EXPERIENCE REPORT

Analyzing the oral performance of Portuguese as a Host Language speakers through the Outcome Achievement measure

Ana Flavia Boeing MARCELINO 🆔
Universidade Federal de Santa Catarina (UFSC)

Raquel Carolina de Souza Ferraz D'ELY 🆔
Universidade Federal de Santa Catarina (UFSC)

ABSTRACT

The teaching of Portuguese as a Host Language has faced many challenges since its establishment as part of the field of Portuguese as a Foreign Language, mainly for presenting specificities unknown to the teachers and researchers in the field (GROSSO, 2010; DEUSDARÁ; ARANTES; BRENNER, 2018). Among these challenges are the methods employed to analyze the oral performance of the Host Language speakers. With this in mind, this report aims at discussing the evaluation of the oral performance of adult immigrants, beginner students of Brazilian Portuguese as a Host Language, based on two descriptive scales of the measure of Outcome Achievement, a multifaceted measure that looks, mainly, at pragmatic aspects of language use in tasks whose main focus is the communicative outcome of the performance. Even though the proposal of criteria that compose the measure of *Outcome Achievement* is based on the interpretation of raters, which might bring incongruences to the evaluations, the measure presents an alternative to evaluating oral performance in contrast with other more traditional measures. In conclusion, the employment of the Outcome Achievement measure to analyze oral performance in tasks to teach Portuguese as a Host Language brings aspects that are

inherent to this context of teaching, taking into account the immediate objectives of language use of the immigrant population.

RESUMO

O ensino de Português como Língua de Acolhimento (PLAC) tem enfrentado inúmeros desafios desde seu estabelecimento como parte da área do ensino de Português como Língua Estrangeira, por apresentar especificidades desconhecidas até então por professores e pesquisadores da área (GROSSO, 2010; DEUSDARÁ; ARANTES; BRENNER, 2018). Dentre estes desafios estão os métodos de avaliação empregados para medir o desempenho oral dos falantes da língua de acolhimento. Com este cenário em vista, este relato tem o objetivo de discutir a avaliação do desempenho oral de estudantes adultos imigrantes, aprendizes iniciantes de Português como Língua de Acolhimento na variante brasileira, com base em duas escalas da medida *Outcome Achievement*, uma medida multifacetada que olha, principalmente, para aspectos pragmáticos do uso da língua em tarefas onde o foco principal é o resultado comunicativo do desempenho. Apesar de a proposta de análise dos critérios que compõem a medida *Outcome Achievement* se basear na interpretação de avaliadores, podendo trazer incongruências nas avaliações, a medida apresenta um alternativa a métodos de avaliação do desempenho oral que se baseiam em medidas mais tradicionais. Conclui-se que o emprego dessa medida à análise do desempenho oral em tarefas para o ensino de português como língua de acolhimento abrange aspectos inerentes a este contexto de ensino, levando em consideração os objetivos imediatos de uso da língua da população imigrante.

KEYWORDS

Portuguese as a Host Language. Evaluation methods. Oral performance. Outcome Achievement.

PALAVRAS-CHAVE

Português como Língua de Acolhimento. Método de avaliação. Desempenho oral. *Outcome Achievement.*

Introduction

The teaching of Portuguese as a Host Language has increased in importance in the latest years and it became an urgent matter for research to embark on investigating the processes taking place in these classrooms. Although it is of common knowledge that the teaching of Portuguese as a Host Language can be inserted alongside other specific contexts of teaching under the umbrella field of Portuguese as a Foreign Language, the Host context brings its specificities to the discussion of how to teach and evaluate language learning, which has not been addressed by the broader area yet (JENSEN, 2002; LOPES, 2009; DEUSDARÁ; ARANTES; BRENNER, 2018).

Among the specificities brought by the teaching of Portuguese as a Host language is a necessity for specialized teachers, for personalized teaching materials (SILVEIRA; XHAFAJ, 2020), and consequently for adequate techniques and methods for analyzing the language development of the speakers in this context. These aspects have to be coherent with the students' context of learning and, primarily, with their learning objectives. However, the tradition in the field of language learning, concerning the evaluation of language development has been strongly connected to standardized tests that look mainly at structural elements of language which might not be the most appropriate approach to the Host Language context – though some important changes have happened in the recent years (LONG, 2015), with the advance of proficiency tests and scales that are more function-oriented instead of grammar-based.

Teaching Portuguese as a Host Language (PHL) involves understanding that the population in focus is composed of immigrants and refugees that have been forcibly displaced[1] from their home countries and are looking for establishing in Brazil for a new life. Therefore, the immediate needs of this population concerning language learning should consider the situations of immediate adaptation to the ways of living in the new country (GROSSO, 2010). In this matter, research in the field has lacked a careful consideration of how to prepare and evaluate language performance considering these populations' immediate communication needs.

Having in mind this context's needs for tailored syllabuses that meet the population's communication needs and, consequently, evaluates their performance accordingly, this study aims at presenting an alternative way of assessing students' oral performance in tasks in the context of PHL for beginners in Brazil. We aim at assessing students' oral performance not through the well-known CALF measures but adopting the measure of Outcome Achievement (based on FARIAS, 2018), which is a multifaceted construct, that does not rely solely on vocabulary use, or grammatical accuracy, for instance, but instead has a strong concern for meaning and contextual elements that may dictate the adequacy of the vocabulary or grammar to be used. This article reports on the process of designing and implementing two oral tasks, and two descriptive scales for Outcome Achievement which were, in turn, designed to meet the characteristics of these two oral tasks.

_____

[1] For the purposes of this article, the terms immigrants, refugees, and forcibly displaced people are used interchangeably, although we understand that they might hold different legal and social statuses.

This article is organized as follows. First, we present an overview of the fields of Portuguese as a Host Language and the traditions in evaluating oral language performance in tasks. Then, we describe the method employed to gather data for the study, focusing on the population's characteristics and the steps followed to first, implement the two tasks in a classroom setting; second, design the two Outcome Achievement descriptive scales; and finally, implement it for analyzing the participants' oral performances. To conclude, we bring the results of the performances in the two tasks and suggest further steps to investigate the measure.

1. Portuguese as a Host Language and the traditions of oral language assessment of performance in tasks

In order to analyze the proposal of a recently conceptualized measure, as is the Outcome Achievement, it is key to look at the traditions in research that have been using different measures to analyze performance, and specifically, in this case, oral language performance. To do so, it is important to look at performance having in mind the context of teaching and learning to which the measure is being proposed here. Therefore, this section is organized to, first, bring an overview of the context of Portuguese learning and teaching as a Host Language; and second, present the background of studies that inform the analyses of oral language performance.

The concept of Host Language has been frequently addressed by some researchers (GROSSO, 2010; LOPEZ, 2018; OLIVEIRA; SILVA, 2017) to debate the recent migration movements taking place in all parts of the world. Although migration has been a common phenomenon throughout the history of humanity, some would say that the last years could be subscribed as "the biggest migration crisis of all times" (DEUSDARÁ; ARANTES; BRENNER, 2018, p. 3228, our translation). According to the United Nations High Commissionaire for Refugees' (UNHCR) 2018 report, 68.5 million people around the world had been recently forced from their homes, which, according to the report, represents the highest levels of displacement on record ever witnessed. Moreover, also according to UNHCR, about 85% of the world's displaced people are in developing countries, such as Brazil.

As a consequence, this growing number of immigrants who have seen Brazil as an alternative place in which they can rebuild their lives has caused important changes in the profile of the population in need to speak Portuguese. The urge to understand the specificities of this population has been growing and it has shown the field's need for language teaching and learning practices that fit this population best. It has been perceived that the tools available, usually used for the teaching of Portuguese as a Foreign Language, do not fit this population's communication needs in the first moments of welcoming in the new country (DEUSDARÁ; ARANTES; BRENNER, 2018).

The need to develop initiatives to assist the immigrants in the societies to which they move is urgent, once the situation of forcibly displaced people highlights necessities related to legal affairs,

housing and working issues, and even more urgent matters nowadays such as access to the health systems, taking into consideration the COVID-19 pandemic. And in all these matters, language is a key aspect that may help or hinder the adaptation of forcibly displaced people to their new home.

In this context, speaking Portuguese is not only a matter of being proficient in interacting with native speakers of the target language in diverse situations; it is a matter of being proficient in solving particular and immediate necessities. According to Grosso (2010), the Host Language concept is linked to the host context and the migratory context. It is generally composed by an adult audience, who might learn Portuguese "due to different contextual needs, often linked to the resolution of urgent survival issues, in which the host language has to be the link of affective (bidirectional) interaction as the first form of integration (in linguistic immersion) to a full democratic citizenship" (GROSSO, 2010, p. 74, our translation).

Arantes, Brenner and Deusdará (2016) also highlight the idea that the "teaching of a host language to refugee youths and adults presents specificities that are not the same as those of foreign language teaching" (p. 1202 *apud* DEUSDARÁ; ARANTES; BRENNER, 2018, p. 3229, our translation) and due to this fact, many of the traditional approaches to teaching and learning a language might not consider aspects that are crucial to this context. One example is the fact that this population might not aim at developing the most accurate speaking skills in the short term, but instead might aim to learn enough Portuguese that could guarantee effective communication in the city hall, with the Federal Police department, with health professionals and other situations that are of an immediate necessity to them (MARCELINO, 2020).

Accordingly, teachers and researchers should find alternatives to prepare the learners of Portuguese as a host language to perform in these situations and to assess their performances in order to develop their oral skills globally, but also giving special attention to those specific communication needs.

