



UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"  
Câmpus de São José do Rio Preto

Jéssica Nunes Caldeira Cunha

O PRÉ-TESTE AO EPPLE (EXAME DE PROFICIÊNCIA PARA  
PROFESSORES DE LÍNGUAS ESTRANGEIRAS): DESAFIOS  
TECNOLÓGICOS, VALIDADE E HISTÓRICO DE  
IMPLEMENTAÇÃO

São José do Rio Preto  
2019

Jéssica Nunes Caldeira Cunha

O PRÉ-TESTE AO EPPLE (EXAME DE PROFICIÊNCIA PARA  
PROFESSORES DE LÍNGUAS ESTRANGEIRAS): DESAFIOS  
TECNOLÓGICOS, VALIDADE E HISTÓRICO DE  
IMPLEMENTAÇÃO

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, Área de Concentração – Linguística Aplicada, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CAPES.

Orientador: Prof. Dr. Douglas Altamiro Consolo  
Co-orientadora: Profa. Dra. Melissa Alves  
Baffi-Bonvino

São José do Rio Preto  
2019

C972p

Cunha, Jéssica Nunes Caldeira

O Pré-Teste ao EPPLE (Exame de Proficiência para Professores de Línguas Estrangeiras): desafios tecnológicos, validade e histórico de implementação /

Jéssica Nunes Caldeira Cunha. -- , 2019

142 f. : il.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp),  
Faculdade de Ciências Farmacêuticas, Araraquara,

Orientador: Douglas Altamiro Consolo

Coorientadora: Melissa Alves Baffi-Bonvino

1. Avaliação. 2. Tecnologia. 3. Professores Formação. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Farmacêuticas, Araraquara. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Jéssica Nunes Caldeira Cunha

O Pré-Teste ao EPPLE (Exame de Proficiência para Professores de Línguas Estrangeiras): desafios tecnológicos, validade e histórico de implementação

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, Área de Concentração – Linguística Aplicada, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CAPES.

Comissão Examinadora

Prof. Dr. Douglas Altamiro Consolo  
UNESP – Câmpus de São José do Rio Preto  
Orientador

Prof<sup>a</sup>. Dr<sup>a</sup>. Maria Helena Vieira Abrahão  
UNESP – Câmpus de São José do Rio Preto

Prof<sup>a</sup>. Dr<sup>a</sup>. Sandra Regina Butros Gattolin de Paula  
UFSCAr – Universidade Federal de São Carlos

São José do Rio Preto  
14 de fevereiro de 2019

*À minha avó, Prof. Ma. Maria Nunes Caldeira, por ter me ensinado o valor da educação.*

*Aos meus pais, avô e irmã por estarem sempre ao meu lado e me apoiarem e incentivarem durante toda a minha formação.*

*Ao meu parceiro Alan, por me ensinar a confiar e investir em mim mesma.*

## AGRADECIMENTOS

Ao Prof. Dr. Douglas Altamiro Consolo, pela oportunidade, pela orientação e dedicação.

À Profa. Dra. Melissa Alves Baffi-Bonvino pela co-orientação e auxílio.

À Profa. Dra. Maria Helena Vieira Abrahão e à Profa. Dra. Suzi Marques Spatti Cavalari pela disponibilidade para debater esta pesquisa no Exame de Qualificação.

À Profa. Dra. Sandra Regina Buttros Gattolin de Paula, pelas reflexões e comentários no IX SELIN em 2017 e no X SELIN em 2018.

À Profa. Dra. Matilde Scaramucci, pelas sugestões e contribuições quando do 66º Seminário do GEL, em seu simpósio sobre avaliação.

À “parceira de temática” Profa. Ma. Camila Sthéfanie Colombo, por sempre estar aberta a ajudar, com gentileza e enorme eficiência, mesmo em momentos de provação em sua vida.

Aos professores do PPGEL e convidados com quem tive contato em suas disciplinas, Profa. Dra. Dora Riestra, Prof. Dr. Douglas Altamiro Consolo, Prof. Dr. Roberto Gomes Camacho, Profa. Dra. Solange Aranha e Profa. Dra. Suzi Marques Spatti Cavalari, pelos valiosos ensinamentos.

Aos participantes da pesquisa, por contribuírem com o fornecimento de dados para o meu trabalho e para o projeto maior do EPPLE.

