da L1, especialmente aqueles muito parecidos (como algumas vogais) ou de difícil articulação (como a fricativa interdental /θ/). O mesmo pode acontecer com padrões de fala ou estruturas silábicas.

O **segundo** é o **processo de transferência grafo-fônica-fonológica**, influenciado pela diferença entre a ortografia e a pronúncia de uma palavra. Tanto a língua inglesa quanto a língua portuguesa utilizam o mesmo sistema alfabético, mas a correspondência entre as letras ou combinações de letras que representam sons no português brasileiro são muito mais previsíveis. Por isso, geralmente a maioria das palavras é escrita da mesma maneira que é falada.

No inglês, no entanto, um único som pode ter muitas representações **escritas diferentes** (como /i:/ em **see**, **sea**, **people**, etc.) ou o **mesmo grafema** por ser pronunciado de várias formas (como em ‘**ough**’ nas palavras **though**, **through**, **rough**, etc.). Conseqüentemente, os estudantes podem aplicar regras ortográficas da sua L1 ao decodificar palavras em inglês, o que impacta sua pronúncia, especialmente na aprendizagem em contexto formal.

Avaliando estudantes brasileiros de inglês como L2

Vogais

O inventário vocálico do inglês é muito maior que o do português. Ele conta com 9 monotongos (sons únicos) e 6 ditongos (combinação de dois sons em sequência). Dentre esses sons, algumas vogais são especialmente difíceis por não fazerem parte do inventário fonológico do português, como é o caso de /ɪ/, /ʊ/ e /æ/.

Monotongos			Ditongos	
/i/ <u>fe</u> at	/æ/ b <u>a</u> d	/ʊ/ p <u>u</u> t	/eɪ/ s <u>a</u> y	/aʊ/ c <u>ow</u>
/ɪ/ f <u>i</u> t	/ɑ/ f <u>a</u> ther	/u/ f <u>oo</u> d	/aɪ/ l <u>i</u> e	/ɔɪ/ t <u>oy</u>
/ɛ/ b <u>e</u> d	/ɔ/ sp <u>o</u> t	/ʌ/ c <u>u</u> t	/oʊ/ n <u>o</u>	/ju/ c <u>ue</u>

Fonte: Zimmer, Silveira e Alves (2009)

Uma das maiores dificuldades dos estudantes brasileiros está presente em distinguir e, consequentemente, produzir o contraste nos seguintes pares de vogais: /ɪ/ e /i/, /ʊ/ e /u/, bem como /æ/ e /e/ (Zimmer; Silveira; Alves, 2009). Alguns desses sons vocálicos são um desafio para os aprendizes porque à primeira vista eles podem soar iguais, mas além de durações diferentes, também são produzidos com força muscular, altura da língua e abertura da boca contrastantes.

Por isso, muitos estudantes tendem a não perceber com precisão a diferença que existe entre essas vogais, o que pode impactar a inteligibilidade de sua pronúncia, principalmente em palavras que são pares mínimos. As vogais mencionadas, seus pontos de articulações, bem como exemplos de como esses desvios podem afetar a comunicação em alguns contextos são apresentados com detalhes a seguir.

Legenda:

Dificuldade advinda da articulação de um som 

Dificuldade advinda da ortografia 

Pares mínimos

/ɪ/ (ship) vs /i:/ (sheep)

Descrição:

/ɪ/ é uma vogal curta e relaxada. Ela não exige grande esforço muscular. A língua permanece relativamente relaxada durante a articulação. É produzida com a língua em posição alta média e anterior, próxima do palato (céu da boca), mas não tão alta quanto em /i:/.

/i:/ é uma vogal longa e tensa. Ela exige maior esforço muscular. A língua é posicionada de forma mais firme e controlada. É produzida com a língua em posição alta e anterior, ainda mais próxima do palato, além dos lábios levemente esticados.

Produção oral de brasileiros:

Assimilação do contraste, ambas produzidas como /i/.

Exemplos de problemas para a comunicação:

Pronunciar “sin” em vez de “seen”.

/ʊ/ (full) vs /u:/ (fool)

Descrição:

/ʊ/ é uma vogal curta e relaxada. Ela não exige grande esforço muscular. É produzida com a língua em posição alta-média e posterior, próxima ao palato mole (parte posterior do céu da boca), e os lábios são levemente arredondados.

