

UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE LETRAS
DEPARTAMENTO DE LÍNGUAS ESTRANGEIRAS E TRADUÇÃO

THE USE OF L1 OR L2 IN THE QUESTIONS OF HIGH-STAKES ENGLISH READING TESTS: LITERATURE REVIEW AND AN EMPIRICAL STUDY

FERNANDO BORGES FURLAN

Brasília - DF
2025

FERNANDO BORGES FURLAN

THE USE OF L1 OR L2 IN THE QUESTIONS OF HIGH-STAKES ENGLISH READING TESTS: LITERATURE REVIEW AND AN EMPIRICAL STUDY

Undergraduate thesis presented to Universidade de Brasília, as an English Teaching course requisite for graduation.

Orientator: Dr. Gladys Quevedo Camargo

Brasília - DF

2025

ABSTRACT

In the interest of aiding the decision-making process when creating English reading exams for a Brazilian audience, this study undertakes two core tasks: First, explain the possible areas where the findings could have implications and review literature previously written on the subject. Second, verify differences in the performances of undergraduate students from English courses in a Brazilian university when taking the same English reading test, once with the items in Portuguese (L1), and once with the items in English (L2), as well as appraising the students' sentiments and opinions about this topic. The experiment had a sample size of 61 students, mostly taking English Language, Teaching, or Translation courses. The initial hypothesis, that there would not be large differences in performance between tests, has been supported by the data. The students seem to view the test with items in Portuguese as easier and less complicated, and the one with items in English as more serious and professional. In the Brazilian context, participants cited accessibility as a reason to keep high-stakes English test items, particularly ENEM's, in Portuguese.

Keywords: reading assessment, ESL, multiple-choice items, high-stakes tests, mother tongue

RESUMO

Com o objetivo de auxiliar o processo de tomada de decisão na criação de provas de leitura em inglês para o público brasileiro, este estudo realiza duas tarefas principais: primeiro, explica as possíveis áreas onde as descobertas podem ter implicações e revisa a literatura já escrita sobre o assunto. Segundo, verifica as diferenças no desempenho de alunos da UnB de diferentes cursos de inglês ao realizarem a mesma prova de leitura em inglês, uma vez com os itens em português (L1) e outra com os itens em inglês (L2), além de avaliar os sentimentos e opiniões dos alunos sobre o tema. O experimento teve uma amostra de 61 alunos, a maioria cursando Bacharelado, Licenciatura ou Tradução em Letras - Inglês. A hipótese inicial, de que não haveria grandes diferenças de desempenho entre as provas, foi corroborada pelos dados. Os alunos parecem considerar a prova com itens em português mais fácil e menos complicada, e a com itens em inglês mais séria e profissional. No contexto Brasileiro, os participantes citaram acessibilidade como um motivo para manter os itens de provas de inglês de alto impacto, como o ENEM, em português.

Palavras-chave: avaliação de leitura, itens de múltipla escolha, ESL, provas de alto impacto, língua nativa

LIST OF TABLES AND GRAPHS

Table 1 - Competencies from the ENEM's framework of reference.....7
Table 2 - Word bank to describe each version of the test.....11
Table 3 - Difficulty indexes of an item.....12
Table 4 - Discrimination of an item.....12
Table 5 - Demographics of the study.....14
Table 6 - Difficulty and Discrimination of the items in the EN1 test.....16
Table 7 - Development of respondents with different scores.....17
Table 8 - Difficulty and Discrimination of the items in the PT1 test.....18
Table 9 - Development of respondents with different scores.....18
Graph 1 - Descriptors chosen by the EN1 group.....21
Graph 2 - Descriptors chosen by the PT1 group.....21
Graph 3 - Aggregated descriptors of the tests.....23
Table 10 - Preferences about test and opinions about admission exams.....24
Table 11 - Comments left by participants.....25

ABBREVIATIONS AND ACRONYMS

CEFR – Common European Framework of Reference
CELTA – Certificate in Teaching English to Speakers of Other Languages
DIF – Differential Item Functioning
EN1 – The group of participants who took the test with items in English first
ENEM – Exame Nacional do Ensino Médio (National High School Exam)
FUVEST – Fundação Universitária para o Vestibular
FURG – Universidade Federal do Rio Grande
L1 – First language (a person's native language)
L2 – Second language
PT1 – The group of participants who took the test with items in Portuguese first
PTE – Pearson Test of English
UnB – Universidade de Brasília
USP – Universidade de São Paulo

SUMMARY

1 INTRODUCTION..... 5
2 THEORETICAL FOUNDATION..... 6
2.1 FUNDAMENTAL CONCEPTS ABOUT READING ASSESSMENT..... 6
2.2 ENEM..... 7
2.2 HIGHER-LEVEL ENGLISH TESTS..... 8
2.3 PREVIOUS STUDIES..... 9
3 METHODOLOGY..... 10
4 RESULTS & ANALYSES..... 14
4.1 DEMOGRAPHICS..... 14
4.2 EN1 ANALYSIS..... 16
4.3 PT1 ANALYSIS..... 17
4.4 COMPARATIVE ANALYSIS..... 19
4.4.1 PERFORMANCE AND DEVELOPMENT..... 19
4.4.2 PARTICIPANT PERCEPTIONS..... 20
4.4.2.1 DESCRIPTORS..... 20
4.4.2.2 PREFERENCES AND OPINIONS..... 23
5 CONCLUSIONS AND FINAL CONSIDERATIONS..... 27
REFERENCES..... 29
APPENDIX..... 32

1 INTRODUCTION

In the name of fairness, high-stakes tests designers must adopt the best practices known when it comes to creating and applying tests. However, there are some areas in which the course of action remains unclear. That is the case with the items in language tests whose primary purposes are not proficiency evaluation. The question is: should the items be in the target language, or in the test-takers' native language? In particular, this study observes differences between items in English (L2) and Brazilian Portuguese (L1). Many admission tests for higher education in Brazil, including the nation-wide ENEM (Exame Nacional para o Ensino Médio), feature one or more texts in English and items about them in Portuguese.

With a firm belief in the importance of the English language as a tool of emancipation and expansion of one's perspectives, the study at hand proposes to contribute to the debate, summarizing relevant discussions around the nature of assessment, validity, purposes of reading, and the role of English tests in their different contexts, and providing an empirical study to investigate the following questions:

1. How does the language (L1 vs English) used in the questions and items of reading comprehension English tests affect test-takers' performance?
2. What do test-takers with an intermediate to advanced level of English feel about the language used in the items when they take a test?

To these means, UnB (Universidade de Brasília) students were invited to take the same English reading comprehension test twice, performed consecutively, the only difference being the items written in L1 or L2, and to give their own opinions on the subject.

2 THEORETICAL FOUNDATION

In this section, fundamental concepts pertinent to the understanding and implications of the study are described. ENEM's situation is addressed in detail because of its high importance. Other relevant English tests in Brazil with items in Portuguese are briefly cited, and a review of previous studies is made.

2.1 FUNDAMENTAL CONCEPTS ABOUT READING ASSESSMENT

More than anything, this study is motivated by assessment validity. "Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (American Educational Research Association et al., 2014, p. 11). The author uses the term "interpretations of test scores" so as to make a clear distinction with the proposed use of the test. In other words, all tests measure something, however, whether any one test is adequate depends on what the test was meant to measure. In the context of this study, what is meant to be measured is an ability: reading comprehension in L2, and the question is to what extent the application of L1 or L2 in the questions (items) is aligned with this ability.

Reading comprehension, however, is still a broad term. Grabe and Stoller (2019) list different purposes for reading, such as to find specific information or understand the general idea. The decision-making supported by the test, as well as its length, usually determine what purposes for reading are demanded by its questions. Grabe (2009) cites time constraints, ascertainment of fairness and usability as limitations of reading assessment. In the Methodology part, the purposes for reading demanded by the study's test will be explored.

When speaking of validity, a concept that is important to take into account is differential item functioning (henceforth DIF). DIF is the propensity of an item to measure anything other than what it is meant to purposefully measure, and by doing so, favor or disfavor a certain group of individuals. (Zumbo 2007 apud Zhu; Aryadoust, 2020). Test-takers with identical ability might get different results which stem from aspects such as age, gender, socioeconomic status, first language, among others. Because of the smaller scope of this study, it does not make use of DIF identification techniques, but it urges researchers with more resources and time to conduct more robust research and employ it in their analyses.

2.2 ENEM

DIF is particularly a reason for concern in high-stakes tests. For this reason, while perhaps the test used in this study's experiment is unlike ENEM in terms of L2 proficiency, the enormous importance of this exam for Brazilian students justifies addressing its situation in detail.

The formulation of ENEM's language questions has been a target of criticism (Rauber, 2012; Araújo, 2017; Maciel, 2025). Moreover, with only 5 dedicated questions among its 180, and the option to forego English entirely in favor of Spanish, it reflects Brazil's difficulty in helping its students learn English through public education (Siqueira; Anjos, 2011).

Oliveira & Bueno (2021) have highlighted Brazil's long-standing status of low proficiency in English in a global ranking. While exploring the myriad of factors that cause this to be the case is outside the scope of this research, it is worth noting that ENEM's most recent framework of reference (matriz de referência), from 2009, only presents 4 competencies directly related to second language usage, in summary:

Table 1 - Competencies from the ENEM's framework of reference.

H5	Associating vocabulary and expressions of a text in L2 to their theme.
H6	Using knowledge of L2 and its mechanisms as a means of expanding the possibilities of access to information, technology and culture.
H7	Relating a text in L2 with its linguistic structures, function and social use.
H8	Recognizing the importance of cultural production in L2 as a representation of cultural and linguistic diversity.