In this aspect, we have seen a growing number of studies that have addressed the teaching of Portuguese as a Host Language, presenting options of practices that allow for a focus on the communication needs of this population, as is the case of studies in the field of Task-Based Language Teaching (TBLT) (SILVEIRA; XHAFAJ, 2020; MARCELINO, 2020; LOPEZ, 2016; 2018; CURSINO; ALBUQUERQUE; FIGUEIREDO SILVA; GABRIEL; ANUNCIAÇÃO, 2016, to cite a few recent studies).

What we need to know about TBLT now, considering the purposes of this article, is that at its core is the idea that people use language to plan, perform and recall the activities done in the "real-world" (LONG, 2015) – meaning the world outside the classroom walls. The language classroom, therefore, assumes this functional perspective for the use of language bearing in mind the *communicative purpose/objective* of the interactions that take place in the everyday life. According to Long (2015), the tasks are the units of "analysis throughout the design, implementation, and evaluation" (p. 6) of a language course, and consequently, its evaluation should be coherent to the kind of tasks students are required to perform in class and, at the same time, outside the classroom.

Moving on to looking at oral tasks in the field of TBLT, there has been a preference for analyzing performance through what is called the CALF measures (which stand for Complexity, Accuracy, Lexical density, and Fluency) (FOSTER; SKEHAN, 1996; PALLOTTI, 2009; SKEHAN, 1996, 1998, 2003).

These measures are applied to analyze oral performance usually by means of transcribed speech and computer programs in charge of the counting of subordination in the sentences, number of errors per clause, number and variety of words used, and number of pauses and false starts, to cite a few possibilities. However, some recent studies have questioned the idea that to perform successfully in an oral task the speech produced has to be complex, accurate, lexically rich, and fluent (REVESZ; EKIERT; TORGERSEN, 2016). More specifically, research that has had its onset in pedagogical contexts has focused on the learners' immediate communication needs and has claimed that to perform tasks successfully, one's main concern is to communicate the message and fulfill that need regardless of some violation of the language rules that might appear in these learners' speeches.

In this sense, the pragmatic aspects of language play a more important role in the communication – and consequently in the performance – than the structural elements of language, such as the adequate use of grammar. We understand that these pragmatic elements are related to the "conditions governing the language use, the linguistic practice" (FIORIN, 2003, p.161, our translation) and how these conditions affect the linguistic choices made by learners. In our point of view, the pragmatic elements of the enunciation are not captured by the CALF measures, such as the people, the time, and the space in which the communication happens. Therefore, there is a need to investigate other assessment tools that consider these aspects when determining communicative success, for instance.

In the field of TBLT, Pallotti (2009), as well as Robinson (2001), have already highlighted the fact that communicative success and adequate communication have been little investigated in the field and, consequently, suggestions of alternative measures have been put under scrutiny, such as *adequacy* and *outcome achievement*.

The measures of adequacy and outcome achievement are known to assess language performance taking into consideration the context in which the language is being used and the interlocutors the speech addresses, for instance. According to Farias (2018), both constructs are very similar and may overlap depending on the nature of the task to which it is applied. Pallotti (2009) states that A*dequacy* "represent[s] the degree to which a learners' performance is more or less successful in achieving the task's goals efficiently" (p. 596), while Farias (2018) explains that "differently from Adequacy, that puts attention on examining if *language* was used adequately for communicative purposes, Outcome Achievement is focused on investigating if the expected outcome of a task was accomplished" (p. 69, our emphasis).

Adequacy and outcome achievement are constructs that place an important part of successful communication in conveying the appropriate genre of communication that the tasks require, by agreeing with the contexts in which the tasks are inserted, for instance. And, due to the fact that contexts of language use entail different aspects of language production, the alternative to assess performance in these terms is to adopt descriptive scales (PALLOTTI, 2009) *tailored* to the tasks learners are required to perform and that consider the pragmatic components of language in use (FIORIN, 2003).

Pallotti (2009) explains that one example of measure scales to rate language production is the well-known Common European Framework of Reference for Languages: Learning, teaching, assessment, that "presents a comprehensive descriptive scheme of language proficiency and a set of

Common Reference Levels (A1 to C2) defined in illustrative descriptor scales..." (COUNCIL OF EU-ROPE, 2020, p. 27). And in the Brazilian context, there is the Exam for the Certification of Proficiency in Portuguese (CELPE-BRAS), which claims to be "based on the idea of proficiency as the adequate use of language to conduct actions in the world. The exam considers textual elements and, mainly, discoursive elements: the context, the purpose and interlocutors involved in the interaction." (MIN-ISTÉRIO DA EDUCAÇÃO, n/d, our translation).

Although the descriptions of the measures of adequacy and outcome achievement resemble CELPE-BRAS's idea of proficiency, the oral part of the exam presents limitations for assessing learners in the host language context of learning. The main limitation being represented, in our point of view, by the fact that the only genre of language use available to assess oral performance in the exam is to have a conversation or interaction[2] about a topic chosen by the interlocutor-evaluator.

Our experience, as language teachers and researchers has shown that immediate beginners are not yet equipped with the language proficiency necessary to maintain a five-minute conversation about a topic they might not be acquainted with. And that the descriptive scale designed to assess oral performance according to this genre in the exam, might not be the best fit to assess performance according to the immediate communicative needs of learners in the context of the host language. In this context, equipping the learners to perform in specific focused-tasks might be more urgent. In order to illustrate, we can think of the learners being able to communicate the symptoms of a disease to a health professional, as a task that is very urgent for learners in the context of the host language.

We understand that every proficiency test presents limitations and are designed with different objectives in mind, and that the descriptive scales provided by the tests are many times used to guide curriculum development and even to guide teachers' choices in language classrooms. Therefore, considering the latter, scales such as CELPE-BRAS and CEFR present a comprehensive description on how to assess performance, and might not be a straightforward source for teachers designing and implementing their assessments in the classroom setting of Portuguese as a host language, specially for beginner learners who need to fulfill their immediate communication needs, irrespective of making language mistakes.[3]

Having considered the different approaches to analyze oral performance in the field of Task-Based Language Teaching, the following section describes the steps involved in the data collection and instruments this article presents.

––––––––––––––––––

[2] Although the situation is characterized as a conversation or interaction, the test taker, many times, might describe or support a point of view, for instance, based on the material dictating the topic of the conversation. Therefore, we agree that different textual genres can be comprised (MARCUSCHI, 2005) in this situation of communication.

[3] We do not intend to criticize the descriptive scales of such tests, but only to point out that in the case of CELPE-BRAS, because the scale presented by the test was designed originally for academic purposes, it might not be adequate to apply to assess the performance of beginner speakers.

2. Method

In order to report the process of assessing students' oral performance in the context of PHL for beginners in Brazil, by adopting the measure of Outcome Achievement (based on FARIAS, 2018), a construct developed in the field of Task-Based Language Teaching that looks at the adequate performance in a task through a pragmatic perspective, in this section, first we describe the context of teaching at which the oral tasks were aimed and follow by presenting the steps taken to create the two scales of Outcome Achievement. Then, we present the procedures for the implementation of the measure to analyze these students' oral performances[4].

First of all, we present the context of teaching on which this study focuses, which comprises a group of students (17 in total) beginner speakers of Brazilian Portuguese, who were immigrants living in *Florianópolis – Santa Catarina* around the first semester of 2019 and who had just started taking language classes at an extension project of the *Universidade Federal de Santa Catarina* (UFSC), entitled *Projeto PLAM*[5].

This group of learners was composed of one Iranian, one Jordanian, two Venezuelans, and 13 Haitian immigrants. The group is far from being considered homogenous, once they vary in sex (10 men, 7 women), nationality (Iranian, Jordanian, Venezuelan, and Haitian), and mainly in age range (18 to 76 years old), schooling (a few learners have college degrees, while some others have not completed secondary school) and language backgrounds (while a few students were monolinguals, others spoke two or three languages)[6].

PLAM project offers classes once-a-week, for three hours, on Saturday mornings. The classes are free of charge for the students and count on the participation of volunteer teachers and teacher assistants. The project usually offers one class for immediate beginner speakers and one class for students who are higher-beginners to intermediate speakers of Portuguese.

The data collection reported in this article took place in two of those classes, in which students were presented to two cycles of tasks, as part of their regular language classes. Therefore, in the first class, the students participated in the class regularly and performed pedagogical pre-tasks designed to prepare students to perform the final target task (FIGURE 1). Not all the students present in the class accepted to participate in the study and some did not complete the activities in class,

_____

[4] It is important to notice that the methods presented here were part of a longer master's investigation conducted by the first author of this article (MARCELINO, 2020) and that only two of the tasks implemented in the original master's dissertation are reported here.

[5] This project is part of NUPLE (*Núcleo de Ensino e Pesquisa em Português Língua Estrangeira*) from UFSC. NUPLE is one of the branches of the Cátedra Sérgio Vieira de Mello, responsible for facilitating the partnership between research and extension projects of Brazilian public universities with the United Nations High Commissionaire for Refugees (UNHCR).

[6] The research project upon which this article is based was approved by the Ethics Committee of the university (CEP - USFC) under the CAAE number: 07060919.2.0000.0121, and under the Report number: 3166979.

which was a requirement to have their performance included in the analysis[7]. This explains why the number of participants is different in the two tasks, once data collection took place in a real classroom setting, the researchers had to rely on students' attendance and their agreement to participate in the lessons' activities and, finally, with students' willingness to record their answers.

The cycles of tasks to which the students were presented before the recording tasks were designed taking into consideration the students' goals for Portuguese language use in their daily-lives outside school[8] and aimed at preparing these students to perform adequately in the final target tasks.