/u:/ é uma vogal longa e tensa. Ela exige maior esforço muscular. É produzida com a língua em posição alta e posterior, ainda mais próxima do palato mole. Os lábios ficam bastante arredondados e para frente, com mais tensão muscular que em /ʊ/.

Produção oral de brasileiros:

Assimilação do contraste, ambas produzidas como /u/.

Exemplos de problemas para a comunicação:

Pronunciar “pool” em vez de “pull”.

/ɛ/ (men) vs /æ/ (man)

Descrição:

/ɛ/ é uma vogal média-baixa e anterior. Ela não exige grande esforço muscular por ser mais relaxada do que /æ/. É produzida com a língua em posição média-alta e anterior, com a boca moderadamente aberta e os lábios afastados.

/æ/ é uma vogal aberta e anterior. Ela exige maior esforço muscular do que /ɛ/ por ser ligeiramente tensa. É produzida com a língua em posição baixa e frontal, com a boca bem aberta e os lábios afastados.

Produção oral de brasileiros:

Assimilação do contraste, ambas produzidas como /ɛ/.

Exemplos de problemas para a comunicação:

Pronunciar “bed” em vez de “bad” ou “said” em vez de “sad”.

Avaliando estudantes brasileiros de inglês como L2

Consoantes

No que diz respeito às consoantes, os estudantes podem apresentar dificuldades em sua pronúncia e, durante a produção oral, cometerem determinados desvios que já podemos prever (Zimmer; Silveira; Alves, 2009). Isso acontece porque, como mencionado anteriormente, há uma forte influência da ortografia na pronúncia, dificuldade de articular alguns sons e complexidade de padrões silábicos (como consoante-vogal-consoante-consoante) que não existem no português.

Entre os desafios mais comuns para os estudantes estão substituir uma consoante esperada por outra que seja mais fácil de articular oralmente, produzir incorretamente as consoantes nasais em posição final de palavras, simplificar encontros consonantais com a adição de uma vogal em vários contextos, como em palavras com o sufixo ‘-ed’, por exemplo. Todos esses desvios podem levar a dificuldades de interpretação da intenção original da fala e ocasionar mal entendidos durante a comunicação. Os processos descritos são explicados com exemplos a seguir.

Substituição de consoantes

/θ/ (think/faith) e /ð/ (the/with) 

Descrição:

/θ/ e /ð/ são consoantes fricativas interdentais. São produzidas com a ponta da língua posicionada atrás dos dentes incisivos superiores.

Produção oral de brasileiros:

/θ/ pode ser produzida como /t/, /f/ ou /s/; /ð/ pode ser produzida como /d/ ou /z/.

Exemplos de problemas para a comunicação:

Pronunciar “sink” em vez de “think”.

Por que isso acontece?

Dificuldade de articulação de um som que não existe no inventário fonológico do português.

/h/ (him/house)

Descrição:

/h/ é uma consoante fricativa glotal. Ela é produzida através da passagem de ar pela glote.

Produção oral de brasileiros:

Produzida como /x/ (de ‘rato’) ou totalmente omitida.

Exemplos de problemas para a comunicação:

Pronunciar “hat” com o /x/ do português ou como “at”.

Por que isso acontece?

Dificuldade de articulação e influência da ortografia (em português, o ‘h’ geralmente é mudo, como em ‘hospital’).

/tʃ/ (chart/chicken)

Descrição:

/tʃ/ é uma consoante pós-alveolar surda. Ela é produzida com a ponta da língua posicionada próxima ao palato duro, mas sem o tocar completamente.

Produção oral de brasileiros:

Produzida como /ʃ/ (de ‘chá’).

Exemplos de problemas para a comunicação:

Pronunciar “sheep” em vez de “cheap”.

Por que isso acontece?

Influência da ortografia (em português, as palavras com ‘ch’ tem som de /ʃ/ (chá) e não /tʃ/ (tchau)).

/ɹ/ (rat/very/car)

Descrição:

/ɹ/ é uma consoante aproximante rótica. Ela é produzida com a língua curvada para trás, sem tocar no palato.