Source: BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Matriz de Referência para o ENEM. Brasília, 2009. (Researcher's translation).

In competencies H5 and H7, L2 reading ability is clearly recognized as something to be developed with students inside the classroom, and considering the other items of the framework that may be applied to L2, text comprehension and interpretation dominate and dictate the classroom setting. (Conceição, 2006)

Competencies H6 and H8 are broad and may be achieved in a number of different ways (a perspective supported by Blanco (2013)). There is no explicit

mention of speaking, listening or even writing competencies. While ENEM is a reading test, and its framework serves as more of a guide to designing its questions, not directing how teaching is conducted in the classroom, the reason to mention this is the washback effect, that is, the impact an evaluative instrument has on teaching (Richards & Schmidt, 2010).

There is a lack of quantitative studies on the washback effect of ENEM in English classes. The ones that have been done have had small sample sizes and varied results: Blanco (2013) reports examples of impact in the classroom stemming from it. Vicentini, in 2014, observed little change in the practices of teachers, but significant change in the textbooks used in public schools. The same author published another study in 2015, in which she observed more change in the practices of private school teachers than in public school teachers', though her study focused on the essay part of the test (which is in L1).

The fact that the most relevant test for college admission in Brazil allows students to do well with little overall English ability does not help students see the importance of truly learning the language. Lima's study (2021) indicates that public school students do want to learn English as a whole, including how to communicate. Alves & Silva (2023) highlight the importance of oral communication to achieve the social use of the language, which is present in ENEM's framework.

Having established the above, it must be acknowledged that the discussion on whether ENEM should begin to feature its English and Spanish test questions in the target language is more nuanced than for other tests. Quevedo-Camargo et al. (2014) present a rich overview of this discussion, so the present study will rather turn its attention to other English reading tests that feature questions in L1.

2.2 HIGHER-LEVEL ENGLISH TESTS

At higher levels of education, there are also English reading tests which are more demanding on takers, but still feature questions in Portuguese. For instance, FUVEST's (Fundação Universitária para o Vestibular) English proficiency test, which gives access to master and doctorate courses, plus its transfer exams into USP (Universidade de São Paulo), part of the English courses' ENADE (Exame Nacional de Desempenho dos Estudantes), and FURG's (Universidade Federal do Rio Grande) English reading proficiency exam, to name a few. Though orientation documents, sometimes detailing the competencies evaluated in each of these tests

are available[1], none of them provide a clear justification about the usage of Portuguese over English in their items. As for other college admission exams (vestibulares), there does not seem to be a standard. Some institutions feature questions in Portuguese, others in English.

2.3 PREVIOUS STUDIES

Zhu and Aryadoust (2020) conducted a study among test-takers of the PTE (Pearson Test of English) to look for DIF between groups of students whose L1s belonged to the Indo-European family and those who didn't, in order to see if those with first languages more similar to English would have better performance on the test. The results showed the mother tongue did not have a significant impact across groups. Some DIF was observed favoring low-ability groups in some items, but according to the paper, that could be attributed to lucky guesses in difficult items and the notion that in higher levels of L2, the advantages the mother tongue could create tend to diminish.

Phisalprapa et al. (2016) looked at the results of fourth-year medical students in Thailand in an exam which had some items in Thai and others in English. The results showed that the use of English in the items did affect test-takers' scores negatively, particularly those of low performance overall.

Closer to the context of the present study, Quevedo-Camargo et al. (2014) employed similar methods to the ones used here to investigate the differences in the performance of Brazilian agronomy students who took the same English reading test twice, once with items in English, then, over a month later, in Portuguese. That study revealed no significant differences between the two performances.

Hoang et al. (2012) observed test results of students taking an introductory economics unit in Australia, and found that non-native speakers were disadvantaged by the use of high lexical density and non-specialized economics vocabulary in the items of that course's test.

---

[1] FUVEST: https://www.fuvest.br/wp-content/uploads/proficiencia-2026_edital.pdf and https://acervo.fuvest.br/fuvest/transferencia/2024/transferencia_2024_manual_atualizado_2023-04-18.pdf;
ENADE: https://www.in.gov.br/en/web/dou/-/portaria-n-274-de-28-de-junho-de-2024-569028210;
FURG: https://proficienciale.furg.br/images/SEI_FURG_-_0384158_-_EDITAL_05_2025.pdf

3 METHODOLOGY

When selecting the texts that have been used in the study's English reading test, the CEFR level was taken into consideration. The CEFR (Common European Framework of Reference for Languages). It is an international standard used to describe and assess language proficiency. Developed by the Council of Europe, it provides a six-level scale: beginner (A1), pre-intermediate (A2), intermediate (B1), upper-intermediate (B2), advanced (C1), and proficient (C2). These levels outline what a learner can do in listening, reading, speaking, and writing, in the form of competencies (Council of Europe, 2001).

The questionnaire used in this study to investigate differences in test-takers performance was a two-step English reading test combined with a survey. The researcher applied it in UnB classes which had mostly undergraduates from the English teaching, English language and literature, and English translation courses. The questionnaire was entirely digital, shared with students through a QR code, and consisted of the following, in order:

1. Questions about students' age, university course, gender, ethnicity, and income. (Table 5)

2. Reading test: a text in B2 level of English followed by 5 multiple-choice questions about it, then a C1 text, also followed by 5 questions (Appendix).

3. Another test – the same texts and items, but this time, the items were in the other language, and the alternatives had been shuffled so as not to have the same answer key as the first test. Half of the participants (group EN1) took the test with the questions in English first, then Portuguese, and the other half took the test with the questions in Portuguese first (group PT1).

4. A final section where students were asked which test they preferred, expressed whether they think multi-subject national exams should feature questions in Portuguese or English, could leave remarks or comments about the study, and chose, from a bank, words they would use to describe each test. (Table 2).

Table 2 - Word bank to describe each version of the test.

Difficult	Easy	Organized	Disorganized
Clear	Ambiguous	Professional	Amateurish
Serious	Informal	Efficient	Inefficient
Useful	Useless	Complicated	Uncomplicated
Better	Worse		

Source: Author's own work.

Both texts and most of the items used were taken from FUVEST's transfer exams, the first from the 2021 test, and the second, from the 2024 one, both available online[2]. As the entity responsible for drafting USP's admission exams, which are highly competitive because of the university's high-ranking status (QS World University Rankings, 2024), FUVEST is regarded as a reliable test producer. The texts, in turn, had been adapted from articles from The New York Times and The Washington Post, respectively. These specific texts were selected to match the reading level of UnB's English courses' students, as identified by Borges-Almeida's (2023) study. The texts were run through AI models (ChatGPT and DeepSeek) to have its CEFR levels identified.

The choice to apply both versions of the test consecutively stems from the understanding that, to maintain reliability of scores between two parallel (equivalent) tests, there must be no opportunity for test-takers to improve their reading comprehension skills or share experiences about the tests. (Hathcoat et al., 2016).

To analyze the results, the researcher applied a combination of Classical Test Theory (CTT) metrics to the tests' items and statistical analyses of the changes in participants' scores. CTT provides psychometric instruments which are able to attribute indexes to an item's difficulty and discrimination in a test. Difficulty is calculated as the percentage of correct answers of an item among its respondents (in this study, each group considered separately). The difficulty index of an item can be interpreted like so (such that 100% difficulty means all participants chose the correct answer to an item, and 0% means all of them answered incorrectly) (Nünez, n.d.):

[2] https://acervo.fuvest.br/fuvest/transferencia.html

Table 3 - Difficulty indexes of an item.

Level	Difficulty Index
Too easy	75% - 100%
Easy	55% - 74%
Intermediate	45% - 54%
Difficult	25% - 44%
Too difficult	0% - 24%

Source: NÜNEZ, n.d., p. 6

The discrimination index of an item refers to its ability to separate the highest-scoring respondents from the lowest-scoring ones. To determine it, the 25% respondents with the best and worst performances are considered (and respectively labeled the high group and low group), then the proportion of correct responses in the high group in relation to the low group are calculated (Salkind, 2017). In this study, the high and low groups were determined using the total score of the test, that is, adding up the score of both EN and PT versions. This index can be interpreted as such (Nünez, n.d.):

Table 4 - Discrimination of an item.

Classification	Discrimination Index
Very high	0.4 - 1.0
High	0.3 - 0.39
Moderate	0.2 - 0.29
Low	0.1 - 0.19
Very low	0.0 to 0.9

Source: NÜNEZ, n.d., p. 6

As for the statistical analysis, the Wilcoxon signed-rank test was utilized, which is appropriate to assess whether pairs of samples (in this case, the scores of the EN and PT tests) varied meaningfully or if the changes could be attributed to chance (Siegel, 2007). This test was chosen over the more common paired samples t-test because of the small sample size of the groups.

Some limitations must be acknowledged. While the test takers were advised in person not to cheat, talk, or change the answers to their first test, data collection has been done in an informal environment with little monitoring.