The first task required students to identify three different images of the flu symptoms and record their answers to the task (with their phones or with an audio recorder provided by the researcher[9]). The rubrics of the task were: "*Daniel chegou no Brasil há uma semana e ainda não fala português. Ele está doente e você vai com ele ao posto de saúde. Explique para o enfermeiro os sintomas de Daniel.*"; as presented in FIGURE 1.

FIGURE 1 – Task 1 (MARCELINO, 2020).

The second task required students to choose an occupation in which they could work, and state previous working experiences that related to the occupation chosen. Once again the students were required to record their answers, this time, to the task: "*Grave uma mensagem de Whatsapp oferecendo seu trabalho e falando sobre suas experiências nessa profissão*", as presented in FIGURE 2.

_____

[7] Students had to complete the task in class to assure that they would not have help from outsiders or that they would not read their responses, instead of saying it spontaneously. Furthermore, the completion of the tasks in class assured that students would have access to an internet connection to send their recordings to the researcher or even have access to the recorders, in case the students did not have a mobile phone available to record their responses.

[8] A Needs Analyses was conducted before the design of the tasks, in order to unveil the situations of language use to which this group of learners was exposed the most. For the complete report of this process see MARCELINO, 2020.

[9] The first author of this article is the researcher who was present during the tasks' implementation and who organized and supervised data collection.

FIGURE 2 – Task 2 (MARCELINO, 2020).

In order to assess the students' performances in these two tasks, two descriptive scales were designed according to the measure of Outcome Achievement. The two scales followed the same rationale, considering the aspects that were important to achieve each task's communicative outcome. The communicative outcome of the first task was to communicate the symptoms of the flu to a health professional, while the communicative outcome of the second task was to choose an occupation to offer to a colleague/friend and state the previous working experiences associated with the occupation wanted. Thus, because the two tasks presented different characteristics, one of the features that composed the measure of Outcome Achievement was adapted to comprise the specificities of each task.

The first task presented a closed-outcome, meaning that learners' responses could be characterized as correct or incorrect based on the matching of their responses with the visual support provided. The second task presented an open outcome that required justification, meaning that students' responses could only be characterized as correct or incorrect depending on the justification provided to their choice for occupation. This characteristic impacted on one of the features of the Outcome Achievement measure, which had to be adapted accordingly. As it is possible to see in TABLE 1, while in the first task there is only one feature for Correctness, the second task is composed of Correctness A and Correctness B, because of this two-factor component of the answer.

I.	Communicative objective	It aimed at giving a general score for the performance, depending on whether the student achieved the task's communicative objective or not.
II.	Correctness (for Task 1) Correctness A Correctness B (for Task 2)	It was specific to each task and concerned whether the students' responses were correct, in Task 1, according to the images they should describe (Correctness), and in Task 2 if a student's alleged professional experience (Correctness B) matched with the position they mentioned they were searching (Correctness A).
III.	Communicative context	It comprised giving indications of the situation of interaction, the place, and the people involved. For example, the response should have contextual clues that the genre of the communication in task 2 – recoding a Whatsapp message to a colleague/friend – characterizes mainly as an informal situation, that implies more colloquial language use;
IV.	Coherence	It dealt with the level of connection of the ideas conveyed by the students, "whether they followed a well-structured sequence of events in their speech, for instance, by first introducing Daniel (the fictitious character in Task 1) and then stating his health problem and not the other way around" (MARCELINO, 2020, p. 59);
V.	Clarity	It comprised the transparency of the meanings being conveyed by the students.
VI	Prosody	In this analysis, it stands for the suprasegmental features of speech: intonation, rhythm, and speed.
VII.	Vocabulary	It aimed at analyzing if the words used in the speech were sufficient and adequate to achieve the Communicative outcome of the tasks.

TABLE 1 – Description of the features the composed the measure of Outcome Achievement.

To create the two scale, the teacher of the class, one volunteer teacher assistant of the same group, along with the first author of this article, analyzed the cycles of tasks to which the students would be presented before performing the task, once the cycle of tasks was designed to prepare the students to perform in these specific tasks. Then they analyzed and raised the key characteristics of each task, and finally, considering the students' profile – immigrant beginner learners of Brazilian Portuguese – they settled seven features (shown in TABLE 1) that should be taken into consideration for the analysis.

The features' descriptions seen in TABLE 1 were inspired by the studies of Farias (2014, 2018), Zaccaron (2017), Specht and D'Ely (2020), and Lima Terres, Torres, and Boeing Marcelino (2020), which have all devised descriptive scales for assessing English learners' performances in tasks in different modes. While Farias (2014, 2018) worked with written narratives, Specht and D'Ely (2020) worked with oral narratives. In turn, Zaccaron, Xhafaj, and D'Ely (2019), Zaccaron (2017) and Lima Terres, Torres, and Boeing Marcelino (2020) assessed oral short messages recorded through means of the Whatsapp mobile application.

In summary, in this study the Outcome Achievement measure comprised seven features implemented to analyze students' oral performance using Likert scales that attributed scores from 0 to 5 (in which 0 was the lowest score and 5 the highest). In the first task 15 learners participated by recording their response, while in the second task, only 12 learners participated. The Outcome Achievement of the learners' answers was analyzed by nine raters. Five of these raters were experienced Portuguese as Additional Language teachers with an average teaching experience of 6.8 years. These five raters had also had experience teaching other languages such as English, German, and Japanese. The other four raters had an average teaching experience of 9.5 years, teaching English as a Foreign Language

REVISTA DA **ABRALIN**

(FL). All of the nine raters had a Teaching and/or Bachelor's degree in *Letras* (English FL and/or Portuguese Mother Language (ML). All of them were native speakers of Brazilian Portuguese.

Each of the raters' analyses took about 2 hours to be completed. The sessions of analyses were accompanied by the first author of this article, who would explain the overall objectives of the study and present the raters with the procedures. The decision to have the researcher present throughout the sessions was made to assure that all raters would have the same or, at least, very similar conditions of evaluation.

First, raters read an instructional text in which concepts such as the Communicative Outcome of a task were explained. Then, the raters could read the descriptive scale of the tasks and read the cycles of tasks that were taught before the performance, followed by a review of the descriptive scale of each task[10]. Raters could solve with the researcher any questions they had concerning the procedures for their evaluation or the constructs involved. The researcher clarified that there was not a correct or incorrect evaluation of the performances and that each rater evaluation would not be judged or compared with other raters' evaluations[11]. Finally, when the rater had no further questions about the proceedings, the researcher would play the first recording twice, wait for the rater to provide scores and justification, and proceed to the following recording.

In summary, raters could listen twice to each recording and then give a score from 0 to 5 to each feature of Outcome Achievement. In addition to giving a score from 0 to 5, raters should also justify their choice for all the scores given to each student in each feature. FIGURE 3 illustrates the instructions and the descriptive scale of Task 1, to which the raters responded.

––––––––––––––––––

[10] All the texts presented to the raters were in a digital format. Raters provided their scores and justifications through an online formulary, which allowed for immediate storage of their evaluations.

[11] Although a few raters tried to discuss their evaluations with the researcher, she replied to them by saying that she was not allowed to share her evaluation of any of the students' performances in order not to influence their judgment of the performances.

Descritores de avaliação - *Outcome achievement* - Tarefa A: Problemas de saúde

Querido/a Avaliador/a, você irá ouvir os áudios com as respostas dos/as estudantes para a tarefa-alvo do segundo ciclo de tarefas da minha pesquisa (que para fins da sua avaliação será denominado, aqui, como Tarefa A). O objetivo desta tarefa-alvo é: *Comunicar dores e sintomas de uma doença a um profissional da saúde.* Seu papel como avaliador é mensurar cada performance estabelecendo uma pontuação de 0 a 5 para cada critério abaixo (sendo 0 a menor pontuação e 5 a maior pontuação).Finalmente, considerando o que as variáveis dos aspectos pragmático e de adequação do resultado comunicativo implicam, eu convido você a criar uma categoria adicional, caso você julgue necessário, ou seja, se você acredita que as categorias estabelecidas deixaram de apresentar algum aspecto importante, por favor, inclua-o abaixo, marque sua avaliação e me comunique.

Para ajudar você a se lembrar da tarefa final de cada participante enquanto você ouve os áudios, copiamos ela abaixo:

* Lembre-se que dizer o nome correto de cada sintoma era parte da tarefa dos estudantes, eles estão aqui pra que você tenha acesso às respostas corretas, enquanto ouve os áudios.

1D - Daniel chegou no Brasil há uma semana e ainda não fala português. Ele está doente e você vai com ele ao posto de saúde. Explique para o enfermeiro os sintomas de Daniel.