Produção oral de brasileiros:

Produzida como /x/ (de ‘rato’).

Exemplos de problemas para a comunicação:

Pronunciar “head” em vez de “red”.

Por que isso acontece?

Dificuldade de articulação e influência da ortografia (em português, o ‘r’ pode ser uma fricativa uvular (como em ‘rei’) ou uma vibrante simples alveolar (como em ‘laranja’), diferente do /ɹ/ de ‘run’ em inglês. No entanto, ortograficamente todos esses sons são representados pela mesma letra).

Vocalização de consoantes nasais no final das palavras

/m/ (spam)

Descrição:

/m/ é uma consoante bilabial. Ela é produzida com os dois lábios se tocando completamente. Há liberação de ar pelo nariz.

Produção oral de brasileiros:

O /m/ não é completamente articulado (com o encontro total dos lábios).

Exemplos de problemas para a comunicação:

“Seem” pode ser confundida com a palavra “sim” do português.

Por que isso acontece?

Dificuldade de articulação (em português, ‘sim’ /sĩ/ termina com uma vogal nasal, sem a completa articulação da consoante /m/).

/n/ (can)

Descrição:

/n/ é uma consoante alveolar. Ela é produzida com a ponta da língua tocando os alvéolos (a região atrás dos dentes superiores). Há liberação de ar pelo nariz.

Produção oral de brasileiros:

O /n/ não é completamente articulado (com o contato da língua com os alvéolos).

Exemplos de problemas para a comunicação:

“Ten” pode ser confundida com a palavra “tem” do português.

Por que isso acontece?

Dificuldade de articulação (em português, ‘glúten’ /'gluteĩ/ termina com uma vogal nasal, sem a completa articulação da consoante /n/).

/ŋ/ (sing)

Descrição:

/ŋ/ é uma consoante velar. Ela é produzida com a parte de trás da língua tocando o palato mole.

Produção oral de brasileiros:

/ŋ/ é produzido como uma vogal nasal OU adiciona-se um /g/ no final da palavra (/sɪŋ/ > /sɪŋg/).

Exemplos de problemas para a comunicação:

“Sing” pode ser confundida com a palavra “sim” do português ou é produzido como parte de um encontro consonantal /ŋg/.

Por que isso acontece?

Dificuldade de articulação e influência da ortografia (em português, o ‘ng’ representa o som /ŋg/, como em ‘singular’).

Simplificação silábica

Redução ou omissão de encontro consonantal, especialmente em final de sílaba

Exemplo no inglês norte-americano:

desk /dɛsk/

Produção oral de brasileiros:

/dɛs/ sem o /k/ final.

Exemplos de problemas para a comunicação:

“Ask” pode ser confundida com “as”.

Por que isso acontece?

Dificuldade de articulação (omite-se a consoante final para facilitar a pronúncia já que em português a estrutura silábica CVCC não é comum).

Adição de uma vogal (geralmente /ɪ/ ou /ə/) no começo ou final de palavras

Exemplo no inglês norte-americano:

book /bʊk/ – study /'stʌdɪ/

Produção oral de brasileiros:

/ˈbʊki/ com adição de /i/ no final – /is'tʌdɪ/ com adição de /i/ no começo.

Exemplos de problemas para a comunicação:

“Milk” pode ser confundida com “milky”, deixando a fala infantilizada.

Por que isso acontece?

Dificuldade de articulação (adiciona-se uma vogal para facilitar a pronúncia já que em português encontros consonantais no começo ou no final de sílabas não são comuns).

Adição de uma vogal (geralmente /e/) entre consoantes após a adição do sufixo ‘-ed’

Exemplo no inglês norte-americano:

looked /lʊkt/ – moved /muvd/

Produção oral de brasileiros:

/lʊked/ com adição de /e/ entre /k/ e /d/ – /muved/ com adição de /e/ entre /v/ e /d/.

Exemplos de problemas para a comunicação:

“Asked” pode soar como “ask it” ou o tempo verbal referente ao passado pode não ser reconhecido.

Por que isso acontece?

Dificuldade de articulação (adiciona-se uma vogal para facilitar a articulação já que em português encontros consonantais (/lʊkt/) não são comuns) e influência da ortografia (ao contrário do português, o /e/ está presente na palavra, mas não é pronunciado).