Aos meus pais e avós, Edson, Miriam, Daniel e Maria (*in memoriam*) pelo apoio durante toda a minha formação, pela crença inabalável nas minhas habilidades e inteligência, por me escutarem e acalentarem quando eu tive dúvidas e medos.

Ao meu parceiro Alan, por me acompanhar de perto durante todo esse processo, me incentivando e me ajudando a me manter firme, focada e confiante. Sua companhia e apoio foram essenciais durante essa etapa tão diferente e nova na minha vida. Outra contribuição não menos importante foi a de suas leituras críticas do trabalho, me ajudando a aprimorá-lo e a melhorar meu texto.

À minha irmã Janaina, que se tornou uma amiga inestimável e “companheira de vida adulta” pelos *insights* sobre a vida acadêmica e particular, durante esse tempo em que nos tornamos mais próximas.

À querida professora Dra. Cida Zolnier, por acreditar em mim e ser meu modelo de professora e de pessoa que contribui com a aprendizagem e com a vida cotidiana de seus estudantes.

À professora Dra. Ana Barcelos pela primeira oportunidade de ser professora, pelas reflexões incitadas, pela colaboração neste trabalho.

Aos colegas mestrandos e doutorandos do Programa de Pós-Graduação em Estudos Linguísticos (PPGEL) do IBILCE/UNESP Aline Mello, Renan Felipe, Sílvia Soggio del Monte, William Danilo Garcia, pelos cafés e conhecimentos compartilhados.

Às amigas Marcela, Ákila, Maggie e Bárbara, mestrandas ou mestres em outras terras, pelo acolhimento, pelas trocas sobre a pós-graduação e pela ajuda com o trabalho.

À Profa. Ma. Eloísa Moriel Valença e sua família, por toda a ajuda e acolhimento quando do processo seletivo do mestrado.

Aos amigos feitos em Rio Preto: Bianca di Santi, Fernando Moraes, Hanae Xavier, Marina Scardoelli, pelos momentos felizes juntos, que iluminaram essa jornada.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

*“You have been told also that life is darkness, and in your weariness you echo what was  
said by the weary.*

*And I say that life is indeed darkness save when there is urge,*

*And all urge is blind save when there is knowledge,*

*And all knowledge is vain save when there is work,*

*And all work is empty save when there is love;*

*And when you work with love you bind yourself to yourself, and to one another, and to*

*God.*

*(...)*

*Work is love made visible.”*

*(Khalil Gibran – The Prophet)*

## RESUMO

O Exame de Proficiência para Professores de Línguas Estrangeiras (EPPLÉ) tem sido foco de muitas pesquisas no Brasil. Recentemente, desenvolveu-se o pré-teste ao teste oral do EPPLÉ com o intuito de fornecer aos candidatos uma informação preliminar sobre sua proficiência, que pudesse prever os resultados do exame. O objetivo deste trabalho é debater a possibilidade de se oferecer um pré-teste ao EPPLÉ oral, com validade de critério em relação a ele, e que seja disponível em dispositivos tecnológicos móveis. Assim, o pré-teste foi aprimorado em relação a uma versão anterior, e aplicado a dois grupos de professores de língua inglesa (em formação), um através de computadores e outro através de dispositivos tecnológicos móveis. Parte dos participantes realizou também o EPPLÉ. Nossos dados provêm dos resultados dessas aplicações, de observações de campo nesses momentos, e de um questionário aplicado aos participantes logo após, em que respondem a perguntas sobre suas impressões gerais do pré-teste e também de seu meio de aplicação (computadores ou celulares). Em nossa análise, avaliamos os resultados do pré-teste em comparação aos do EPPLÉ buscando evidências de validade preditiva e paralela. Além disso, investigamos os potenciais desafios e vantagens em se utilizar as tecnologias propostas para a aplicação do teste. Os dados apontaram para um paralelismo, embora não totalmente direto, entre os resultados do EPPLÉ oral e do pré-teste, o que sugere que o pré-teste possa ser usado como instrumento de previsão. Além disso, não houve diferença significativa nos resultados das aplicações através de celulares e de computadores, embora haja comentários dos participantes específicos a cada tecnologia, descrevendo vantagens e desvantagens. Os resultados, portanto, apontam para que seja possível produzir um pré-teste válido e disponível em celulares, apesar das limitações encontradas. Nossa análise nos levou também a possíveis caminhos para aprimorar ainda mais o pré-teste, a partir da versão desenvolvida neste trabalho, para que possa ser um instrumento cada vez mais confiável de previsão dos resultados do EPPLÉ oral.