FEBRE   DOR DE CABEÇA   DOR DE GARGANTA

I.   Em termos gerais, o estudante alcançou o objetivo comunicativo da tarefa.
( )0   ( )1   ( )2   ( )3   ( )4   ( )5
Por que você escolheu essa pontuação?_____

II.   São mencionados no áudio os exatos sintomas que aparecem nas imagens.
( )0   ( )1   ( )2   ( )3   ( )4   ( )5
Por que você escolheu essa pontuação?_____

III.   O contexto comunicativo é levado em consideração na resposta. (pense na situação do posto de saúde: como se dá essa comunicação, quem são os sujeitos envolvidos, etc).
( )0   ( )1   ( )2   ( )3   ( )4   ( )5
Por que você escolheu essa pontuação?_____

IV.   A informação foi transmitida de forma coerente.
( )0   ( )1   ( )2   ( )3   ( )4   ( )5
Por que você escolheu essa pontuação?_____

V.   A informação foi transmitida de forma clara - Você pode compreendê-la sem fazer esforço.
( )0   ( )1   ( )2   ( )3   ( )4   ( )5
Por que você escolheu essa pontuação?_____

VI.   A entonação, ritmo e velocidade com que a informação foi transmitida são adequadas.
( )0   ( )1   ( )2   ( )3   ( )4   ( )5
Por que você escolheu essa pontuação?_____

VII.   O vocabulário usado é suficiente e adequado para cumprir a tarefa efetivamente.
( )0   ( )1   ( )2   ( )3   ( )4   ( )5
Por que você escolheu essa pontuação?_____

VIII.   Sua sugestão:_____.
( )0   ( )1   ( )2   ( )3   ( )4   ( )5
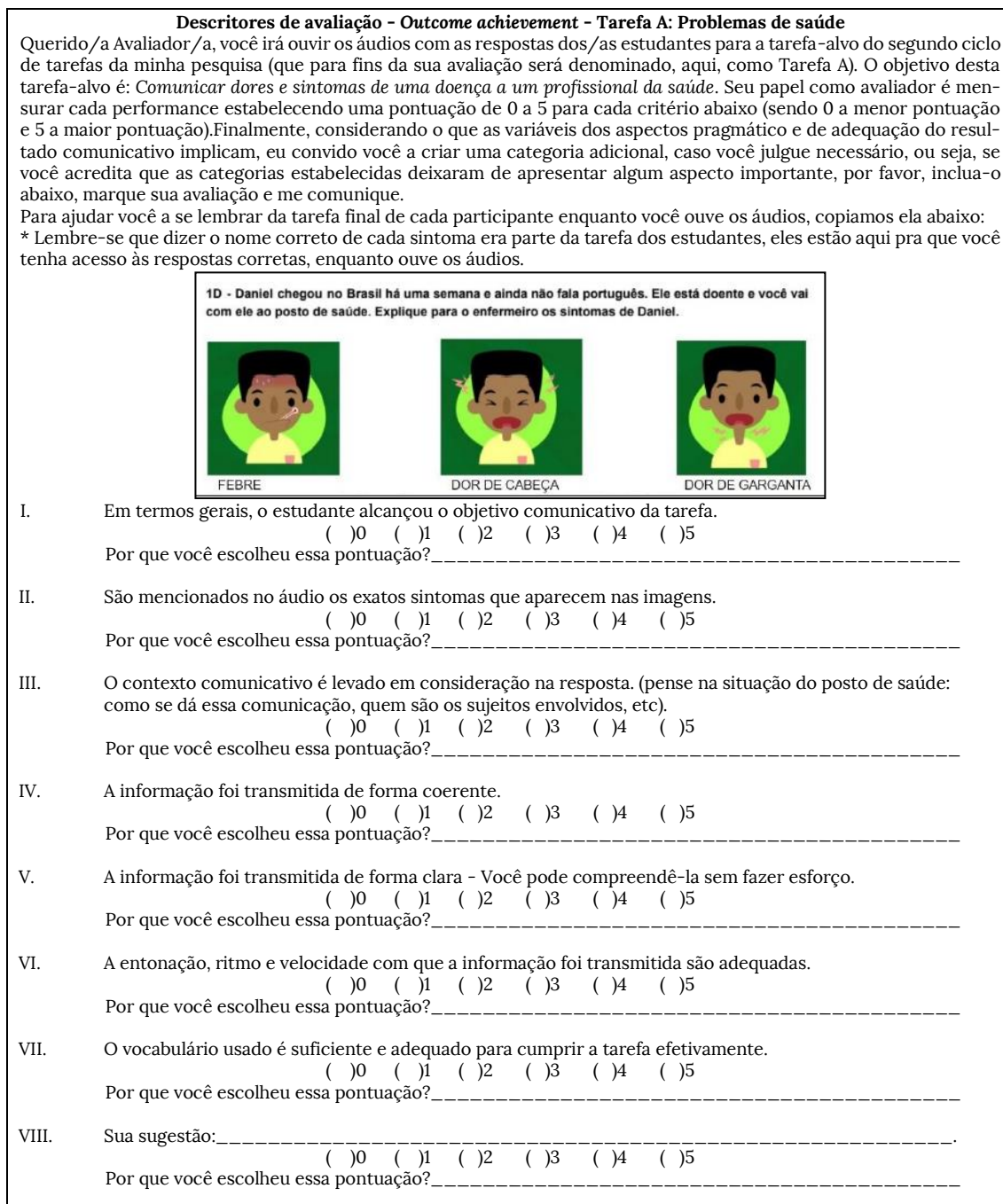Por que você escolheu essa pontuação?_____

FIGURE 3 – Descriptive scale of Outcome Achievement for Task 1 (MARCELINO, 2020).

As seen in FIGURE 3, raters had the option of creating one further feature for analyzing the students' performances; however, none of them suggested a topic of evaluation that was not already comprised by the predefined features presented.

After evaluations were completed, in order to check for interrater reliability, we ran a Cronbach's alpha test on SPSS (23.0). According to Taber (2018), reliability tests demonstrate "the

extent to which an instrument can be expected to give the same measured outcome when measurements are repeated" (p. 1274). Also according to him, many studies in science education have long adopted Cronbach's alpha value as an indication that instruments (usually scales and tests) are fit for their purpose. The tests ran with raters' raw data of Tasks 1 and 2 indicated a robust result once Cronbach's alpha value for Task 1 was .802 and for Task 2 was .822 as shown in FIGURE 4, assuming that the Cronbach's alpha benchmark traditionally used in the field is .70 – .80 (LARSON-HALL,2016). These results showed a strong correlation between raters' answers.

Task 1

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.803	.802	9

Task 2

Reliability Statistics

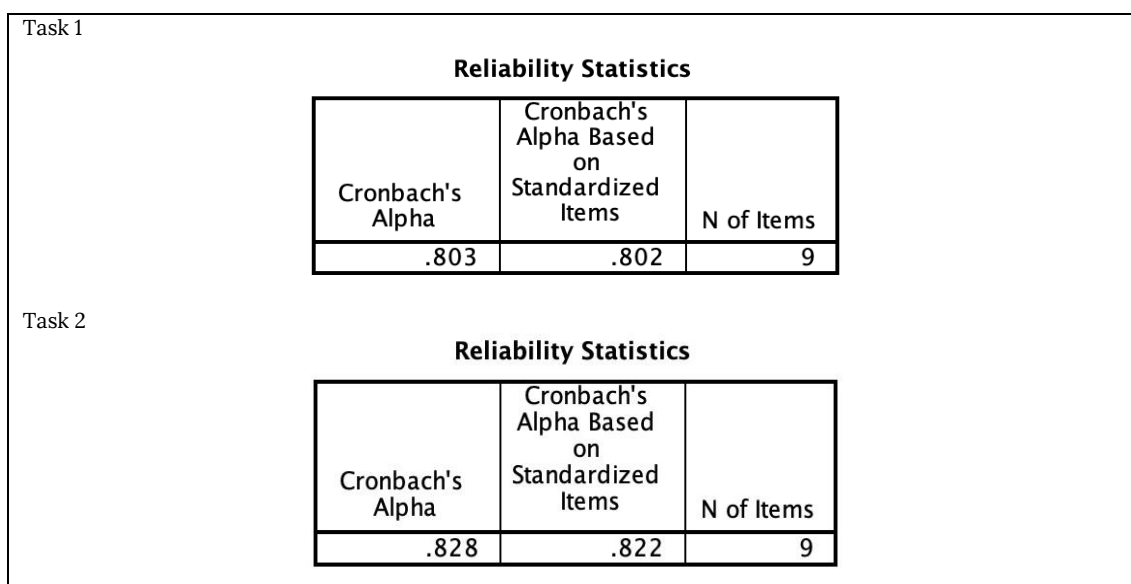| Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.828	.822	9

FIGURE 4 – Interrater reliability tests for Tasks 1 and 2 (MARCELINO, 2020).

Regarding the analysis of each learner's performance and the group mean values in each feature of Outcome Achievement, the procedures are explained in the following section, alongside the tables presenting all the results of the raters' assessment.

To conclude, the following section presents the results of the students' oral performances to tasks assessed through the Outcome Achievement measure scales, designed according to the characteristics of the tasks presented in FIGURES 1 and 2.

3. Results and discussion

This section presents the results from students' recorded answers to the two tasks, and discusses these results in light of previous studies, bringing possible impacts to understanding speaking performance in the Host Language context.

As mentioned previously, the students' oral performances were evaluated by nine raters according to the Outcome Achievement measure. In Task 1, 15 students recorded their answers.

In order to achieve the score for each learner in each feature, all the scores attributed by the nine raters to each of the seven features of Outcome Achievement were added and divided by nine, the results of this calculation are presented in TABLE 3. Then the 15 scores of all learners for each feature were added, and finally divided by 15, to achieve the group's mean value to each feature of Outcome Achievement, presented in TABLE 2 below.

Feature	N	M	Min	Max	SD	SEM
Communicative Objective	15	3,41	1,56	5,00	1,03	0,27
Correctness	15	3,62	0,67	5,00	1,46	0,38
Context	15	3,33	1,78	5,00	0,93	0,24
Coherence	15	3,42	1,67	4,89	1,00	0,26
Clarity	15	3,67	0,78	5,00	1,13	0,29
Prosody	15	3,79	1,44	4,67	0,85	0,22
Vocabulary	15	3,34	1,56	4,67	0,92	0,24

N: Number of participants; **M**: mean scores of all 15 participants attributed by the nine raters' mean score for each participant; **Min**: Minimum scores among the 15 participants; **Max**: Maximum scores among the 15 participants; **SD**: Standard Deviation of the dispersion of scores in comparison to the Mean scores; **SEM:** Standard Error of the Mean

TABLE 2 – Descriptive Statistics of group scores for performance in Task 1 (MARCELINO, 2020).