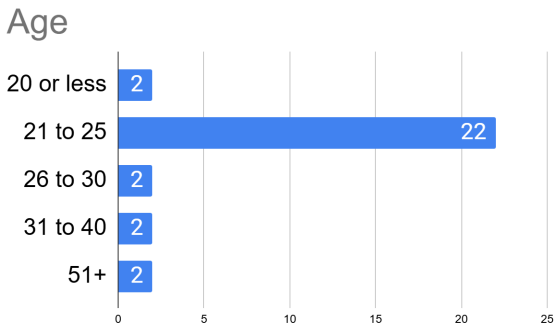
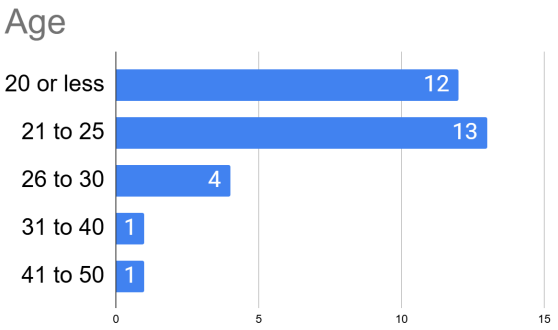
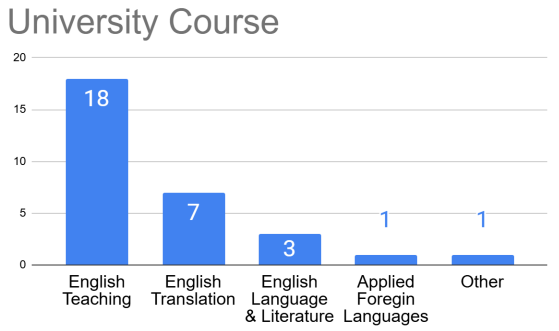
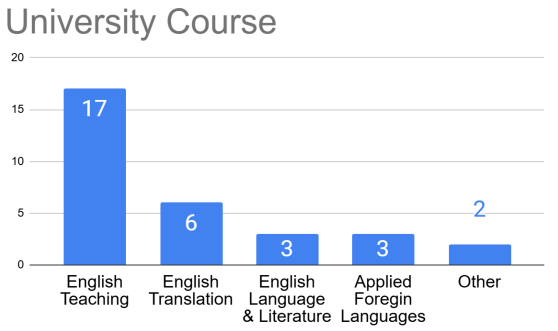
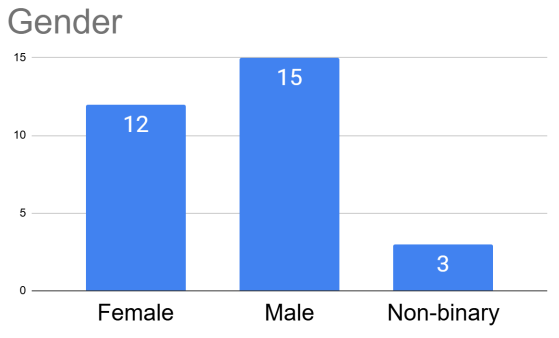
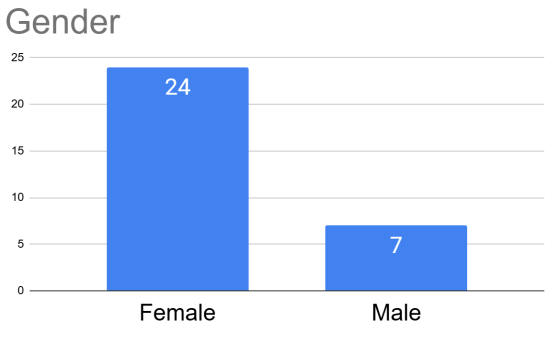
4 RESULTS & ANALYSES

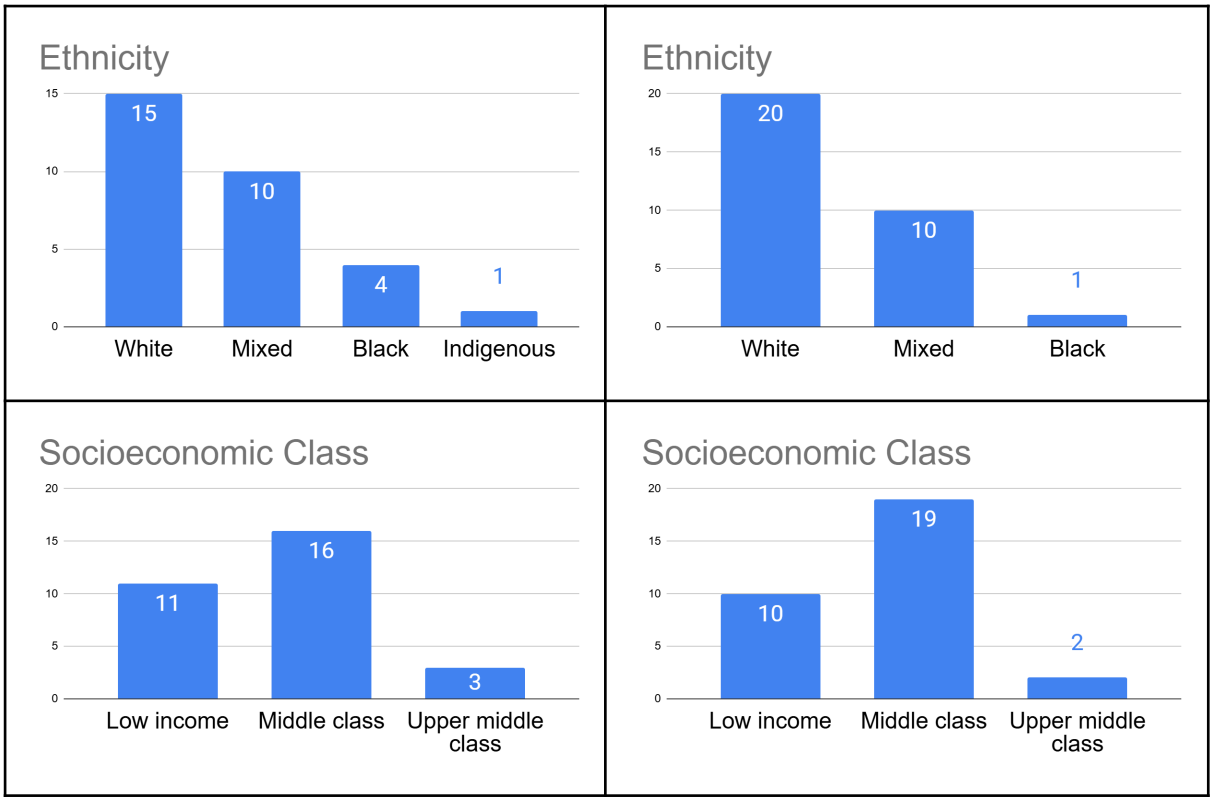
The following analysis will detail the demographics, then look at the groups individually, focusing on the behavior of the items and scores, and finish by comparing data between the two groups using statistical procedures and exploring the qualitative, opinion-based part of the questionnaire.

4.1 DEMOGRAPHICS

The total number of participants was 61. The detailed demographics are presented in Table 5:

Table 5 - Demographics of the study.

EN1 (30 respondents)	PT1 (31 respondents)
	
	
	



Source: Author's own work.

The vast majority of participants (49) were under 25 years old, and over half (35) were enrolled in the English Teaching course. Female students were overrepresented in the PT1 group (24 to 7). 57% of total respondents identified as white, with less representation of other ethnicities in the PT1 group. A third of the participants, with similar representation in both groups, reported being from low income backgrounds. No respondents selected the options "Very low income" or "Upper class".

Taking the sample sizes and profile of participants (with a higher level of English) into consideration, as well as the fact that both tests were taken sequentially, the initial hypothesis is that the language will not have a big impact on test-takers' performance.

The majority of questions in the test (Appendix), with the exception of items 8 and perhaps 4, demanded a combined ability of scanning (finding the part of the text where the relevant piece of information can be found) and abstraction (synthesizing the text's ideas and matching them with the item's paraphrasing), which taps into vocabulary and grammar knowledge. Under Grabe's (2019) perspective, the purposes for reading would be reading to search for information and reading for

general comprehension. As the author defends, reading for general comprehension is not a straightforward process, and may even be more difficult than learning to read.

4.2 EN1 ANALYSIS

The following table presents the difficulty (Dif) and discrimination (Dis) of each item in each test.

Table 6 - Difficulty and Discrimination of the items in the EN1 test.

Item		English	Portuguese	Item	English	Portuguese
1	Dif	0.90 - Too easy	0.97 - Too easy	6	0.47 - Intermediate	0.50 - Intermediate
	Dis	0.29 - Moderate	0.14 - Low		0.86 - Very High	0.71 - Very High
2	Dif	0.90 - Too easy	0.90 - Too easy	7	**0.83 - Too easy**	**0.80 - Too easy**
	Dis	0.29 - Moderate	0.29 - Moderate		0.57 - Very High	0.57 - Very High
3	Dif	0.83 - Too easy	0.90 - Too easy	8	0.83 - Too easy	0.87 - Too easy
	Dis	0.43 - Very High	0.29 - Moderate		0.71 - Very High	0.43 - Very High
4	Dif	0.83 - Too easy	0.83 - Too easy	9	**0.67 - Easy**	**0.63 - Easy**
	Dis	**0.29 - Moderate**	**0.43 - Very High**		0.71 - Very High	0.71 - Very High
5	Dif	0.67 - Easy	0.70 - Easy	10	0.77 - Too easy	0.80 - Too easy
	Dis	0.71 - Very High	0.57 - Very High		0.43 - Very High	0.29 - Moderate

Source: Author's own work.

In regards to difficulty, 9 items were easy or too easy. This is unsurprising, considering the profile of respondents (English courses' students). Items 6 to 10 showed considerable discrimination, which is expected due to the more advanced CEFR classification of the second text. For 7 out of 10 items, there was an increase in the number of correct answers in the second test, and also a decrease in discrimination. This can be explained by the fact that most respondents who changed their answers in the second test changed them to a correct answer, some of them belonging to the low group.