**Palavras-chave:** avaliação de proficiência; tecnologia; formação de professores; línguas estrangeiras

## ABSTRACT

The Proficiency Exam for Foreign Language Teachers (EPPL) has been the subject of much research in Brazil. Recently, a pretest to the oral test of EPPL was developed in order to provide preliminary information on the candidates' proficiency that could predict the results of the exam. The objective of this work is to discuss the possibility of offering a pretest to the EPPL oral test, with criterion validity, and available on mobile devices. Thus, the pretest was improved over an earlier version, and taken by two groups of Brazilian English teachers (in pre-service education). The first group took the test through computers and the second through smartphones. Part of these participants also took the EPPL examination. Our data is provided by the scores in both tests, field observations during the applications, and a survey answered by participants shortly after taking the pretest, in which they are asked about their general view of the pretest and also about the means of application (computers or mobile devices). In our analysis, we compared results of the pretest to the EPPL's, looking for evidence of predictive and parallel validity. In addition, we investigated potential challenges and advantages of using the proposed technologies in the application of the pretest. The results pointed to a parallel, although not quite direct, relationship between the results of the oral EPPL and the pretest, which suggests that the pretest can be used as a prediction tool. Besides, there was no significant difference in the results of the applications through cell phones and computers, although in the surveys the participants pointed out characteristics of each technology, describing advantages and disadvantages. The results therefore indicate that it is possible to develop a valid pretest that is also available on mobile devices, despite the limitations found during the study. Our analysis has also directed us to some possibilities to improve the pretest even more, based on the version developed in this research, so that it can be a more reliable prediction instrument for the oral test in the EPPL examination.

**Keywords:** proficiency assessment; technology; teacher development; foreign languages

## SUMÁRIO

<b>INTRODUÇÃO.....</b>	<b>15</b>
Tema de pesquisa.....	16
Justificativa e relevância do tema.....	19
Objetivos e perguntas de pesquisa.....	20
Organização da dissertação.....	21
<b>1 METODOLOGIA.....</b>	<b>23</b>
<b>1.1 Natureza da pesquisa.....</b>	<b>24</b>
<b>1.2 O EPPLE .....</b>	<b>27</b>
<b>1.3 Desenvolvimento e alocação do pré-teste.....</b>	<b>28</b>
<b>1.4 Faixas de proficiência do pré-teste.....</b>	<b>35</b>
<b>1.5 Perfil dos participantes.....</b>	<b>38</b>
1.5.1 Grupo MG.....	39
1.5.2 Grupo SP <sub>1</sub> .....	40
1.5.3 Grupo SP <sub>2</sub> .....	40
<b>1.6 Coleta de dados.....</b>	<b>41</b>
1.6.1 Observações de campo.....	41
1.6.2 Aplicações e questionários.....	42
1.6.3 Resultados das avaliações.....	45
<b>1.7 Procedimentos de análise.....</b>	<b>47</b>
<b>2 FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>49</b>
<b>2.1 Proficiência linguística e avaliação.....</b>	<b>50</b>
2.1.1 Proficiência linguístico-comunicativo-pedagógica.....	52
<b>2.2 Pré-testagem e testes diagnósticos.....</b>	<b>53</b>
2.2.1 Associação entre avaliações de línguas.....	55
2.2.2 Associação entre avaliações de construtos distintos.....	58
<b>2.3 Avaliação e tecnologia.....</b>	<b>61</b>
2.3.1 CALT – <i>Computer-Assisted Language Testing</i> .....	64
2.3.2 Avaliação <i>web-based</i> .....	69
<b>2.4 Validade.....</b>	<b>72</b>
2.4.1 Validades de critério.....	76