Posição silábica da consoante

Aspiração de oclusivas surdas em sílabas tônicas seguidas de uma vogal

Exemplo no inglês norte-americano:

/p/: pot /p^hɒt/ - /t/: attack /ə't^hæk/ - /k/: car /k^hɑr/

Produção oral de brasileiros:

/p/, /t/ e /k/ sem aspiração: /pɒt/, /ə'tæk/ e /kɑr/.

Exemplos de problemas para a comunicação:

“**P**at” pode ser confundida com “**b**at”, além de demandar mais esforço do ouvinte para compreender o início de uma palavra.

Por que isso acontece?

Dificuldade de articulação (em português, essas consoantes são produzidas sem a explosão forte do ar, como em ‘pode’, ‘tenho’ e ‘casa’).

Delateralização fazendo o /l/ em posição final de palavras virar uma vogal ou semivogal

Exemplo no inglês norte-americano:

well /wɛl/

Produção oral de brasileiros:

/wɛw/, com o som final sendo o da semivogal /w/.

Exemplos de problemas para a comunicação:

“Cool” pode ser confundida com “coo”.

Por que isso acontece?

Dificuldade de articulação (em português, o /l/ em posição final é vocalizado, sendo produzido como /w/, como nas palavras ‘mal’ e ‘papel’).

Desvozeamento de consoantes sonoras em posição final de palavras (/b/, /d/, /g/, /v/, /ð/, /z/, /ʒ/)

Exemplo no inglês norte-americano:

mad /mæd/ - dog /dɔg/

Produção oral de brasileiros:

/mat/ substituindo o /d/ (vozeado) por /t/ (surdo) e /dɔk/ substituindo o /g/ (vozeado) pelo /k/ (surdo).

Exemplos de problemas para a comunicação:

“Lea**v**e” pode ser confundida com “lea**f**”.

Por que isso acontece?

Dificuldade de articulação (em português, consoantes sonoras geralmente não aparecem no fim de palavras, então para facilitar elas são produzidas como consoantes surdas (/b/, /t/, /k/) que não exigem vibração das cordas vocais).

Como atribuir notas com a rubrica?

Para transformar os níveis de desempenho da rubrica em notas numéricas, alguns **cuidados** e **decisões** precisam ser considerados. É importante destacar que a conversão direta da escala da rubrica (que apresenta seis níveis) para números pode alterar o significado pedagógico da avaliação e não representar com clareza o progresso do estudante.

Por exemplo, um estudante avaliado no nível mais alto de desempenho da rubrica (nível 6) não deve receber automaticamente a nota 6 em um sistema tradicional de notas, já que esse valor indicaria uma aprendizagem ainda em desenvolvimento, com vários aspectos a melhorar.

Atualmente, não há consenso na literatura sobre a melhor forma de transformar o desempenho visto em rubricas em notas numéricas. Ainda assim, reconhecemos a importância dessa conversão para a prática docente, especialmente em contextos em que o professor precisa registrar notas no sistema escolar.

Os passos descritos a seguir foram pensados **apenas para esta rubrica** e partem do pressuposto de que você deseja obter uma nota de pronúncia para compor a nota final da produção oral em L2 dos seus estudantes.

1º passo

Anote os níveis de desempenho alcançados pelo estudante em cada critério da rubrica. Por exemplo:

Vogais: 3

Consoantes: 4

Prosódia: 4

Fluência e habilidade de autocorreção: 5

Formação silábica: 4

Compreensibilidade: 5

Se quiser saber mais sobre rubricas e notas, veja o capítulo 11 de Brookhart (2014).

2º passo

Coloque os níveis anotados em ordem crescente:

3 4 4 4 5 5

3º passo

Calcule a mediana dos valores. A mediana é usada porque representa melhor o desempenho geral do estudante, sem ser influenciada por valores muito altos ou muito baixos em critérios específicos.

Como a quantidade de números de critérios é par, observe os dois valores que ficam exatamente no meio da sequência.

No exemplo, esses valores são **4 e 4**.

Agora, some os dois números e divida o resultado por 2:

$$4 + 4 = 8$$

$$8 \div 2 = 4$$

O valor **4** será a mediana do desempenho do estudante.