Considering that raters evaluated students' performance in a six-point scale, from 0 to 5, where 0 was the lowest and 5 the highest score that could be attributed by each rater, the group obtained, in Task 1, a mean score higher than 3 in all seven features of Outcome Achievement, as presented in TABLE 2. The Standard Deviation values show the dispersion of the scores compared to the Mean which indicates that students' scores fluctuated around the mean and are spread around diverse values (LARSON-HALL, 2016). This is mainly explained by the fact that the score range used to analyze the performance is small (0 to 5). The Standard Error of the Mean (SEM), also presented in the table, stands for the standard deviation of the sample mean and indicates that the Mean estimate is highly precise since according to Larson-Hall (2016) "the smaller the SE is, the more precise the estimate" (p. 84). In Task 1, the feature with the highest mean score was Prosody and the lowest mean score was Context (3,79 and 3,33 respectively), as shown in TABLE 2.

It is interesting to note that when comparing the minimum and maximum scores of the participants for Task 1, five out of the fifteen participants (namely, P1, P5, P6, P7, and P11 in TABLE 3) received maximum scores for some of the features, while only two out of fifteen (P9 and P15) shared the minimum scores. This means that the number of participants who did well in the task outnumbered considerably the participants who did not perform the task satisfactorily.

Participant 6 is one of the students who received the highest scores for Outcome Achievement; therefore, according to the raters, her performance is an example of what would be an adequate speech that aimed at communicating the flu symptoms to a health professional: "*Bom dia! Eu e-eu estou aqui porque meu filho Daniel está sentindo problemas de saúde. Ele tem dor de garganta, muita*

segreçã-segregação nasal e se siente muito cansado. Eu estoy muito assustada". This performance received the highest grade for the Context feature, taking into consideration that she even created a story to Daniel's character by saying he was her son. However, raters penalized her, mainly for her trouble in pronouncing the symptom "*secreção nasal*". It is interesting to trace a comparison between the performances of Participant 6 and Participant 5, because, among the learners who performed task 1, Participant 5 received the highest score for Outcome Achievement, even though one could argue that his answer was a lot simpler than Participant 6's answer for instance. Participant 5's answer was: "*O Daniel tem febre, dor de cabeça e dor de garganta*". This might be explained by the fact that, possibly, raters evaluated more highly performances that were free of errors, instead of performances that were more creative or detailed, as can be seen in TABLE 2 that Participant 5 received a total score of 33,67, while Participant 6 received a total score of 32,44.

Participant	Com. O.	Corr	Cont	Cohe	Clar	Pros	Voca	Total
P1	4,78	4,89	3,89	4,33	4,89	4,67	4,56	32,00
P2	3,78	3,56	3,44	3,67	4,00	4,00	3,56	26,00
P3	4,00	3,78	4,67	4,56	3,22	3,78	4,11	28,11
P4	3,33	4,44	2,67	3,22	3,89	3,78	3,22	24,56
P5	5,00	5,00	4,44	4,89	5,00	4,67	4,67	33,67
P6	4,56	4,33	5,00	4,78	4,67	4,67	4,44	32,44
P7	4,00	5,00	3,67	4,11	3,89	3,56	3,78	28,00
P8	3,33	4,56	2,56	3,33	4,22	3,89	3,00	24,89
P9	2,00	0,67	3,11	2,33	4,33	4,22	2,78	19,44
P10	3,11	3,67	2,89	2,89	3,11	2,56	2,89	21,11
P11	3,89	5,00	4,00	3,22	4,11	4,33	4,00	28,56
P12	2,00	1,33	2,44	1,78	1,78	3,44	2,22	15,00
P13	3,11	4,33	2,33	3,33	3,78	4,00	2,56	23,44
P14	2,78	1,78	3,11	3,22	3,44	3,89	2,78	21,00
P15	1,56	2,00	1,78	1,67	0,78	1,44	1,56	10,78
Com. O.: Communicative Objective. **Corr:** Correctness. **Cont:** Content. **Cohe:** Coherence. **Clar:** Clarity. **Pros:** Prosody. **Voca:** Vocabulary. **Total:** all features added.								

TABLE 3 - Students' individual scores by feature for Task 1 (MARCELINO, 2020).

In the opposite end of the scale was Participant 15, who received the lowest mean scores for most features and consequently the lowest score for Outcome Achievement. His answer was as follows: "*Daniel [inaudible] doctora. Febre, dor de garanta e se tá + é, ahm, + é segre-ção nasal*". This participant did not mention the correct symptoms and had great difficulty in communicating it clearly, which is reflected in his lowest score for Clarity (see TABLE 3). Raters justified their choices by saying that this performance was filled with pauses between sentences and in the middle of words, resulting as well in a bad score for Prosody, for instance. All these aspects hampered the comprehension of the entire message and resulted in a poorer outcome.

Moving on to analyzing performances in Task 2, this time only 12 students recorded their performances. As shown in TABLE 4, one more time, the mean scores of the entire group were higher

than 3 for most features of Outcome Achievement, except for the Correctness A feature. Standard Deviation values were low, which indicated a low variance of scores throughout the scale. The lowest and highest means were Correctness A and Prosody (2,97 and 3,44, respectively).

Feature	N	M	Min	Max	SD	SEM
Communicative Objective	12	3,30	1,67	4,56	0,83	0,24
Correctness A	12	2,97	0,67	4,67	1,10	0,32
Correctness B	12	3,26	1,44	4,56	0,95	0,27
Context	12	3,09	1,67	3,78	0,61	0,18
Coherence	12	3,27	1,56	4,11	0,85	0,25
Clarity	12	3,06	1,33	4,33	1,08	0,31
Prosody	12	3,44	2,11	4,22	0,78	0,23
Vocabulary	12	3,09	1,33	4,11	0,92	0,26

N: Number of participants; **M**: mean scores of all 13 participants attributed by the nine raters' mean score for each participant; **Min**: Minimum scores among the 13 participants; **Max**: Maximum scores among the 13 participants; **SD**: Standard Deviation of the dispersion of scores in comparison to the Mean scores; **SEM**: Standard Error of the Mean.

TABLE 4– Descriptive Statistics of group scores for task performance in TASK 2 (MARCELINO, 2020).

Task 2, as well as Task 1, presented a higher number of participants that received the maximum scores (P4, P5, P8, P9, and P12) in comparison with few participants sharing the minimum scores (P7 and P10, see TABLE 5).

Participants	Com. O.	CorrA	CorrB	Cont	Cohe	Clar	Pros	Voca	Total
P1	3,67	3,11	3,22	3,00	3,67	3,22	3,78	3,67	24,17
P2	3,67	4,22	3,89	3,22	3,67	4,22	4,11	3,67	26,61
P3	2,56	2,78	2,11	2,56	2,33	1,56	2,33	2,11	15,89
P4	3,67	4,67	3,33	3,11	3,67	3,78	3,89	3,33	25,44
P5	3,78	3,00	3,44	3,56	3,56	4,11	4,22	4,00	26,44
P6	3,56	1,67	3,89	3,11	4,00	2,33	3,56	3,56	22,89
P7	1,67	2,56	2,22	1,67	1,56	1,78	2,11	1,78	12,94
P8	4,56	4,11	4,56	3,67	4,11	3,67	3,89	4,11	28,33
P9	3,67	3,44	3,67	3,78	3,78	3,67	3,89	3,33	25,67
P10	2,00	0,67	1,44	3,44	1,89	1,33	2,44	1,33	13,50
P11	3,11	2,89	2,89	2,44	3,22	2,78	2,89	2,56	19,89
P12	3,67	2,56	4,44	3,56	3,78	4,33	4,22	3,67	26,72

Com. O.: Communicative Objective. **CorrA**: Correctness A. **CorrB**: Correctness B. **Cont**: Content. **Cohe**: Coherence. **Clar**: Clarity. **Pros**: Prosody. **Voca**: Vocabulary. **Total**: all features added.
Note: To calculate the Total, features of Correctness A and B were added and divided by two.

TABLE 5 - Students' individual scores by feature for Task 2 (MARCELINO, 2020).

Participant 8 received the best evaluation for Outcome Achievement. In his answer: "*Eu bom para vaga comércio porque minha experiências com comércio de três anos. É, eu, é, estudei Universidade em Língua e, e Comercial também. Eu gosto também desse trabalho*", raters justified their

scores by saying the student was successful in presenting an occupation clearly and in connecting it adequately to the previous working experience mentioned, which is reflected in his good scores for Correctness A and Correctness B. Although the raters said they needed to put some effort to understand his message, which is reflected in his low score for Clarity, the communicative objective of his message was attained almost completely if we look at the Communicative Objective feature of his performance.

In turn, although Participant 7 received the lowest score for Outcome Achievement, her performance in some ways is not comparable to the performance of the other participants. This is explained by the fact that, in class, she was aided by a teacher assistant and a colleague, to complete the recording. Therefore, although some raters considered that Participant 7 was able to pronounce some of the words spelled out by others, other raters considered that once she did not perform by herself she did not achieve the outcome of the task, resulting in the lowest score given to a student in this task.