<b>3 ANÁLISE E DISCUSSÃO DE DADOS.....</b>	<b>79</b>
<b>3.1 O pré-teste e seus resultados.....</b>	<b>80</b>
3.1.1 Impressões sobre o teste.....	80
3.1.2 Resultados e validade de critério.....	85
3.1.3 Validade paralela.....	87
3.1.4 Validade preditiva.....	91
3.1.5 A associação entre os dois instrumentos.....	93
<b>3.2 Uso da tecnologia.....</b>	<b>100</b>
3.2.1 Aplicações no computador.....	100
3.2.2 Aplicações no celular.....	103
3.2.3 Comparação entre PT2 no computador e no celular.....	107
<b>3.3 Discussão e encaminhamentos.....</b>	<b>110</b>
3.3.1 Sobre a pesquisa.....	110
3.3.2 Sobre o pré-teste (PT2).....	112
<b>CONSIDERAÇÕES FINAIS.....</b>	<b>116</b>
<b>REFERÊNCIAS.....</b>	<b>121</b>
<b>ANEXOS.....</b>	<b>128</b>
<b>Anexo 1 - Modelo do Termo de Consentimento.....</b>	<b>129</b>
<b>Anexo 2 - Modelo do questionário.....</b>	<b>131</b>
<b>Anexo 3 - Script do pré-teste 2.....</b>	<b>135</b>
<b>Anexo 4 – Construto do EPPLE.....</b>	<b>137</b>
<b>Anexo 5 – Faixas de proficiência do Pré-teste 2.....</b>	<b>138</b>
<b>Anexo 6 – Faixas de proficiência do EPPLE oral.....</b>	<b>140</b>

## LISTA DE QUADROS

<b>Quadro 1</b> – Comparação entre PT1 e PT2.....	31
<b>Quadro 2</b> – Faixas de proficiência do PT2.....	36-37
<b>Quadro 3</b> – Características gerais dos participantes da pesquisa.....	41
<b>Quadro 4</b> – Aplicações do pré-teste e do EPPLÉ.....	42
<b>Quadro 5</b> – Correspondência esperada entre escalas de proficiência do PT2 e EPPLÉ.....	47
<b>Quadro 6</b> – Descritores de competência comunicativa na produção escrita e oral.....	59
<b>Quadro 7</b> – Atributos do pré-teste em relação à tecnologia.....	65
<b>Quadro 8</b> – Relação entre descritores das faixas de proficiência do pré-teste e do EPPLÉ oral.....	85-86
<b>Quadro 9</b> – Resultados – Validade paralela – Pré-teste como referência.....	87-88
<b>Quadro 10</b> – Resultados – Validade Paralela – EPPLÉ <sub>1</sub> oral como referência	90
<b>Quadro 11</b> – Resultados – Validade Preditiva – Pré-teste como referência....	91-92
<b>Quadro 12</b> – Resultados – Validade Preditiva – EPPLÉ <sub>2</sub> oral como referência.....	93-94
<b>Quadro 13</b> – Relações globais de resultados do pré-teste e do EPPLÉ oral com base nas análises.....	96
<b>Quadro 14</b> – Previsões do pré-teste ao EPPLÉ oral com base nos resultados analisados.....	96
<b>Quadro 15</b> – Síntese entre previsões baseadas na análise de dados e na análise teórica das faixas de proficiência.....	98
<b>Quadro 16</b> – Sugestões e críticas ao pré-teste no computador.....	101-102
<b>Quadro 17</b> – Sugestões e críticas ao pré-teste no celular.....	104-106
<b>Quadro 18</b> – Síntese dos principais problemas nas aplicações via computador e via celular.....	109

## LISTA DE ILUSTRAÇÕES

<b>Figura 1</b> – Tela do PT1.....	32
<b>Figura 2</b> – Tela do PT2 em computador.....	32
<b>Figura 3</b> – Telas do PT2 em celular.....	33
<b>Figura 4</b> – Participantes da pesquisa.....	39
<b>Figura 5</b> – Construtos do EPPL oral e do PT2.....	58
<b>Figura 6</b> – Tipos de validade.....	75
<b>Gráfico 1</b> – Impressões sobre o pré-teste – MG.....	81
<b>Gráfico 2</b> – Impressões sobre o pré-teste – SP.....	82

## LISTA DE ABREVIATURAS E SIGLAS

**DTMs** – Dispositivos Tecnológicos Móveis (celulares, *tablets*...)

**EPPLE** – Exame de Proficiência para Professores de Línguas Estrangeiras

**LI** – Língua Inglesa

**MG** – grupo que contém 13 participantes que realizaram o pré-teste em computador e o EPPEL<sub>1</sub> em uma universidade pública em MG<sub>s</sub>.