4º passo

Use a tabela de conversão abaixo para transformar a mediana da rubrica em uma nota numérica. Por exemplo: Se a mediana do estudante for 4, a nota atribuída será 8.

Mediana	Nota
6	10
5	9
4	8
3	7
2	6
1	5

Se você estiver utilizando apenas um critério da rubrica em uma avaliação e quiser atribuir uma nota ao desenvolvimento do estudante, a tabela de conversão também pode ser utilizada.

Recursos para estudo

Se você tiver dúvidas sobre algum critério da rubrica ou quiser estudar mais sobre componentes fonético-fonológicos, recomendamos o site **Teaching Pronunciation with Confidence**.

Os materiais estão disponíveis apenas em inglês, mas o site é bastante intuitivo. Ao acessar o link e rolar para baixo, você pode escolher o aspecto da pronúncia que deseja estudar (vogais, consoantes ou prosódia), ler as explicações, ouvir exemplos em áudio e realizar atividades práticas com feedback.

TEACHING PRONUNCIATION WITH CONFIDENCE

Chapter 4

Consonants

Agata Guskarska

Learning Objectives

- To define consonants and their characteristics
- To explore innovative ways to teach consonant pronunciation to students
- To introduce teachers to minimal pairs and tasks to use them in the classroom
- To provide a database for most commonly mispronounced vowels

Outline

- 4.1 What are consonants?
- 4.2 Why are they important in English speech?
- 4.3 Are some consonants more important than others?
- 4.4 How to approach teaching consonants?
- 4.5 Technology Corner
- 4.6 Activities
- 4.7 References

4.1 What are consonants?

Consonants are sounds that are produced with narrowing in the mouth, which causes some obstruction of air stream. They differ from each other based on place and manner of articulation and voicing. Figure 4.1 below illustrates this classification of the consonants in GA English.

MAINER	VOICING	Bilabial	Labiodental	Dental	Alveolar	Palatal	Velar	Glottal
plosive	voiceless	p			t		k	ʔ
plosive	voiced	b			d		g	
fricative	voiceless		f	θ	s	ʃ		
fricative	voiced		v	ð	z	ʒ		
affricate	voiceless					tʃ		
affricate	voiced					dʒ		
nasal	voiced	m			n		ŋ	
lateral fricative	voiced				l			
fricative (liquid)	voiced					r [ɹ]		
slide	voiced	w				ɹ	ɹ [ɹ]	

Table 4.1 Consonants classification

Glossário de termos

Acento tônico (word stress)

Ênfase dada a uma sílaba específica de uma determinada palavra, fortalecendo-a, deixando-a mais longa ou alta que as restantes (TAbLe, reLAX, comPUter, PArent, reveLAtion).

Acento frasal (sentence stress)

Ênfase geralmente dada a palavras importantes e funcionais (substantivos, verbos, adjetivos, advérbios) em uma frase ("They are WATCHing a MOVie toNIGHT").

Adição/simplificação de sons

Acrescentar um som que não existe na palavra (/Istap/ em vez de /stap/ na palavra "stop") ou simplificar um grupo de sons esperado (/askid/ em vez de /æskt/ na palavra "asked").

Aspiração

Sopro de ar presente na pronúncia de consoantes oclusivas surdas como /p/, /t/ e /k/ em sílaba tônicas, seguida por uma vogal ("pet" apresenta uma aspiração no /p/).

Compreensibilidade

Facilidade para entender o que foi dito.

Confiabilidade

Grau em que uma atividade/instrumento gera resultados consistente e estáveis em diferentes momentos ou entre diferentes avaliadores (por exemplo, avaliar os estudantes sem critérios pré-estabelecidos claros pode fazer com que diferentes professores atribuam notas diferentes a um mesmo estudante, gerando resultando inconsistentes e não confiáveis).

Entoação

Envolve o uso de tom para adicionar significado a mensagem dita pelo falante. Para o ouvinte, indica se o falante está afirmando algo, fazendo uma pergunta, demonstrando alguma emoção específica (insegurança, imparcialidade, provocação, etc.).

Inteligibilidade

Decodificação da informação (sinal sonoro) que foi produzida por um falante.