The second lowest score was attributed to the performance of Participant 10, whose answer was as follows: "*Oi, ahm, como você, a, como você está? Tudo bem? Mi profi-mi profissão é cozinheiro. Eu tenho experiência de quatro anos, ok?*". According to the raters, although Participant 10 started his performance well by "setting the mood" and greeting the interlocutor, which gave him a good score for Context, they agreed that the learner had not achieved a clear communication of the occupation neither of the experience related to it, once he only mentions the time of his experience. This is reflected in the low scores attributed mainly to Clarity, and Correctness A and B.

Furthermore, once we analyzed students' performances and learned what raters considered and penalized the most, it is interesting to notice the features with the highest and lowest evaluations. We believe that fruitful discussions can be raised from the fact that, in Task 1, Context received the lowest mean score, considering the group's mean (3,33). We understand that this feature of the Outcome Achievement aimed at evaluating whether the situation of interaction was present in students' performances, meaning that the students' choices of language could provide contextual clues of the place where the interaction happened and who were the people involved, which is related to the pragmatic features of language (FIORIN, 2003), so important in this context of teaching and learning. Unsuccessful performances considering this feature might indicate that students were not made aware of the importance of context in the interaction, and how it may affect the language used to communicate their messages.

In Task 2, the feature that received the lowest mean score was Correctness A (2,79), and interestingly, it was not accompanied by Correctness B (3,26), once the two were expected to be connected. It is possible to suggest that students preferred to describe their work experiences in more detail, leading raters to judge the correspondence with the job wanted based on the experiences described. Most of the times, when the experience did not match the job chosen, or students described solely their past working experiences, it was Correctness A that was penalized, as is the case of Participant 3: "*É, eu sou, eu sou secretária al-[inaudible] alfânde-alfândega. Experiência em saúde, [inaudible] enfemela, hum de posto de saúde já. Ahm, dois, experiência em dois anos*". This performance

showed that the learner chose an occupation that would not be typically connected to her experience as a health professional.

On the other hand, looking at the feature that received the highest scores for the groups' mean in both Task 1 and Task 2, we see Prosody (scoring 3,79 and 3,50, respectively). Prosody was the given name for three suprasegmental features considering speech: intonation, rhythm, and speed. On one hand, it could be argued that the highest scores this feature received are connected to the fact that both tasks allowed students to achieve an adequate outcome by using short sentences or even isolated words, like in Task 1, in which considering a "real-world" situation of communication, arriving at a health center and naming isolated symptoms could lead to successful communication, considering its objectives. An interesting further analysis would consider investigating whether there is a correlation between Prosody as an Outcome Achievement feature and the measure of Fluency from the CALF measurements.

To conclude, we should remember that the nine raters, even being all from the same field, native speakers of Portuguese, and being language teachers, diverged in their evaluations. This fact might yield two possible discussions: first of all, the use of raters to validate pedagogical practices might not be adequate as well, since the process of learning is not determined by only one single performance, but is permeated through teaching and learning practices inside classrooms, of which outside evaluators might be unaware. Second, the fact that the components of Outcome Achievement are subjective according to each rater concept of successful communication is an issue that deserves further investigation.

Measures such as the one used in this article to assess oral performance have been just recently included in the field of performance analyses in TBLT and there are a growing number of studies that have chosen to take Outcome Achievement, or Adequacy, in contrast with other well-established measures such as CALF. We agree with Pallotti (2009) that Outcome Achievement, or Adequacy, "should be seen as both an independent construct based on task success and as a way of interpreting CAF measures" (p. 599), once, for example, Context is relevant for determining the adequate level of accuracy expected of a message.

In this article, it was seen that the differences between raters analyses of participants' performances were based on the different proceduralization of the Outcome Achievement measure, indicating that although the measure prioritizes the achievement of the communicative outcome of the tasks, raters' evaluation is influenced by their perception of fluency and accuracy, reflected in the scores attributed to the performances' lack of errors, smooth or intricate speech, for instance. This reinforces that this evaluation, as well as others, might be impacted by some degree of subjectivity. Although raters equally understood the features, they presented different evaluations of what was an adequate performance for each of them. This might indicate the importance of having dual-task research designs that comprise a cognitive approach and, as well, outcome fulfillment, as suggested by Skehan (2003). This dual-task design might illuminate the relationship between rater's subjective perceptions of fluency and accuracy, for instance, and allow for the correlation of their perceptions with measures that provide straightforward scores.

4. Final remarks

The growth in the field of teaching and learning of Portuguese as a Host Language, due to an increasing number of immigrants who find in Brazil a place to rebuild their lives, once many have been forcibly displaced from their previous homes, has made urgent the need to understand more deeply the specificities of this context (GROSSO, 2010; SILVEIRA; XHAFAJ, 2020, MARCELINO 2020). One of the specificities of this context is the inadequate use of traditional measures concerning oral performance. Mainly in the field of Task-Based Language Teaching the tradition to analyze oral performance has been to adopt the CALF measures (Complexity, Accuracy Lexical density, and Fluency), once these measures are believed to be part of and influence the communicative competence of speakers (FOSTER; SKEHAN, 1996; SKEHAN 1998; 2003). We know that the number of studies showing the efficacy of the CALF measures is a lot higher than the number of studies critiquing them. However, we reiterate that assessing only complexity, accuracy, lexical density and fluency might not be enough to determine if learners are being successful or not in their communication (PALLOTTI, 2009, REVESZ; EKIERT; TORGERSEN, 2016).

Moreover, considering the specific context of the Host Language addressed in this article, in which speakers are mostly focused on solving one specific problem of communication or using language adequately enough to "get one's message across", adopting the CALF measures to assess oral performance might be inadequate. In this context, this study aimed at presenting an alternative way of assessing oral performance in tasks that do not look at the CALF measures but that takes into consideration the Outcome Achievement measure, which is a multifaceted construct, that does not rely solely on vocabulary use, or grammatical accuracy, for instance, but instead has a strong concern with meaning and contextual elements, that may dictate the adequacy of the vocabulary or grammar to be used. To do that, 17 adult immigrants, beginner speakers of Brazilian Portuguese performed two oral tasks and had their performances evaluated by nine raters, native speakers of Brazilian Portuguese and experienced language teachers.

These students obtained different scores for Outcome Achievement, reflected in the varied scores attributed to each of the seven features that compose the measure. These seven features were thought to comprise the pragmatic aspects of language (FIORIN, 2003) involved in a situation of communication such as the ones presented in the two tasks investigated here. We believe that this article supports the idea that the process of understanding the communicative needs of the speakers, designing focused tasks according to their needs, and evaluating the speakers' performances according to the features of these specific tasks is very challenging and can be a complicating factor when a teacher in the classroom has to take decisions on his/her own about how to evaluate the students.

In this sense, the field of TBLT seems to shed light on what to prioritize in the moment of evaluation, and on what to prioritize during the preparation of these students to perform a task, once in the context of the Host Language especially, issues related to the pragmatic elements of language (FIORIN, 2003) are fundamental.

Acknowledgement

REFERENCES

COUNCIL OF EUROPE. *Common European Framework of Reference for Languages: Learning, teaching, assessment* – Companion volume, Council of Europe Publishing, Strasbourg, 2020.

CURSINO, C.; ALBUQUERQUE, J.; FIGUEIREDO SILVA, M. C.; GABRIEL, M.; ANUNCIAÇÃO, R. M. de. Português Brasileiro para Migração Humanitária (PBMIH): reflexões linguísticas e pedagógicas para o ensino de PLE em contexto de migração e refúgio. In: RUANO, B. P.; SANTOS, J. M. P.; SALTINI, L. M. L. (Orgs). *Cursos de português como língua estrangeira no Celin-UFPR: práticas docentes e experiências em sala de aula.* Curitiba: Ed. UFPR, 2016.

DEUSDARÁ, B.; ARANTES, P. C. C.; BRENNER, A. K. "É um problema de todo mundo": conceitos, métodos e práticas no ensino de português para refugiados. *Fórum Linguístico*, v.15, n. 3, p. 3226–3240, 2018. DOI http://dx.doi.org/10.5007/1984.

FARIAS, P. F. TASK-TEST: *What lies beyond implementing a Task-Based assessment: Comparing learners' performance and unveiling learners' perception in a testing situation.* 2014. Dissertação (Mestrado em Inglês: Estudos Linguísticos e Literários) - Centro de Comunicação e Expressão da Universidade Federal de Santa Catarina, Florianópolis, 2014.

FARIAS, P. F. *Critical EFL and Critical Literacy: the impacts of designing and implementing a cycle of tasks in a public school setting for Critical Language and Critical Literacy development.* 2018. Dissertação (Doutorado em Inglês: Estudos Linguísticos e Literários) - Centro de Comunicação e Expressão da Universidade Federal de Santa Catarina, Florianópolis, 2018.

FIORIN, J. L. Pragmática. In: FIORIN, J. L. *Introdução a Linguística II: Princípios de análise.* Editora Contexto, v. 02, 2003, p. 161-186.

FOSTER, P.; SKEHAN, P. The influence of planning and task type on second language performance. *Studies in Second Language Acquisition.* v 18, nº 3, p. 299-323, 1996. DOI http://www.jstor.org/stable/44487894.

GROSSO, M. J. dos R. Língua de acolhimento, língua de integração. *Horizontes de Linguística Aplicada*, v. 9, n. 2, p. 61–77, 2010. DOI https://doi.org/10.26512/rhla.v9i2.886.

JENSEN, J. A Problemática das Variações Sociolinguísticas no Ensino do Português como Língua Estrangeira (PLE). *Revista de Letras*, v. 1, n. 24, p. 12, 2002.

LARSON-HALL, J. *A Guide to Doing Statistics in Second Language Research Using SPSS and R*. 2. ed. New York: Routledge, 2016.