With this in consideration, three items (emphasized in bold in the table) behaved atypically: 4, 7, and 9. Looking at the individual results of both versions, the changes in level or classification of each item have been caused by a difference of a

single response, given the small sample size. This, combined with the possibility of guessing and the effect of fatigue, lead to the conclusion that these items are not representative of the respondents' overall development.

The following table contains the development of the respondents who had a different score in each test. Fourteen respondents are not represented, as they had the same score in both tests.

Table 7 - Development of respondents with different scores.

Respondent	Score in the EN test	Score in the PT test	Score difference	Respondent	Score in the EN test	Score in the PT test	Score difference
1	3	7	+4	9	4	5	+1
2	8	10	+2	10	3	4	+1
3	8	10	+2	11	10	9	-1
4	9	10	+1	12	9	8	-1
5	8	9	+1	13	8	7	-1
6	7	8	+1	14	8	7	-1
7	7	8	+1	15	5	4	-1
8	5	6	+1	16	10	6	-4

Source: Author's own work.

In this group, 14 respondents had the exact same performance, 10 saw improvement and 6 had lower scores. To assess whether the differences in score were statistically significant, a Wilcoxon signed-rank test was conducted, resulting in $W = 48.0$, $p = 0.323$. Although there was slightly more improvement, these results suggest that, for the EN1 group, the change from EN to PT did not lead to a meaningful variation in performance.

4.3 PT1 ANALYSIS

The difficulty (Dif) and discrimination (Dis) of each item in the PT1 tests are presented in Table 8, with PT before EN to match the order of administration:

.

Table 8 - Difficulty and Discrimination of the items in the PT1 test.

Item		Portuguese	English	Item	Portuguese	English
1	Dif	0.84 - Too easy	0.87 - Too easy	6	0.48 - Intermediate	0.68 - Easy
	Dis	0.57 - Very High	0.43 - Very High		0.43 - Very High	0.57 - Very High
2	Dif	0.81 - Too easy	0.81 - Too easy	7	0.71 - Easy	0.74 - Easy
	Dis	0.43 - Very High	0.43 - Very High		0.43 - Very High	0.43 - Very High
3	Dif	0.87 - Too easy	0.77 - Too easy	8	0.84 - Too easy	0.81 - Too easy
	Dis	0.43 - Very High	0.43 - Very High		0.29 - Moderate	0.29 - Moderate
4	Dif	0.58 - Easy	0.58 - Easy	9	0.65 - Easy	0.48 - Intermediate
	Dis	0.86 - Very High	0.86 - Very High		1,00 - Very High	1,00 - Very High
5	Dif	0.81 - Too easy	0.71 - Easy	10	0.77 - Too easy	0.77 - Too easy
	Dis	0.29 - Moderate	0.43 - Very High		0.43 - Very High	0.43 - Very High

Source: Author's own work.

Similarly to the EN1 group, 8 items were easy or too easy for this profile of respondents. However, in this group, there was a slight overall decrease in performance, which explains why there is not a consistent pattern of difficulty from the first test to the second. Discrimination was high in this group, with only two items (5 and 8) not registering very high levels across the two versions.

As before, the following table presents score development, minus the ones that stayed the same:

Table 9 - Development of respondents with different scores.

Respondent	Score in the EN test	Score in the PT test	Score difference	Respondent	Score in the EN test	Score in the PT test	Score difference
1	5	8	+3	10	8	7	-1
2	8	10	+2	11	9	8	-1
3	9	10	+1	12	9	8	-1
4	8	9	+1	13	7	5	-2
5	8	9	+1	14	7	5	-2
6	7	8	+1	15	8	6	-2
7	6	7	+1	16	9	7	-2
8	5	6	+1	17	10	7	-3
9	1	0	-1				

Source: Author's own work.

In this group, 14 respondents had the same performance, 8 saw improvement and 9 had lower scores. This time, the overall performance fell slightly, but as before, the statistical analysis does not clearly point to the language of the items heavily influencing the outcome for most participants of this profile. Wilcoxon signed-rank test results: W = 62.5, p = 0.517.

4.4 COMPARATIVE ANALYSIS

Before the beginning of the study, the researcher envisioned a single group of test takers, similarly to Quevedo-Camargo et al.'s (2014), and pondered whether to have respondents take the PT or EN test first. Here, the opportunity arose to instead have two groups and compare them.

4.4.1 PERFORMANCE AND DEVELOPMENT

Strictly speaking, the statistical analyses do not support one language making the test easier than the other, no matter which test was administered first. This, however, is not definitive proof that the language used does not make a difference. A number of factors must be taken into consideration: firstly, 61 participants do not constitute a large sample size. Secondly, as outlined by Zhu and Aryadoust (2020) and in line with the findings of Phisalprapa (2016), in higher proficiency levels, the differences tend to diminish. Finally, considering the numbers through another lens, we can verify that the sum of PT test scores was greater than those of EN tests in both groups (the average score of tests with the items in EN was 7.46, and for PT tests, 7.62). Moreover, analysis of the low groups (comprised of 7 students each), reveals that in EN1, 4 of the participants' scores improved and one declined when switching to Portuguese, while in PT1, 2 participants saw a decline in their scores, and one saw improvement switching to English. This could point, albeit inconclusively, to L1 indeed providing an advantage to the participants, particularly the ones with lower proficiency.

Taking a more qualitative approach, in both groups there have been participants who benefited from the change in language and those who have been hindered by it. This points towards the first question of this study – whether the language of the items helps or hidners – having a distinct answer based on each individual's thought process and experience with the language.

Finally, a more pessimistic interpretation could lay doubts about the validity of the layout of the questionnaire. After all, the high number of negative developments in both groups (particularly the second) is unexpected. One would assume that, once participants realized that the texts and questions were the same, they would give extra focus to items they had been uncertain about in the first test, thereby improving their score.

4.4.2 PARTICIPANT PERCEPTIONS

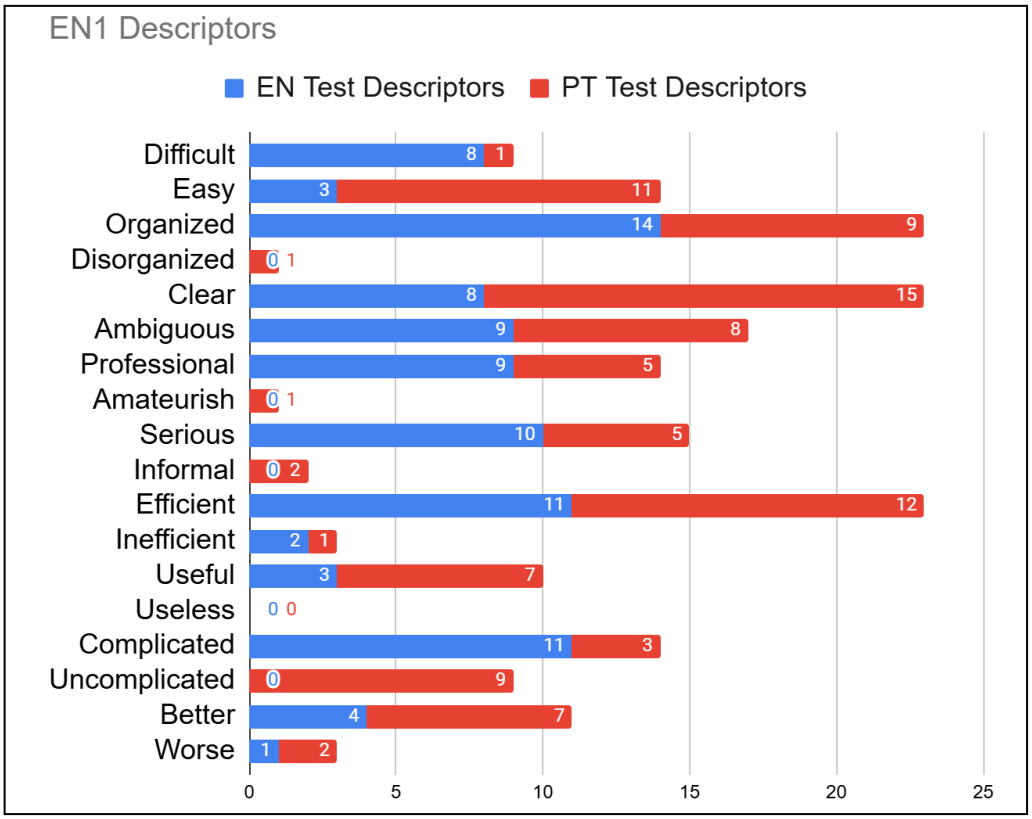
This study also sought to map the perceptions of English courses' students about the topic. The following questions were asked in the end of the questionnaire (translated from Portuguese):

- In your opinion, what words do you associate more with the test with the questions in English? Choose all that apply.
- In your opinion, what words do you associate more with the test with the questions in Portuguese? Choose all that apply.
- Which version of the test did you prefer?
- In your opinion, tests such as university admission exams and ENEM (which cover subjects other than English) should have English test questions in which language?
- Would you like to leave a comment about the test, the study or your sentiments regarding the language used in the items of English exams?