Fluência

Capacidade de falar um grande número de palavras de forma contínua sem pausas desnecessárias.

Ritmo

Padrões de tempo da fala. Na língua inglesa, há uma alternância entre sílabas longas e fortes (tônicas) e curtas e fracas (átonas).

Omissão de sons

Excluir um som que deveria estar na palavra (falar "tes" em vez de "test").

Glossário de termos

Pares mínimos

Duas palavras que podem ser diferenciadas por um único som (“bit” /ɪ/ vs “beet” /i:/).

Prosódia

Elementos da fala que incluem ritmo, entonação e acento em segmentos individuais (vogais e consoantes) ou maiores (frases inteiras).

Substituição de sons

Trocar o som esperado por outro (“fink” em vez de “think”).

Validade

Grau em que uma atividade/instrumento mede a habilidade que se propõe a avaliar (por exemplo, pedir que o estudante leia um texto em voz alta para avaliar sua produção oral espontânea em L2 é um exemplo de uma atividade com baixa validade, uma vez que o que está sendo medido aqui é sua capacidade de compreensão escrita e articulação).

Acesse a rubrica de avaliação da pronúncia

Para acessar ou fazer o download da **Rubrica de Avaliação da Pronúncia em L2 para Estudantes Brasileiros de Inglês**, desenvolvida neste estudo, clique **aqui**.

Para acessar **dois áudios de brasileiros selecionados do Speech Accent Archive** e praticar suas avaliações utilizando a rubrica, clique **aqui**.

O parágrafo que os falantes estão lendo inclui palavras comuns do cotidiano, mas em **sequências complexas** que podem ser difíceis. O texto contém praticamente **todos os sons** da língua inglesa.

"Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station".

Em caso de dúvidas, entre em contato com a pesquisadora pelo WhatsApp ou e-mail.

Obrigada!

Referências

ANDRADE, H. L. What Is Next for Rubrics?. *In*: GOLSAVES, C.; PEARSON, J. (Ed.). **Improving Learning Through Assessment Rubrics: Student Awareness of What and How They Learn**, 2023. p. 314-326.

ARTER, J.; MCTIGHE, J. **Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance**. Corwin Press, 2001.

BROOKHART, S. M. **How to create and use rubrics for formative assessment and grading**. ASCD, 2014. Disponível em: [https://www.geocities.ws/bdktraining/pdfkur/How%20to%20Create%20and%20Use%20Rubrics%20for%20Formative%20Assessment%20and%20Grading%20\(%20PDFDrive%20\).pdf](https://www.geocities.ws/bdktraining/pdfkur/How%20to%20Create%20and%20Use%20Rubrics%20for%20Formative%20Assessment%20and%20Grading%20(%20PDFDrive%20).pdf).

BROOKHART, S. M. Using Rubrics in Basic Education: A Review and Recommendations. **Estudos em Avaliação Educacional**, v. 35, 2024.

CELCE-MURCIA, M. *et al.* **Teaching pronunciation: A course book and reference guide**. Cambridge University Press, 2010.

ISELL, D. R.; SAKAI, M. Pronunciation assessment in classroom contexts. *In*: LEVIS, J. M.; DERWING, T. M.; SONSAAT-HEGELHEIMER, S. (Ed.). **Second language pronunciation: Bridging the gap between research and teaching**. Wiley Blackwell, 2022. p. 194-214.

GUSKAROSKA, A.; ZAWADZKI, Z.; LEVIS, J. M. **Teaching pronunciation with confidence: A resource for ESL/EFL teachers and learners**. Iowa State University Digital Press, 2024.