**MG<sub>s</sub>** - subgrupo de MG, com 7 de seus participantes, que realizou também o EPPEL<sub>2</sub>.

**PT1** – Primeira versão do pré-teste (AGUENA, 2016)

**PT2** – Segunda versão do pré-teste, desenvolvida no presente trabalho e utilizada na coleta de dados (também nos referimos a essa versão simplesmente como “pré-teste”)

**PT3** – Terceira versão do pré-teste, apenas idealizada a partir das limitações do PT2

**QCER** – Quadro Comum Europeu de Referência

**SP** – grupo que inclui todos os 19 componentes dos grupos SP<sub>1</sub> e SP<sub>2</sub>, cuja característica comum é terem realizado o pré-teste através de aparelhos celulares.

**SP<sub>1</sub>** – grupo composto de 8 participantes que realizaram apenas o pré-teste em celular.

**SP<sub>2</sub>** – grupo composto de 11 participantes que realizaram o pré-teste em celular e o EPPEL<sub>1</sub>.

# **INTRODUÇÃO**

## TEMA DE PESQUISA

As avaliações, de modo geral, permeiam os mais variados aspectos da nossa vida pessoal e profissional. Em nossa vida pessoal, tomamos decisões o tempo todo com base em como avaliamos determinadas situações. Por outro lado, as avaliações institucionalizadas são muitas vezes relacionadas aos contextos profissionais ou estudantis. Nesses âmbitos, avaliações muitas vezes fazem o papel de “*gatekeeping*” para oportunidades de trabalho e estudo, tais como para a concessão de certificados e diplomas, seleção para instituições de ensino superior, aceitação de candidatos a uma vaga de emprego, dentre outras. Assim, é de suma importância que tais avaliações sejam tão apropriadas e justas quanto for possível, o que comumente significa dizer, no jargão da área, que sejam válidas e confiáveis.

Para que se mantenha adequada a seus propósitos, uma avaliação deve ser constantemente pesquisada, e, quando necessário, ajustada. Tais pesquisas e ajustes chamamos **processo** de validação porque assim o é, sendo impossível alcançar o ponto em que podemos dar o trabalho por terminado. Uma avaliação sempre poderá e deverá ser adaptada e melhorada face ao surgimento de novas circunstâncias, novo público de examinandos, novo uso, dentre outras possibilidades.

Em se tratando de linguagem, é uma área relativamente nova no Brasil a do desenvolvimento de testes de proficiência linguística (CONSOLO, 2004; CONSOLO et al, 2009; SCARAMUCCI, 2011). Avaliar o conhecimento linguístico é uma empreitada ambiciosa e complexa, visto que o processo de produção de uma boa avaliação exige um longo trabalho de pesquisa, e o objeto de avaliação, a própria linguagem, é multifacetada e se inter-relaciona com vários outros aspectos da vida em sociedade.

O Exame de Proficiência para Professores de Línguas Estrangeiras (EPPE) é um exame desenvolvido por pesquisadores brasileiros para avaliar a proficiência linguística de professores de línguas no Brasil (CONSOLO et al, 2009). O exame de língua inglesa (LI) é composto de dois testes, um escrito e um oral, e ambos têm sido foco de muitos estudos nos últimos anos, visando seu aprimoramento cada vez maior (BAFFI-BONVINO, 2010; ANCHIETA, 2015; FERNANDES, 2016, dentre outros). A proposta vem ao encontro de uma demanda por uma certificação linguística para professores de línguas e o estabelecimento de um padrão do que se espera em termos de proficiência linguística de um bom profissional nesta carreira.

Além disso, o EPPLÉ surgiu em um contexto brasileiro de formação linguística precária de professores de línguas. Tratando-se em especial de professores de LI, sua formação é considerada inadequada em uma parcela significativa dos cursos de licenciatura, tanto em seu aspecto pedagógico quanto em seu aspecto linguístico (ALMEIDA FILHO, 1992; CELANI, 2010, 2008; CONSOLO et al, 1999). A boa formação linguística destes professores é essencial para o sucesso do ensino, uma vez que a língua é (ou deveria ser) o instrumento e o conteúdo das aulas de línguas. Para um bom desempenho de suas atividades, é imperativo que estes profissionais tenham nível de proficiência adequado da língua que ensinam(rão), necessária para que possam produzir insumo linguístico oral e escrito, atuando como modelo para seus alunos e promovendo o uso da língua estrangeira em sala de aula, o que contribuiria para o desenvolvimento linguístico desses alunos (CONSOLO, 2006a). Dessa forma, considerando-se o impacto que instrumentos de avaliação possuem sobre contextos formativos (efeito retroativo), é esperado que o EPPLÉ tenha uma influência positiva na formação desses professores.