4.4.2.1 DESCRIPTORS

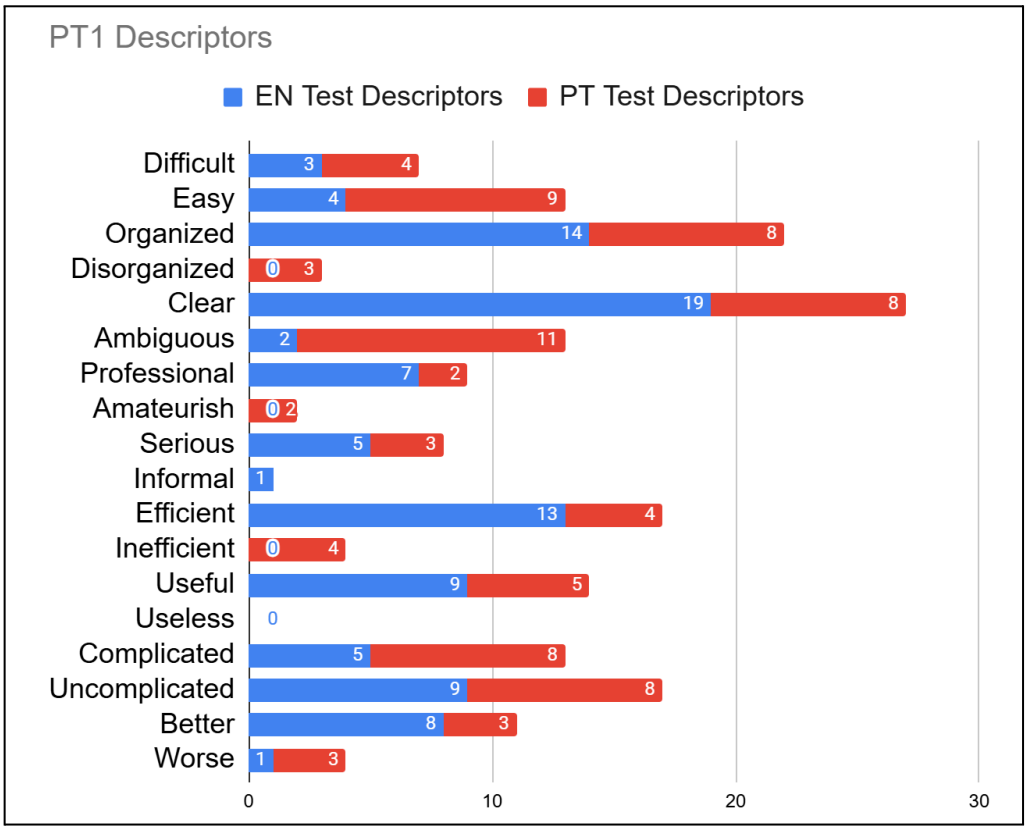
Below, the graphs (1 and 2) showing the results of the first two questions are explored. In this study, we have decided to name these chosen words "descriptors".

Graph 1 - Descriptors chosen by the EN1 group.



Source: Author's own work.

Graph 2 - Descriptors chosen by the PT1 group.



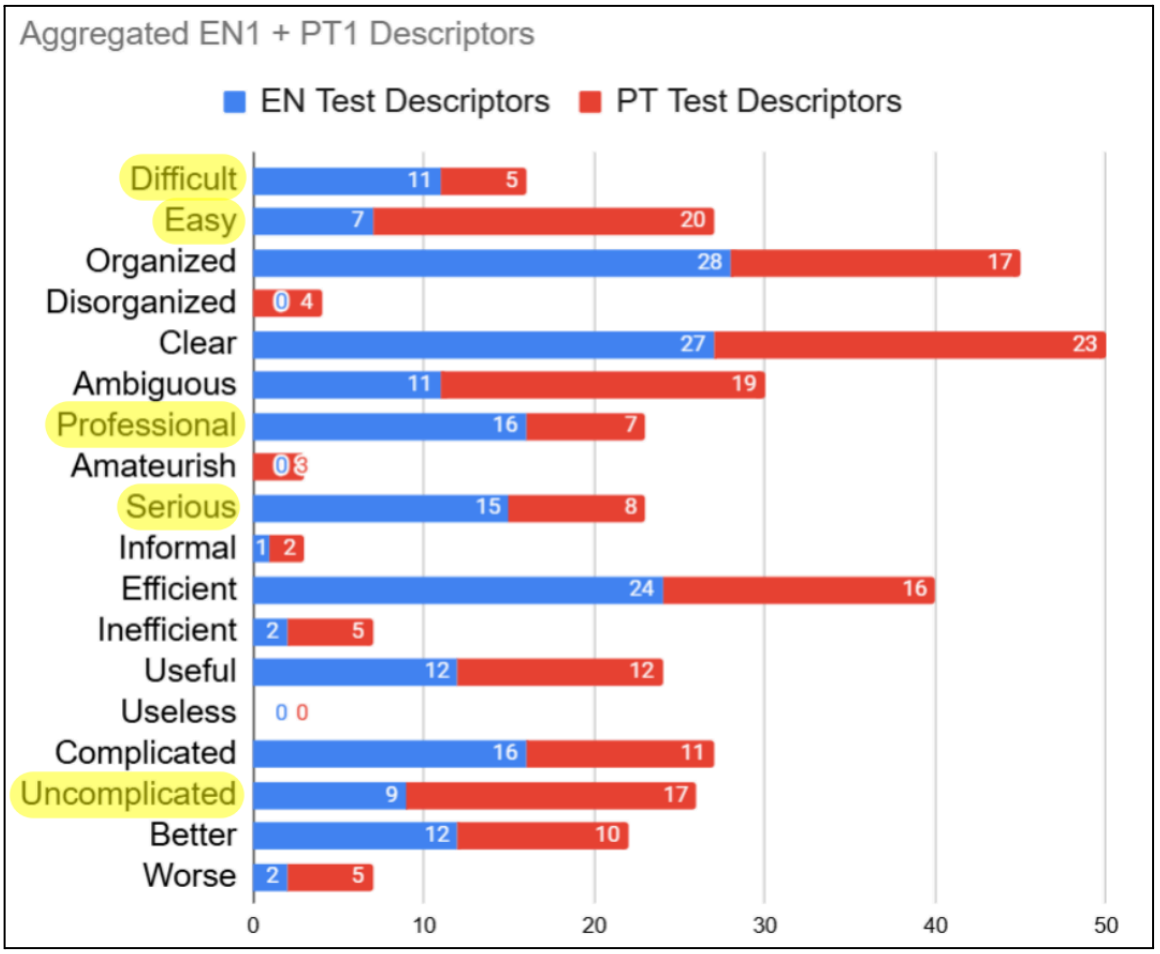
Source: Author's own work.

For EN1, *efficient* stands out as being highly picked for at least one of the tests. Analysing the data reveals that those students tend to see the EN test as more serious, organized and professional, but also more difficult and complicated. The opposite is observed for the PT test: clear, uncomplicated and easy were chosen much more often. These observations, however, are only tendencies; it must be taken into consideration, among other factors, that a considerable number of respondents did view the EN test as clear and the PT test as organized.

Curiously, in group PT1, it's possible to observe a pattern of more positive words chosen for the EN test (such as *organized, professional, efficient, useful,* and *better*), and more negative ones for the PT test (*ambiguous*, *inefficient*). Upon closer inspection, we discovered this had been the case even among students who had a decline in performance (5 out of 9 such cases).

One might propose that the order of administration of the tests could have influenced participant perception. Given this assumption, it may be more fruitful to explore the aggregated descriptors and focus on the largest differences. In the graph below, the descriptors that have been chosen more than 10 times and had a ratio very close to or higher than 2:1 for either test have been highlighted. In plain terms, students see the PT test as easier, more uncomplicated, while the EN version is seen as more difficult and professional/serious. Generally speaking, students see both as useful and efficient.

Graph 3 - Aggregated descriptors of the tests.



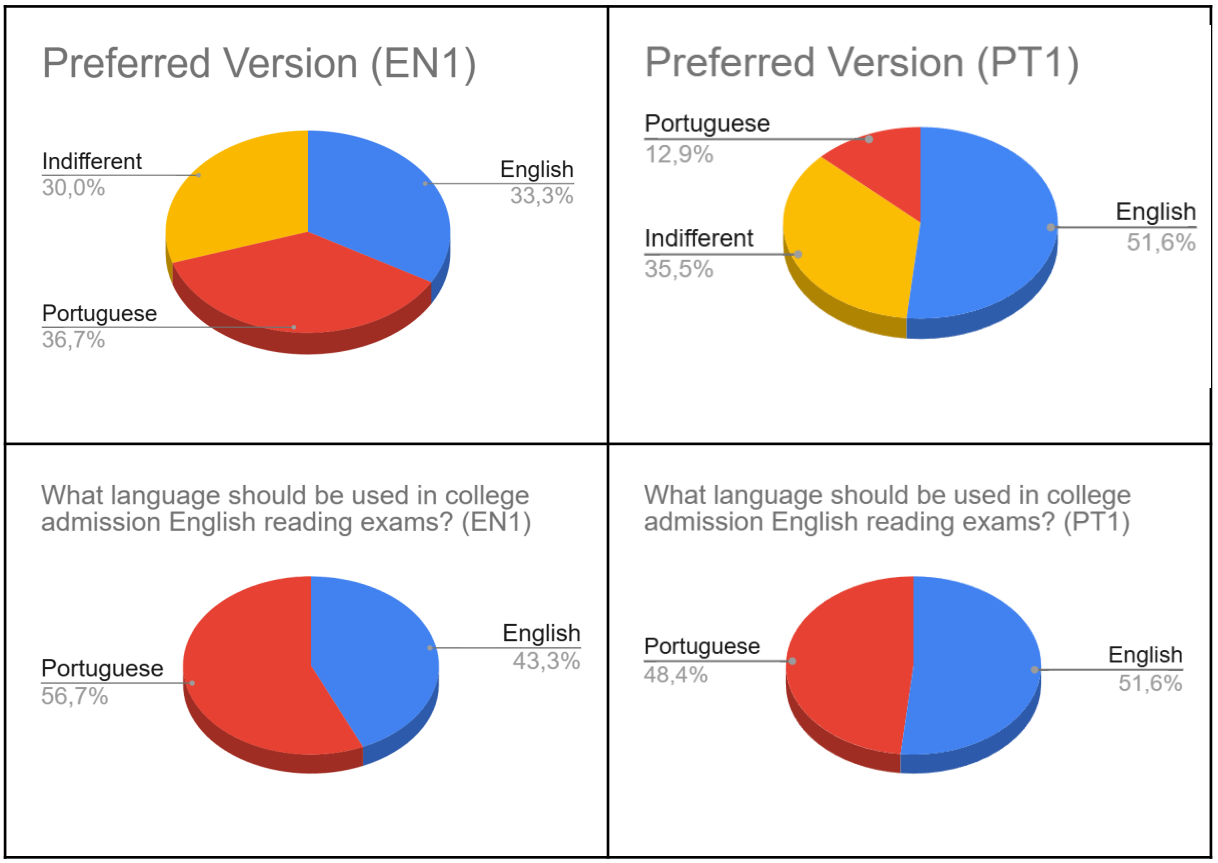
Source: Author's own work.