LIMA TERRES, M.; TORRES, M. C.; BOEING MARCELINO, A. F. O impacto do planejamento colaborativo de uma tarefa oral na performance de aprendizes no contexto de EFL. *Domínios de Lingu@gem*, p. 1-28, 13 jul. 2020.

LONG, M. *Second Language Acquisition and Task-based Language Teaching*. West Sussex: John Wiley and Sons, 2015.

LOPES, J. H. "Materiais didáticos de Português para Falantes de Outras Línguas: do levantamento de produções brasileiras a uma nova proposta". *In*: FURTOSO, V. Aparecida Bagio. (Ed.). *Formação de Professores de português para falantes de outras línguas*. Londrina: EDUEL, 2009. p. 127–156.

LOPEZ, A. P. A. *Subsídios para o planejamento de cursos de Português como Língua de Acolhimento para imigrantes deslocados forçados no Brasil*. Dissertação (Mestrado em Linguística Aplicada) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2016.

LOPEZ, A. P. de A. Portuguese as a Welcoming Language for forcibly displaced immigrants in Brazil: some principles for teaching in light of Interculturality. *Revista Brasileira de Linguística Aplicada*, v. 18, n. 2, p. 389–416, jun. 2018. DOI http://dx.doi.org/10.1590/1984-6398201812048.

MARCELINO, A. F. B. *TBLT and Portuguese as a Host language: analyzing learners' oral performance in terms of Outcome Achievement and investigating the task implementation process through the learners' and the teacher's perspectives*. 2020. Dissertação (Mestrado em Inglês: Estudos Linguísticos e Literários) - Centro de Comunicação e Expressão da Universidade Federal de Santa Catarina, Florianópolis, 2020.

MARCUSCHI, L. A. Gêneros textuais: configuração, dinamicidade e circulação. In: KARWOSKI, A. M. (Org.). *Gêneros textuais: reflexões e ensino*. Palmas e União da Vitória, PR: Kaygangue, 2005.

MINISTÉRIO DA EDUCAÇÃO. *Certificado de Proficiência em Língua Portuguesa para Estrangeiros (Celpe-Bras)*. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Disponível em: https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/celpe-bras. Acesso em: 30 outubro 2020.

OLIVEIRA, G. M. de; SILVA, J. I. da. Quando barreiras linguísticas geram violação de direitos humanos : que políticas linguísticas o Estado brasileiro tem adotado para garantir o acesso dos imigrantes a serviços públicos básicos? *Gragoatá*, v. 22, n. 42, p. 131–153, 2017. DOI dx.doi.org/10.22409/gragoata.2017n42a909.

PALLOTTI, G. CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, v. 30, n. 4, p. 590–601, 1 dez 2009. DOI 10.1093/applin/amp045.

REVESZ, A.; EKIERT, M.; TORGERSEN, E. The Effects of Complexity, Accuracy, and Fluency on Communicative Adequacy in Oral Task Performance. *Applied Linguistics*, v. 37, n. 6, p. 828–848, 2016. DOI 10.1093/applin/amu069.

ROBINSON, P. "Task complexity, cognitive load, and syllabus design". *In*: ROBINSON, Peter. (Ed.). *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press, 2001, p. 287-318. DOI doi:10.1017/CBO9781139524780.012

SILVEIRA, R.; XHAFAJ, D. C. P. The use of tasks in the teaching of Portuguese as a Second Language. *Linguagem e Ensino*. v 23, n. 2. Universidade Federal de Pelotas, 2020. DOI https://doi.org/10.15210/rle.v23i2.17427.

SKEHAN, P. A Framework for the Implementation of Task-Based Instruction. *Applied*

Linguistics, n. 17, p. 38–62, 1996. DOI https://doi.org/10.1093/applin/17.1.38.

SKEHAN, P. A *Cognitive Approach to Language Learning*. Oxford: Oxford University Press, 1998.

SKEHAN, P. Review article Task-based instruction. *Language and Technology*, v. 36, p.1–14, 2003. DOI 10.1017/S026144480200188X.

SPECHT, A.; D'ELY, R. C. de S. F. Enhancing strategic planning through Strategy instruction: the effect of two types of Strategy instruction on learners' oral planned performance**.** *Ilha do Desterro*. v. 73, nº 1, p. 043-061, Florianópolis, jan/abr 2020. DOI: http://dx.doi.org/10.5007/2175-8026.2020v73n1p43.

TABER, K. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, v. 48, n. 6, p. 1273–1296, 2018. DOI 10.1007/s11165-016-9602-2.

UNHCR. *Figures at a glance*. Disponível em: http://www.unhcr.org/en-us/figures-at-aglance.html. Acesso em: 19 nov 2018.

ZACCARON, R.; XHAFAJ, D. C. P.; D'ELY, R. C. de S. F. Só mais um minutinho, teacher": planejamento estratégico colaborativo e individual para tarefas orais em l2 em uma escola pública. *Ilha do Desterro*. v. 72, nº 3, p. 401-425, Florianópolis, set/dez 2019. DOI http://dx.doi.org/10.5007/2175-8026.2019v72n3p401.

ZACCARON, R. *The More The Merrier (?): the impact of Individual and Collaborative Strategic Planning on Performance of an Oral Task by Young Learners of English as an L2 in Brazil*. Dissertação (Mestrado em Inglês: Estudos Linguísticos e Literários) - Centro de Comunicação e Expressão da Universidade Federal de Santa Catarina, Florianópolis, 2017.

Appendix 1 - Transcription of Students' performances in Task 1.

Speakers' identification (RS: Researcher, Participants are referred by P and the sequence number, TA: Teacher Assistant, I: Interpreter).

Conventions: (+: short pause, ++: long pause, [ ]: overlapping speech, (( )): analyst's comments, XXX: inaudible, ?: questioning intonation, CAPITAL: stressed word, { }: translated answers, ( ): translator's comments). Based on: GREGGIO; GIL, 2007, and SULLIVAN, 2000.

P1: Está com febre, dor de cabeça e ele tem de segre-segreç-segreção nasal.
P2: Está com febre, tosse e dor de garganta.
P3: Oi, doutor. Ele está com febre, calição, febre-ca-cabeça, secreção nasal.
P4: Dor de cabeça, dor de garganta, febre.
P5: O Daniel tem febre, dor de cabeça, dor de garganta.
P6: Bom dia! Eu e-eu estou aqui porque meu filho Daniel está sentindo problemas de saúde. Ele tem dor de garganta, muita segreçã-segregação nasal e se siente muito cansado. Eu estou muy assustada.
P7: Eu tosse.
      TA: Está com
          Daniel está tosse, + secreção de na-secreção nasal, + cansa-cansaço.
P8: Tosse, secreção nasal, cansação.
P9: Da-daniela? Daniel. Daniel está doente. Ele está, é, com gripe.
P10: Boa tarde, boa tarde. Olá, sim-sim-sim. Primeira ser febre, sec-febre. Sec-segunda: se-dor de cabeça. Terceira: espiracio-espiramação + D'*accord*? Ta bém.
P11: Oi Daniel, tudo bem? Ahm, ela dor de cabeça, ela, ela dor de garganta, ela dois tá febre.
P12: Oi, Daniel. Tô com febre, dor de barriga, gravida, vomisto, tá?
P13: Re, dor de cabeça e dor de garganta.
P14: É, Daniel diz doente. Tem pobrema de saúde, gripe. Tem sistoma tusse.
P15: Daniel XXX doctora. Febre, dor de garanta e se tá + é, ahm, + é segre-ção nasal.

Source: MARCELINO, 2020.

Appendix 2 - Transcription of Students' performances in Task 2.

Speakers' identification (RS: Researcher, Participants are referred by P and the sequence number, TA: Teacher Assistant, I: Interpreter). **Conventions:** (+: short pause, ++: long pause, [ ]: overlapping speech, (( )): analyst's comments, XXX: inaudible, ?: questioning intonation, CAPITAL: stressed word, { }: translated answers, ( ): translator's comments). Based on: GREGGIO; GIL, 2007, and SULLIVAN, 2000.

P1: Eu sou uma boa pessoa para essa vaga de vendedor porque eu tenho mais de dois anos de experiência e eu também, eu tenho habilidade na computa-computador. Eu + te, eu já ajudei a fazer excursão na escola primária também.

P2: Profissão é, eu sou cozinheiro. Eu tenho sete meses. E minha experiência eu, eu já fui trabalhar com serviço de limpeza de quinze meses.

P3: É, eu sou, eu sou secretária al-e-XXX alfânde-alfândega. Experiência em saúde, XXX enfemela, hum de posto de saúde já. Ahm, dois, experiência em dois anos.

P4: Profissão: gerente de loja. Experiência: eu trabalho em uma abrícia de turismo por um anos. Numa empresa de emprego em dois anos.

P5: Minha profissão é serviço de animação. Como experiência já trabalhava numa empresa de gráfico quase duas ano.

P6: Eu sou contadora. Ahm, eu sou egreçada. Eu estava na Universidade. Ahm, eu estou, a, eu estou. Eu sou boa para essa vaga. Ahm, porque eu fui, e-eu fui formada em Contadoria, eu traba-trabalhei em la Universidad como professora. Ahm, eu + fiz Contadoria, Con-contabilidade. Aula contabilidade conta?   Ahm, eu trabalhei na Universidade diez añ-años. E depois eu trabalhei autônoma em, em Contabilidade Tributária.