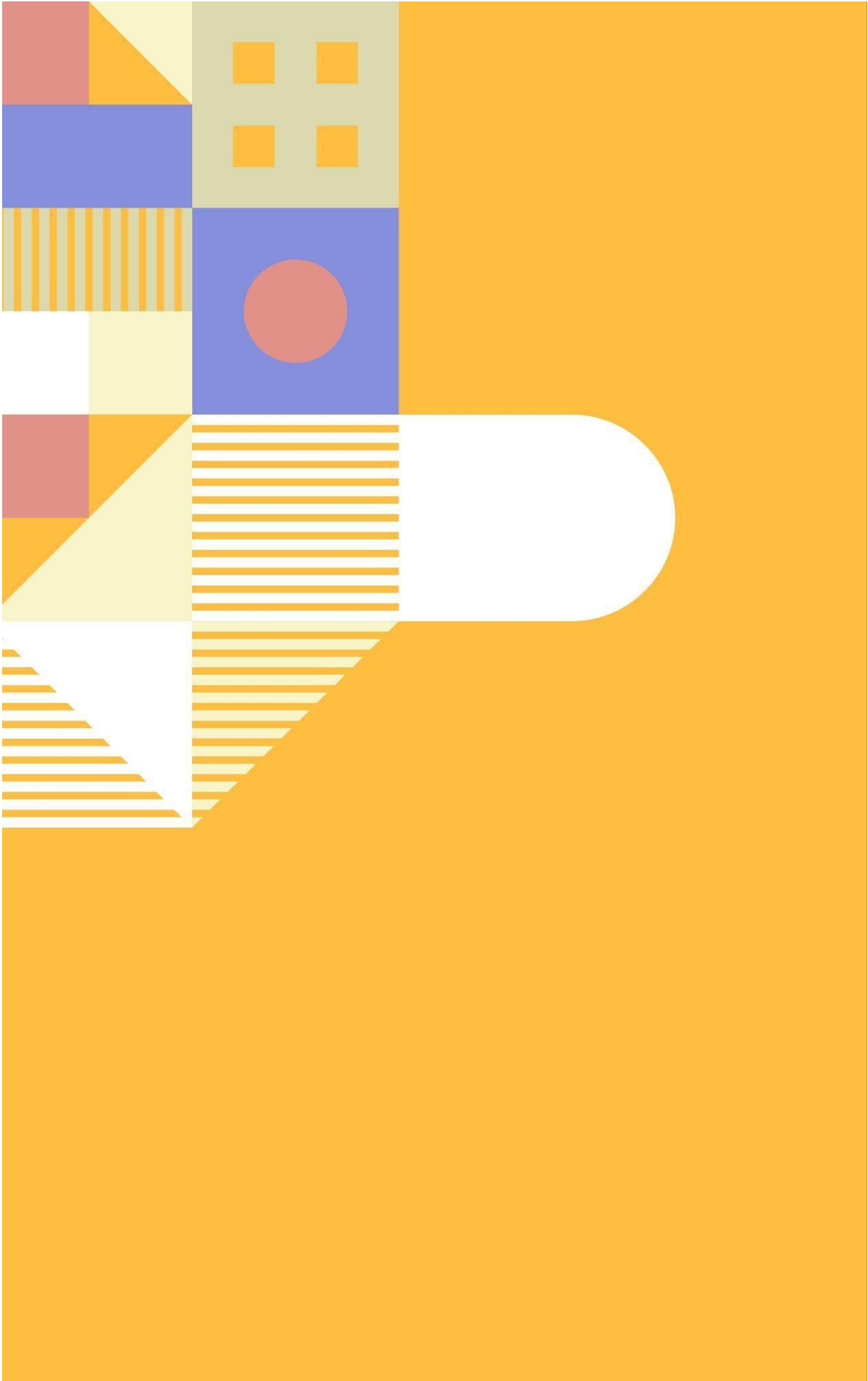
LEVIS, J. M. **Intelligibility, oral communication, and the teaching of pronunciation**. Cambridge University Press, 2018.

WEINBERGER, S. **Speech Accent Archive**. George Mason University, 2015. Disponível em: <http://accent.gmu.edu>. Acesso em: 13 Set. 2025.

WILLIAM, D. Assessment: The bridge between teaching and learning. **Voices from the Middle**, v. 21, n. 2, p. 15-20, 2013.

MCTIGHE, J.; BROOKHART, S.; GUSKEY, T. The value of descriptive, multi-level rubrics. **Educational Leadership**, v. 81, n. 6, p. 1-8, 2024.

ZIMMER, M. C. SILVEIRA, R.; ALVES, U. K. **Pronunciation instruction for Brazilians**, 2009.