4.4.2.2 PREFERENCES AND OPINIONS

Regarding preference, in EN1 participants were equally divided on which test they preferred, while in PT1, only 4 students preferred the Portuguese version.

In regard to students' opinions concerning the language used in admission exams, it's important to highlight the exact wording used in the survey – "In your opinion, tests such as university admission exams and ENEM (which cover subjects other than English) should have English test questions in which language?". The mention of ENEM could have been a key factor which made students choose Portuguese over English in EN1 (Table 11).

Table 10 - Preferences about test and opinions about admission exams



Source: Author's own work.

At the end of the questionnaire, participants were invited to leave comments about the study, the test, or their own thoughts regarding the topic of this study. The comments (left in Portuguese, as they were written) most pertinent to the study are displayed and systematized below (Table 12). Certain relevant passages have been highlighted.

Table 11 - Comments left by participants.

Positive toward EN
Sinto que quando o texto está em inglês e as perguntas também, se torna mais **fácil de identificar a parte do texto** a qual a questão se refere.
Achei curioso como pra mim foi **fácil de compreender** e responder a prova com os itens em inglês.
Algumas **respostas dos itens em inglês foram mais fáceis de encontrar** no texto, outras estavam, aparentemente, no mesmo nível das em português.
Positive toward PT
Meu cérebro ficou meio confuso fazendo as mesmas questões em línguas diferentes e comecei a duvidar das minhas respostas da primeira vez, além de também duvidar da minha capacidade cognitiva. Eu ainda acho que deixar os itens em Português é melhor de alguma forma, funciona de forma **mais clara para muitas pessoas**. É assim que provas de **inglês Instrumental** são desenvolvidas.
Creio que o enunciado em inglês e as alternativas em português seria mais efetivo para a elaboração **das provas do ENEM**.
Acredito que **por ser minha língua materna eu tive melhor compreensão** e fluidez respondendo as perguntas em português.
Comments weighing advantages and disadvantages of each test and others
Acho que a decisão de fazer os enunciados depende muito do **objetivo da avaliação**. Nem sempre **o enunciado demanda os mesmos conhecimentos para ser interpretado que o texto**, então sendo esse o caso, talvez não seja muito interessante fazê-lo em inglês. **O Enem, por exemplo, tem objetivos muito específicos** com suas questões. Ao mesmo tempo, tenho a impressão de que a **qualidade das avaliações feitas completamente em inglês é melhor**.
As duas são muito boas, mas devem ser usadas para **diferentes propósitos**.
Mesmo que eu **tenha respondido a parte em inglês mais rápido**, eu considero que as perguntas devem ser feitas em português porque eu acho que **assim é mais acessível para mais pessoas**.
Acho um pouco complicado ter que **mudar o idioma em que estou pensando** para responder as questões. Contudo, também acho que as **questões em português devem ser mais fáceis para outras pessoas**.
Eu acredito que os enunciados e itens em inglês sejam ótimos pra **nós que temos um domínio maior da língua inglesa**, mas tenho medo que mudá-los em provas como o ENEM possa **causar dificuldade pela quantidade de estudantes de ensino médio sem um domínio** tão grande assim da língua inglesa que a escolhem como língua estrangeira por falta de alternativa.
Apesar de acreditar que provas de inglês com as perguntas em inglês **sejam mais claras e podem discorrer melhor sobre o assunto da prova**, é uma visão de alguém que é fluente em inglês desde os 15 anos. Ou seja, em uma prova de concurso nacional (como o ENEM), **onde nem todos sabem fluentemente a língua**, uma prova com perguntas em inglês se torna mais eficiente para comprovar **apenas o conhecimento pleno de inglês** de quem faz a prova.
Acredito que esse é um estudo muito importante de ser feito. Penso muito nessa questão do idioma utilizado nos enunciados de provas de pequena e larga escala, bem como no idioma utilizado na elaboração dos livros didáticos adotados em escolas públicas. Acredito que **a escolha do idioma deve levar em conta principalmente o contexto** em que aquela prova e material

didático está inserido. Provas de larga escala como o **ENEM e livros didáticos elaborados para as escolas públicas** brasileiras **deveriam conter mais instruções e explicações em português**.

Acredito que é importante, em provas vestibulares, **cobrarem tanto a gramática quanto a interpretação de leitura**, pois é necessário que o indivíduo **saiba e possa responder sobre o idioma em áreas diferentes**.

Creio que o que interfere é o **nível de inglês do candidato** e não realmente a formulação da prova.

Source: Compiled by the author.

Looking at the comments left by some of the participants, one opinion stands out, having been expressed by several students: that test with items in L1 make the test more accessible to test-takers with a lower level of English. This opinion was held even by some participants who preferred the English version, and is in line with the notion that the language the item is written in affects low-proficiency test-takers more than intermediate and high proficiency ones (Zhu; Aryadoust, 2020; Phisalprapa, 2016). As for the advantage of items in English, some students mentioned clarity and/or less difficulty in finding the relevant passages in the text, which naturally can not be expected from students with a lower level of English.

5 CONCLUSIONS AND FINAL CONSIDERATIONS

With the intention of aiding the discussion about what language should be used in the items of English tests, this study showed examples of tests for which these results could be relevant, detailed ENEM's situation amid this discussion, compiled the results of similar studies done in different parts of the world, and applied a test to UnB's students to verify the effect of the language used, as well as give students a means to express how they viewed the situation.

The studies reviewed showed mixed results, but a constant emerged: items in L1 seem to aid students with lower proficiency. Some of this study's participants, already convinced of the value of this concept, argued that keeping items in L1 would make tests more accessible (particularly ENEM).

For this profile of participants, the tests had a high ratio of correct answers, and no statistically relevant changes were observed in the participants' scores between tests, no matter which one was administered first. Participants tended to see the test with items in PT as easier and less complicated, and the EN test as more difficult, but also more serious and professional.

Although not the main focus of the study, it was intriguing to see a pattern of affinity towards English in PT1 (both in preferences and descriptors), even among participants whose performances fell. It's unclear why this group favored EN and showed disfavor towards PT. Again, order of test administration could have been a factor, as well as a general affinity towards English this profile of test-takers may naturally have. Further qualitative research is needed.

Another inadvertent finding of this study was the difficulty classification (based on the CEFR and made by AI models) of many of the English texts used in FUVEST's transfer exams, which resulted in very high levels of C1 texts. It's unclear whether the decision to use such complex texts is deliberate, and, one might argue, incompatible with the usage of items in L1 (whose primary argument in its favor seems to be accessibility).

Readressing the question of the validity of the questionnaire, perhaps it would be fruitful, in future studies, to explore the outlying cases further in depth, that is, making a similar experiment but conduct interviews and follow-up work with the participants who had the biggest differences in score. This would also allow for investigation of preferences towards the first test or the second, and could have implications in other fields of knowledge.

Furthermore, this study also calls for explicit teaching of reading strategies in ESL classrooms, which have been shown to aid in comprehension of texts in English L2 settings (Ludji, 2020; Mekuria, 2024). After handling the same items multiple times, it became very clear to the researcher that a participant could more easily find the right answers by first identifying the relevant part in the text, then meticulously checking whether the wording of each alternative was supported by such part. If a test-taker does not work under the assumption that all alternatives of an item refer to the same part of the text directed by the stem, it can create uncertainty and increase fatigue and time spent on the item.

To conclude this paper, I would like to offer my own perspective on the subject. As an ESL teacher who has received both formal education and informal “best-practices” training on assessment, my view is that, at least in exams that demand higher proficiency reading levels, the items should indeed be in L2. Not only because this paper and others published before showed little difference in performance stemming from item language in higher levels, but also, I would argue, because these kinds of tests can actually benefit from the supposed added discrimination of the use of L2 in the items. Being able to understand written English at a high level entails understanding questions and decoding complex writing, no matter if the wording comes from the text or the item (which is also a text in its own capacity).

REFERENCES

ALVES, Regiane Queiroz; SILVA, Reinaldo Ferreira da. **A oralidade no ensino de língua inglesa no Fundamental II**. In: SEMINÁRIO FORMAÇÃO DE PROFESSORES E ENSINO DE LÍNGUA INGLESA, 7., 2023, São Cristóvão. Anais [...]. São Cristóvão: Universidade Federal de Sergipe, 2023. p. 224–233.