P7: Eu sou...
    TA: Vendedora.
      Eu sou vendedora.
  P2: No Haiti.
      No Haiti. Trabalho no Haiti.
  P2: Já fui trabalhando um ano
      Já fui trabalhando um anos.
  P2: Minha experiência.
      Minha experiência.
  P2: No Haiti.
      No Haiti. Vendedora.
   TA: [Vendedora.] Quanto tempo você trabalhou como vendedora? No Haiti, quantos anos?
  P2: Años? Annes?
   TA: Quanto?
    Um ano.
   TA: Um a-um ano? Então você coloca aqui na experiência: um ano de experiência.

P8: Eu bom para vaga comércio porque minha experiências com comércio de três anos. É, eu, é, estudei Universidade em Língua e, e Comercial também. Eu gosto também desse trabalho.

P9: Eu já trabalhei professor de Física. Se tem uma aula, eu tenho um anos de experiência. Eu gosto muito profissão também.

P10: Oi, ahm, como você, a, como você está? Tudo bem? Mi profi-mi profissão é cozinheiro. Eu tenho experiência de quatro anos, ok?

P11: Eu sou cozinheira.
    RS: Aham.
      Cozinheira há três anos.
    RS: Aham, aonde?
      Aonde? três anos.
    RS: No Brasil ou no Haiti?
    TA: No Haiti.
      Si, ahm, no Haiti.
    RS: No Haiti.
      No Haiti, sim!
    RS: [Ok.]

P12: Eu já traba motorista e mecânico. Eu tenho dezenove años experiência de motorista e seis años meca, de experiência de mecânico também.

Source: MARCELINO, 2020.

Appendix 3 - Descriptive scale task 2.

Descritores de avaliação - *Outcome achievement* - Tarefa B: Procurando trabalho

Querido/a Avaliador/a, você irá ouvir os áudios com as respostas dos/as estudantes para a tarefa-alvo do terceiro ciclo de tarefas que compõem a minha pesquisa (que para fins da sua avaliação será denominado, aqui, como Tarefa B). O objetivo desta tarefa-alvo é: *Enviar uma mensagem de áudio para um colega/conhecido oferecendo seus serviços em um emprego de sua escolha, falando sobre suas experiências no ramo*. Seu papel como avaliador é mensurar cada performance estabelecendo uma pontuação de 0 a 5 para cada critério abaixo (sendo 0 a menor pontuação e 5 a maior pontuação). Finalmente, considerando o que as variáveis dos aspectos pragmático e de adequação do resultado comunicativo implicam, eu convido você a criar uma categoria adicional, caso você julgue necessário, ou seja, se você acredita que as categorias estabelecidas deixaram de apresentar algum aspecto importante, por favor, inclua-o abaixo, marque sua avaliação e me comunique.

Para ajudar você a se lembrar da tarefa final de cada participante enquanto você ouve os áudios, copiamos ela abaixo:
* Lembre-se que essa tarefa é aberta e podem haver diversas respostas diferentes. Não havia um modelo de resposta que os estudantes devessem/pudessem seguir.

> **4A -** Você está procurando um emprego e pede ajuda a suas professoras e colegas para encontrar uma vaga. Grave uma mensagem de WhatsApp oferecendo seu trabalho e falando sobre suas experiências nessa profissão. Use o quadro abaixo para organizar sua mensagem.
>
> **Profissão:**
> **Experiências:**
> _____
> _____

I.   Em termos gerais, o estudante alcançou o objetivo comunicativo da tarefa.
     ( )0   ( )1   ( )2   ( )3   ( )4   ( )5
     Por que você escolheu essa pontuação?_____

II.   A - Compreende-se claramente a profissão que o "candidato" procura.
     ( )0   ( )1   ( )2   ( )3   ( )4   ( )5
     Por que você escolheu essa pontuação?_____

     B - Compreende-se claramente as experiências de trabalho anteriores do "candidato" e sua relação com a vaga buscada atualmente.

     ( )0   ( )1   ( )2   ( )3   ( )4   ( )5
     Por que você escolheu essa pontuação?_____

III.   O contexto comunicativo é levado em consideração na resposta (pense no gênero da mensagem de Whatsapp: como se dá essa comunicação, quem são os sujeitos envolvidos, etc).
     ( )0   ( )1   ( )2   ( )3   ( )4   ( )5
     Por que você escolheu essa pontuação?_____

IV.   A informação foi transmitida de forma coerente.
     ( )0   ( )1   ( )2   ( )3   ( )4   ( )5
     Por que você escolheu essa pontuação?_____

V.   A informação foi transmitida de forma clara - Você pode compreendê-la sem fazer esforço.
     ( )0   ( )1   ( )2   ( )3   ( )4   ( )5
     Por que você escolheu essa pontuação?_____

VI.   A entonação, ritmo e velocidade com que a informação foi transmitida são adequadas
     ( )0   ( )1   ( )2   ( )3   ( )4   ( )5
     Por que você escolheu essa pontuação?_____

VII.   O vocabulário usado é suficiente e adequado para cumprir a tarefa efetivamente.
     ( )0   ( )1   ( )2   ( )3   ( )4   ( )5
     Por que você escolheu essa pontuação?_____

VIII.   Sua sugestão:_____.
     ( )0   ( )1   ( )2   ( )3   ( )4   ( )5
     Por que você escolheu essa pontuação?_____

Source: MARCELINO, 2020.

Appendix 4 - Instructions for Raters evaluation of responses to Tasks 1 and 2.

Querido/a Avaliador/a,
Primeiramente, muitíssimo obrigada pela gentileza de tirar um tempo pra me ajudar nessa tarefa!!!

O principal foco do meu estudo é entender o impacto e as implicações da implementação de ciclos de tarefa em aulas de Português como Língua de Acolhimento para imigrantes adultos que frequentam o projeto PLAM. Para tanto, dois, dos quatro ciclos de tarefas implementadas tiveram como tarefa-alvo a gravação de uma mensagem de voz, cada uma referente a um ciclo de tarefas diferente. O primeiro com o tema "Problemas de saúde" e o segundo com o tema "Procurando trabalho". No campo de estudos sobre o ensino de línguas, atividades orais podem ser avaliadas de diversas maneiras, assim como em outros trabalhos conduzidos em salas de aulas intactas, no meu estudo eu escolhi avaliar a produção oral dos estudantes somente com relação a quão adequado foi seu **resultado comunicativo**. Na área de Tarefas, em inglês, se dá o nome a esse resultado de *communicative outcome*, o que NÃO confere a esse resultado simplesmente uma característica de resposta ou solução a um problema proposto, mas leva em consideração, e isso é central, a função COMUNICATIVA dessa resposta, e sua adequação tanto ao contexto em que ela foi abordada, quanto às características exigidas em cada contexto. Ellis (2003), um dos principais especialistas na área de tarefas, define *outcome* como o objetivo principal de uma tarefa, ou seja, o que se espera que os aprendizes façam (com a língua) ao final da tarefa. A realização desse resultado comunicativo abarca o fato de que, possivelmente, muitas vezes, independente do uso inadequado de estruturas gramaticais aqui e ali, se o estudante é capaz de realizar a tarefa, seu resultado comunicativo (*communicative outcome*) pode ser atingido, considerando que a tarefa possui uma natureza voltada, principalmente, para o significado.
Sua tarefa agora é ouvir os áudios dos participantes da minha pesquisa e selecionar na tabela descritiva uma pontuação à cada descritor, levando em consideração aspectos pragmáticos do uso da língua, que sejam adequados ao gênero discursivo de cada tarefa.
Pra você entender o que era esperado dos estudantes, é muito importante que você entenda como as gravações foram coletadas. Cada gravação aconteceu ao final dos seus respectivos ciclos de tarefas, que tinham o objetivo de preparar os estudantes linguisticamente para realizar a tarefa oral adequadamente. Como a atividade de gravação aconteceu dentro de uma sala de aula intacta, ambas gravações sempre acabaram acontecendo nos últimos minutos das aulas. A professora explicava a atividade, os estudantes tinham alguns minutos para se preparar, fazer anotações **se quisessem** por exemplo, e depois gravavam suas respostas em seus próprios celulares ou em gravadores com o auxílio das professoras. Eles não tinham tempo mínimo nem máximo de gravação.
Então, agora que já estabelecemos seu papel nessa pesquisa e você pôde entender um pouco melhor sobre como o estudo foi conduzido, é hora de botar a mão na massa! Pra você se familiarizar com os instrumentos usados nessa pesquisa, você receberá acesso aos **ciclos de tarefas** implementados, assim como às **mensagens orais gravadas** pelos estudantes. E, finalmente, você acessará as duas tabelas onde deverá marcar sua avaliação.
De maneira a "guiar" sua preparação para a avaliação, você pode seguir os seguintes passos:

1. Leia a tarefa A e pense sobre o que está sendo proposto, para que você entenda o que era esperado dos estudantes.

2. **Faça anotações sobre dúvidas** e/ou perguntas que você possa ter sobre o ciclo e sobre a tarefa-alvo e solucione-as com a pesquisadora **antes de iniciar** a avaliação.

3. Leia a tabela de critérios estabelecidos para avaliar as tarefas orais.

4. Agora, abra o questionário de avaliação. Ouça a resposta do participante nº 1 e estabeleça sua avaliação para cada critério.

5. Depois de atribuir a pontuação, por favor, inclua os comentários que julgar necessário para justificar sua escolha

6. Repita os procedimentos **4)** e **5)** com os demais áudios.

7. Evite comparar a performance dos estudantes.

8. Quando você terminar com todos os áudios salve a tabela de avaliação e envie-a para a pesquisadora por e-mail.

9. Repita o mesmo processo e passo-a-passo com a tarefa B.

Se você tiver qualquer dúvida, por favor, me avise.
Mais uma vez, MUITÍSSIMO OBRIGADA!

Source: MARCELINO, 2020.