APPENDIX F - MESSAGES SENT TO PARTICIPANTS

After participants completed the Background Questionnaire and the Assessment Survey

Oi, ____! Tudo bem?

Eu me chamo Manuella e recentemente você aceitou participar da minha pesquisa de mestrado. Estou entrando em contato para explicar como participar da próxima fase do estudo.

Você precisa acessar o Moodle Grupos, onde terá acesso à rubrica e ao manual de avaliação da pronúncia para a sala de aula que desenvolvi.

Como acessar:

Clique no link: moodle.com

Faça login (ou cadastre-se, se ainda não tiver conta no Moodle).

Digite a senha de inscrição: _____

Na página "Núcleo de Pesquisa em Fonética e Fonologia Aplicada à Língua Estrangeira", clique no ícone "Pesquisa de Avaliação da Pronúncia em L2" para assistir ao vídeo tutorial e fazer download do material.

Prazo:

Você tem cerca de 1 semana para conhecer o material. Após isso, no dia __/__, volto a entrar em contato para enviar o último questionário da pesquisa onde você poderá falar da sua experiência com a rubrica e dar sugestões sobre como melhorá-la.

Peço que confirme o recebimento desta mensagem, por favor.

Qualquer dúvida estou à disposição! Muito obrigada de novo pela sua participação!

One week after participants joined the Moodle page and viewed/downloaded all the study resources

Oi, ____! Tudo bem?

Verifiquei que você conseguiu acessar o Moodle e os materiais da pesquisa. Por isso, estou te enviando o último questionário da coleta de dados, onde você avaliará a rubrica e o manual desenvolvidos. Como você irá avaliar a fala de três estudantes, é recomendado usar um computador e fones de ouvido para respondê-lo. Recomendo, também, reservar cerca de 25 a 35 minutos para completar tudo com calma.

Peço que envie suas respostas até __/__, mas caso precise de mais tempo é só me avisar.

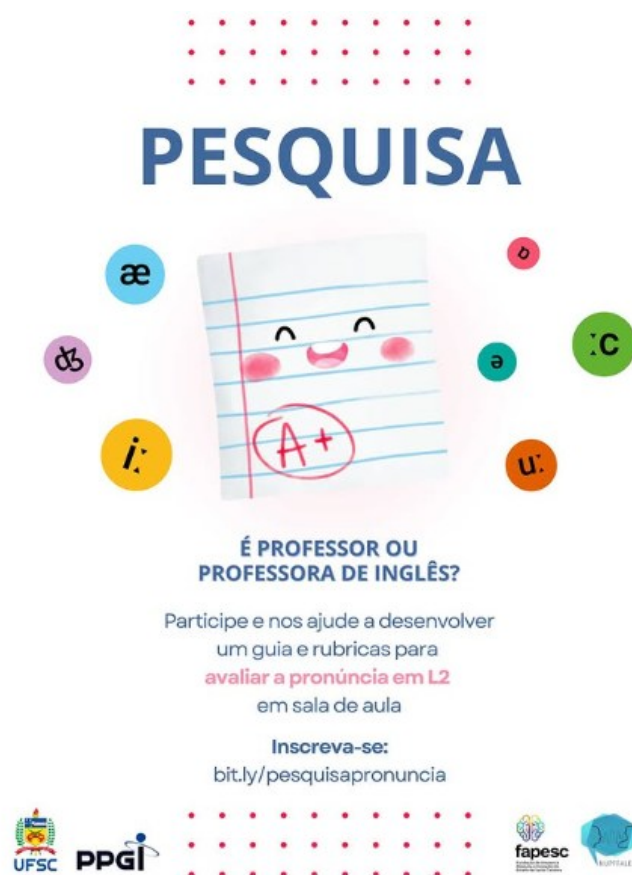
Link para responder ao último questionário: forms.com

Peço que confirme o recebimento desta mensagem, por favor.

De novo, muito obrigada pela sua participação!

APPENDIX G - INVITATION VIA SOCIAL MEDIA

Image shared on Instagram



APPENDIX H - DATA SPREADSHEETS

The data analyzed in this study are available at:
https://docs.google.com/spreadsheets/d/1wjkomluMP78KkquEVZ83U_ZLKRX_BEGbaOW4TTnTnEE/edit?usp=sharing.