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION; AMERICAN PSYCHOLOGICAL ASSOCIATION; NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION. **Standards for educational and psychological testing**. Washington, DC: American Educational Research Association, 2014.

ARAÚJO, Sara Domingos de Sousa. **A validade de construto das questões de língua inglesa do ENEM: um olhar para o conteúdo e a matriz de referência**. 2017. Dissertação (Mestrado em Linguística Aplicada) – Universidade de Brasília, Instituto de Letras, Departamento de Línguas Estrangeiras e Tradução, Brasília, DF, 2017.

BLANCO, J. **A avaliação de língua inglesa no Enem: efeitos de seu impacto social no contexto escolar**. 2013. 132f. Dissertação (Mestrado em Linguística) – Universidade Federal de São Carlos, São Carlos, 2013.

BORGES-ALMEIDA, Vanessa. **Mapeando a proficiência em leitura e compreensão oral em inglês de alunos de Letras**. Trabalhos em Linguística Aplicada, Campinas, v. 62, n. 1, p. 140–157, jan./abr. 2023.

BRASIL. Ministério da Educação. **Matriz de Referência do Novo Enem**. Brasília: MEC, 2009. Disponível em: http://portal.mec.gov.br/dmdocuments/matriz_referencia_novoenem.pdf. Acesso em: 28 maio 2025.

CONCEIÇÃO, Mariney Pereira. **Experiências de aprendizagem: reflexões sobre o ensino de língua estrangeira no contexto escolar brasileiro**. Revista Brasileira de Linguística Aplicada, v. 6, n. 2, p. 185–204, 2006.

COUNCIL OF EUROPE. **Common European Framework of Reference for Languages: learning, teaching, assessment**. Cambridge: Cambridge University Press, 2001.

GRABE, William; STOLLER, Fredricka L. **Teaching and Researching Reading**. 3. ed. Londres: Routledge, 2019.

GRABE, William. Teaching and testing reading. In: LONG, Michael H.; DOUGHTY, Catherine J. (eds.). **The Handbook of Language Teaching**. Chichester: Wiley-Blackwell, 2009. p. 441–462. DOI: 10.1002/9781444315783.ch24.

HATHCOAT, John D.; SANDERS, Courtney B.; GREGG, Nikole. **Selecting and designing instruments: item development, reliability, and validity**. Harrisonburg,

VA: James Madison University, Center for Assessment and Research Studies, 2016.

HOANG, Viet-Ngu; MAY, Lynette A.; TANG, Tommy. **The effects of linguistic factors on student performance on economics multiple choice questions**. Social Science Research Network, [s. l.], 2012. 16 p. DOI: 10.2139/ssrn.2121816.

LIMA, Joana Gomes da Silva. **A relação entre teoria e prática no ensino de Língua Inglesa no ensino médio de uma escola pública: a necessidade da implementação de uma prática oral cotidiana**. Trabalho de Conclusão de Curso (Licenciatura em Letras – Língua Inglesa) – Universidade Federal da Paraíba, Mamanguape, 2021.

LUDJI, Indah Welasasih. **Extending the reading performance of low-level readers through the use of a diagnostic assessment tool in explicit strategic reading instruction**. Jurnal Ilmiah Lingua Idea, [S.l.], v. 11, n. 1, p. 16–29, jun. 2020. DOI: 10.20884/1.jli.2020.11.1.2645.

MACIEL, Mayan Figueiredo Ribeiro. **O gênero lírico nos itens de língua inglesa do ENEM e habilidades de leitura**. Trabalho de Conclusão de Curso (Graduação em Letras) – Universidade de Brasília, Instituto de Letras, Departamento de Línguas Estrangeiras e Tradução, Brasília, DF, 2025.

MEKURIA, Asalifew; BUSHISHO, Elias Woemego; WUBSHET, Hailu. **The effects of reading strategy training on students' reading strategy use and critical reading ability in EFL reading classes**. Cogent Education, v. 11, n. 1, p. 1–21, fev. 2024. DOI: 10.1080/2331186X.2024.2310444.

NICKERSON, Raymond S. **Confirmation bias: a ubiquitous phenomenon in many guises**. Review of General Psychology, v. 2, n. 2, p. 175–220, 1998. DOI: 10.1037/1089-2680.2.2.175.

NÜNEZ, I.B. **Instrumentos de avaliação: provas de perguntas discursivas**. s.d. Disponível em: http://www.sistemas.ufrn.br/shared/verArquivo?idArquivo=670232&key=8f571209bd3523d4ee8cd349145b63a6. Acesso em: 26 jun. 2025.

OLIVEIRA, Jossane Rodrigues de; BUENO, Ivonete. **English proficiency in Latin America: facing inequalities**. Revista EntreLínguas, Araraquara, v. 7, n. 00, p. e021003, 2021.

PHISALPRAPA, Pochamana; MUANGKAEW, Wayuda; ASSANASEN, Jintana; KUNAVISARUT, Tada; THONGNGARM, Torpong; RUCHUTRAKOOL, Theera; KOBWANANTHAKUN, Surapon; DEJSOMRITRUTAI, Wanchai. **The effect of English language on multiple choice question scores of Thai medical students**. Journal of the Medical Association of Thailand, v. 99, n. 4, p. 446–452, abr. 2016. DOI: não disponível.

QUEVEDO-CAMARGO, G.; GIMENEZ, T.; FURTOSO, V. **Itens em português ou na língua alvo na avaliação da compreensão escrita em língua estrangeira: um**

estudo preliminar. Revista Contexturas, n. 23, p. 125–148, 2014. ISSN 0104-7485.

RAUBER, B. B. **Avaliação em língua estrangeira (inglês) no acesso ao Ensino Superior: o Enem em discussão**. 2012. 108f. Dissertação (Mestrado em Linguística) – Universidade Federal de São Carlos, São Carlos, 2012.

RICHARDS, J. C.; SCHMIDT, R. **Longman dictionary of language teaching and applied linguistics**. 4. ed. Great Britain: Pearson Education Ltd., 2010.

SALKIND, Neil J. **Tests & Measurement for People Who (Think They) Hate Tests & Measurement**. 3. ed. Thousand Oaks, CA: Sage Publications, 2017. ISBN 978-1506368382.

SIEGEL, Sidney. **Non-parametric statistics for the behavioral sciences**. New York: McGraw-Hill, 2007. Original de 1956. p. 75–83. ISBN 9780070573482.

SIQUEIRA, Domingos Sávio Pimentel; ANJOS, Flávius Almeida dos. **Ensino de inglês como língua franca na escola pública: por uma crença no seu (bom) funcionamento**. 1. ed. Ponta Grossa: Muitas Vozes, 2011.

TOP UNIVERSITIES. **QS World University Rankings**. QS Quacquarelli Symonds, 2024. Disponível em: https://www.topuniversities.com/world-university-rankings. Acesso em: 6 jun. 2025.

VICENTINI, M. P.. **A redação no ENEM e a redação no 3º ano do ensino médio: efeitos retroativos nas práticas de ensino da escrita**. Dissertação de mestrado, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, SP, 2015.

VICENTINI, Monica Panigassi. **O efeito retroativo da redação do Enem: uma análise das práticas de duas professoras do terceiro ano do Ensino Médio**. In: Anais do SIELP, Uberlândia: EDUFU, v. 3, n. 1, p. 1-9, 2014.

ZHU, Xuelian; ARYADOUST, Vahid. **An investigation of mother tongue differential item functioning in a high-stakes computerized academic reading test**. Computer Assisted Language Learning, [S.l.], v. 33, n. 8, p. 894–916, 2020. DOI: 10.1080/09588221.2019.1704788.

ZUMBO, Bruno D. **Three generations of DIF analyses: considering where it has been, where it is now, and where it is going**. Language Assessment Quarterly, v. 4, n. 2, p. 223–233, 2007. DOI: 10.1080/15434300701375832.

APPENDIX

Text 1

So far, the fastest anyone has run is about 27.5 miles per hour, a speed reached (briefly) by sprinter Usain Bolt just after the midpoint of his world-record 100-meter dash in 2009.

This speed limit probably is not imposed by the strength of our bones and tendons. Rather, a 2010 study suggested that the limit comes from our bipedal stride, in particular how quickly we can rearrange our legs while still leaving time to push off from the ground.

Peter G. Weyand, a biomechanics researcher and physiologist at Southern Methodist University and one of the authors of the study, said that our running speed is limited because we are in the air for most of our stride. During the brief moments that our feet are touching the ground, we have to exert a lot of force.

During the period of ground contact, our legs must push us forward and push us upward to support our body weight. That's a lot of force to exert in a short time - and it's why humans can skate faster than they run, Dr. Weyand said. Keeping your skates on the ground longer helps support your body during the glide phase, taking some of the load off the pushing leg.

I asked Dr. Weyand how he would redesign humans to run faster. I tossed out four options: longer legs, really wide hips, extra legs, or extra knees.

"Adding more knees is probably the trickiest one," he said. Extra knees might let you extend your legs to stay in contact with the ground longer. But if your feet get too far out from under your body, it's hard to generate enough leverage to push down against the ground. Longer legs might help, he said, but the best option would be more legs, so that we could have one or two of them on the ground most of the time, like a quadruped.

The New York Times. January 20, 2020. Adaptado.

Items in English	Items in Portuguese
1. According to the text, the maximum speed that a human being can reach when running (A) assumes constant breaking of records. (B) should soon achieve stabilization of records. **(C) involves aspects related to force and surface.** (D) depends on the size of the bones and tendons. (E) benefits from technological advances in measurement.	1. Segundo o texto, a velocidade máxima que o ser humano pode alcançar em uma corrida (A) depende do tamanho dos ossos e tendões. (B) pressupõe constantes quebras de recordes. (C) se beneficia dos avanços tecnológicos de medição. **(D) envolve aspectos relativos à força e superfície.** (E) deve, em breve, alcançar estabilização de marcas.
2. According to the text, researcher Peter G. Weyand believes that the main factor that limits human speed is associated with **(A) the short time that the foot is in contact with the ground.** (B) the amplitude of the athletes' strides and jumps. (C) the conditions of the available running tracks. (D) the decrease in progression over the years. (E) individual genetic characteristics.	2. Conforme o texto, o pesquisador Peter G. Weyand acredita que o principal fator que limita a velocidade humana está associado (A) às características genéticas individuais. (B) às condições das pistas de corrida disponíveis. **(C) ao pouco tempo que o pé está em contato com o solo.** (D) à amplitude das passadas e saltos dos atletas. (E) à queda da progressão com o passar dos anos.
3. Skating is mentioned in the text to show that compared to running, in this sport there is (A) a set of risk factors for regular practice. (B) a balance between strength and laterality. (C) heavy tension in the knees and hip region. (D) changes in the properties of inactive muscles. **(E) reduced load on the leg responsible for pushing.**	3. O "skate" é mencionado no texto para mostrar que, nessa modalidade esportiva, em comparação com a corrida, observa-se (A) equilíbrio entre força e lateralidade. (B) conjunto de fatores de risco para a prática regular. (C) alteração nas propriedades de músculos inativos. (D) tensão acentuada nos joelhos e região dos quadris. **(E) redução da sobrecarga na perna responsável pelo impulso.**

4. Considering the context, the expression "the trickiest" (last paragraph) has a similar meaning to **(A) the most complex.** (B) the most gullible. (C) the smoothest. (D) the weirdest. (E) the farthest.	4. Considerado o contexto, a expressão "the trickiest" (último parágrafo) tem o sentido de **(A) mais complexa.** (B) mais estranha. (C) mais fácil. (D) mais ingênua. (E) mais longe.
5. Regarding the interviewer's suggestion of adding knees to run faster, Dr. Weyand (A) states that it is the best way to gain speed. **(B) contrasts the pros and cons of the suggestion.** (C) reformulates the idea, giving it greater precision. (D) mentions animals that have more folds in their limbs. (E) leaves the idea unexplored for considering it ineffective.	5. A respeito da sugestão do entrevistador de adicionar joelhos para correr mais rápido, Dr. Weyland (A) mantém a ideia inexplorada por considerá-la ineficaz. (B) afirma que é a melhor forma de se ganhar velocidade. (C) reformula a ideia, dando-lhe maior precisão. (D) cita animais que possuem mais dobras nos membros. **(E) contrasta os prós e contras da sugestão.**

Text 2

The stress of pandemic lockdowns prematurely aged the brains of teenagers by at least three years and in ways similar to changes observed in children who have faced chronic stress and adversity, a study has found. The study, published Thursday in Biological Psychiatry: Global Open Science, was the first to compare scans of the physical structures of teenagers' brains from before and after the pandemic started, and to document significant differences, said Ian Gotlib, lead author on the paper and a psychology professor at Stanford University.

Researchers knew teens had higher "levels of depression, anxiety and fearfulness" than "before the pandemic. But we knew nothing about the effects on their brains," said Gotlib, who is director of the Stanford Affect, Laboratory. "We thought there might be effects similar to what you would find with early adversity; we just didn't realize how strong they'd be."

By comparing MRI scans of a group of 128 children, half taken before and half at the end of the first year of the pandemic, the researchers found growth in the hippocampus and amygdala, brain areas that respectively control access to some memories and help regulate fear, stress and other emotions. They also found thinning of the tissues in the cortex, which is involved in executive functioning. These changes happen during normal development; however, the pandemic appeared to have accelerated the process, Gotlib said.

Premature aging of children's brains isn't a positive development. Before the pandemic, it was observed in cases of chronic childhood stress, trauma, abuse and neglect. These adverse childhood experiences not only make people more vulnerable to depression, anxiety, addiction and other mental illnesses, <u>they</u> can raise the risk of cancer, diabetes, heart disease and other long-term negative outcomes.

The pre-pandemic images of teen brains came from a longitudinal study that Gotlib's team began eight years ago, with the original goal of better understanding gender differences in depression rates among adolescents. The researchers recruited 220 children ages 9 to 13, with a plan to take MRI scans of their brains every two years. As they were collecting the third set of scans, the pandemic shut down all in-person research at Stanford, preventing the scientists from collecting brain scan data from March 2020 until late that year.

As they debated how to account for the disruption, the scientists saw an opportunity to investigate a different question: how the pandemic itself may have impacted the physical structure of the children's brains and their mental health. They matched pairs of children with the same age and gender, creating subgroups with similar puberty, socioeconomic status and exposure to childhood stress.

The Washington Post. December 1, 2022. Adaptado.

Items in English	Items in Portuguese
6. According to the text, the study led by Professor Gotlib hypothesized that the effects of the pandemic among adolescents would be (A) increases in levels of depression, anxiety, and fear. (B) changes in the relationship with stress during early childhood. (C) differences in access to memories and control of fear, stress, and other emotions. **(D) consequences in their brains similar to those of other negative experiences, on different levels.** (E) changes in the brains of adolescents resulting from high levels of depression, anxiety, and fear.	6. De acordo com o texto, o estudo liderado pelo professor Gotlib tinha a hipótese de que os efeitos da pandemia entre os adolescentes seriam (A) alterações na relação com o estresse durante a primeira infância. (B) acréscimos nos níveis de depressão, ansiedade e medo. (C) modificações nos cérebros dos adolescentes decorrentes de níveis de depressão, ansiedade e medo elevados. **(D) consequências em seus cérebros parecidas àquelas de outras experiências negativas, em diferentes escalas.** (E) diferenças no acesso a memórias e no controle do medo, do estresse e de outras emoções.
7. When comparing the tests of the children before and after the first year of the pandemic, the researchers observed that (A) certain areas of the participants' brains were underdeveloped. (B) some children have lost part of the ability to retrieve memories and control emotions. **(C) the pandemic appears to have made the development of some areas of the brain faster.** (D) there was formation of foreign bodies in the hippocampus and amygdala of the patients. (E) the differences are inconclusive, since the growth of the observed areas is normal for that age.	7. Ao comparar os exames das crianças realizados antes e depois do primeiro ano da pandemia, os pesquisadores observaram que (A) algumas crianças perderam parte da capacidade de resgatar memórias e controlar emoções. **(B) a pandemia parece ter tornado o desenvolvimento de algumas áreas do cérebro mais rápido.** (C) havia formação de corpos estranhos no hipocampo e amígdala dos pacientes. (D) certas áreas do cérebro dos participantes estavam subdesenvolvidas. (E) as diferenças são inconclusivas, já que o crescimento das áreas observadas é normal para a idade.
8. In the text, the pronoun "they" (4th paragraph) refers to (A) "the pandemic". (B) "children's brains". (C) "long-term negative outcomes".	8. No texto, o pronome "they" (4º parágrafo) refere-se a **(A) "adverse childhood experiences".** (B) "children's brains". (C) "the pandemic".

(D) "adverse childhood experiences". (E) "depression, anxiety, addiction, and other mental illnesses".	(D) "depression, anxiety, addiction, and other mental illnesses". (E) "long-term negative outcomes".
9. According to the text, the adverse experiences that are related to brain development studied by Gotlib's team refer to (A) events that resulted in high levels of depression, anxiety, and fear. (B) events during childhood and adolescence that generated chronic stress later on. (C) conditions that resulted in increased risks of cancer, diabetes, and heart disease. **(D) cases that corresponded to stress, trauma, abuse, or neglect throughout childhood.** (E) effects that altered brain executive functions during normal adolescent development.	9. De acordo com o texto, as experiências adversas que têm relação com o desenvolvimento cerebral estudado pelo time de Gotlib referem-se a (A) acontecimentos durante a infância e a adolescência que geraram estresse crônico posterior. (B) eventos que resultaram em altos níveis de depressão, ansiedade e medo. **(C) casos que corresponderam a estresse, trauma, abuso ou negligência ao longo da infância.** (D) efeitos que alteraram as funções executivas cerebrais durante o desenvolvimento adolescente normal. (E) condições que decorreram de maiores riscos de câncer, diabetes e doenças cardíacas.
10. According to the text, the pandemic changed the course of the study because (A) the available MRI machines recorded data only until March 2020. (B) the 220 children between the ages of 9 and 13 who were recruited missed the in-person tests. **(C) in-person data collection was interrupted, leading the researchers involved to reformulate the research's main question.** (D) the combination of pairs of the same age and gender generated subgroups with profiles incompatible with scientific research. (E) the change in the nature of the experiment prevented the adolescents from continuing to participate, due to the variation in age range.	10. Segundo o texto, a pandemia alterou o curso do estudo, pois **(A) a coleta presencial de dados foi interrompida, levando os pesquisadores envolvidos a reformularem a pergunta da pesquisa.** (B) as 220 crianças entre 9 e 13 anos recrutadas faltaram aos testes presenciais. (C) a mudança na natureza do experimento impediu a continuidade da participação dos adolescentes, devido à variação de faixa etária. (D) as máquinas de ressonância magnética disponíveis registraram dados somente até março de 2020. (E) a combinação de pares de mesma idade e gênero gerou subgrupos com perfil incompatível com a investigação científica.