O EPPLÉ busca avaliar a proficiência linguística do professor considerada tanto em seu aspecto global como em aspectos específicos da profissão de professor de línguas, tais como utilizar a língua estrangeira para dar instruções, esclarecer conteúdos, dentre outros. O conjunto destas proficiências é chamado de proficiência linguístico-comunicativo-pedagógica por Consolo & Silva (2014), e constitui o domínio do exame EPPLÉ.

Além das mencionadas pesquisas sobre o EPPLÉ que foram desenvolvidas no decorrer dos anos, especialmente no âmbito do projeto de pesquisa “Avaliação da proficiência linguístico-comunicativa-pedagógica do professor de línguas: operacionalização de construto no Exame de Proficiência para Professores de Língua Estrangeira (EPPLÉ)”, o trabalho de iniciação científica de Agüena (2016) introduziu a proposta de um pré-teste para o exame. O pré-teste seria um pequeno teste que indicaria um resultado preliminar ao candidato, permitindo que este tomasse uma decisão informada sobre sua participação no exame.

A crescente onda de avanços tecnológicos teve um impacto semelhante tanto na avaliação quanto no ensino e aprendizagem de línguas. No ensino, sua utilização tem sido pesquisada extensivamente, dada a enorme variedade de possibilidades e o grande potencial positivo na aprendizagem. No entanto, a tecnologia por si só não melhora o ensino e a aprendizagem de línguas, de modo que muitos trabalhos também foram publicados sobre os novos papéis do professor de línguas frente a essas mudanças e a importância da adequação do uso da tecnologia nos diversos contextos (PAIVA, 2001; TUDOR, 2002; KENSKI, 2007,

dentre outros). Em avaliação, que é o nosso foco, são duas as principais correntes de estudos com tecnologia: eficiência e inovação (CHAPELLE & VOSS, 2016). Em nosso caso, a eficiência está na busca por acessibilidade e praticidade no formato de entrega do teste, e a inovação está na utilização de uma plataforma compatível com *smartphones*, ainda pouco utilizados para a aplicação de testes.

Sobre a utilização de tecnologias móveis, a UNESCO publicou em 2014 as *Diretrizes de políticas para a aprendizagem móvel*, o que é mais um passo em direção à integração desse tipo de dispositivo no ensino e aprendizagem. Há vários trabalhos utilizando dispositivos tecnológicos móveis (DTMs), por exemplo, na educação à distância (VIBERG, 2015), ou em escolas públicas brasileiras (COSTA, 2013), ou ainda analisando a utilização de aplicativos de ensino (GUO, 2014). No entanto, encontramos poucos trabalhos sobre o uso de DTMs na avaliação de línguas, chamado de *Mobile-Assisted Language Assessment* (por exemplo, TARIGHAT; KHODABAKHSH, 2016 e LABORDA; ROYO; BAKIEVA, 2016).

Neste estudo, portanto, o pré-teste do EPPLÉ é nosso objeto. Partindo de uma versão inicial desenvolvida por Agüena (2016), revisamos seu desenho para permitir maior acessibilidade e praticidade aos candidatos para obterem sua avaliação de proficiência prévia ao EPPLÉ. Uma vez que o pré-teste é um teste de baixo impacto e com o propósito apenas de informar o candidato, buscamos maximizar esses aspectos (acessibilidade e praticidade) por meio do uso de tecnologias presentes em nosso cotidiano, como computadores e aparelhos celulares ou *tablets*.

Assim, nossa pesquisa gira em torno de um questionamento sobre se é possível oferecer um pré-teste ao EPPLÉ oral que possua validade de critério e que possa ser acessado através de computadores ou DTMs. Iniciamos o trabalho desde a elaboração dessa nova versão do pré-teste, sua pilotagem utilizando tanto computadores como DTMs, sua comparação com resultados obtidos no próprio EPPLÉ e, por fim, informações coletadas pela observação das aplicações e de questionários aplicados aos participantes após a realização do pré-teste. Não perdemos de vista durante o projeto os fatores “acessibilidade” e “praticidade”, que resultaram em algumas limitações, entre elas a própria seleção da plataforma de hospedagem do pré-teste.

Antes de iniciarmos as próximas seções do texto, cabe ressaltar que nosso trabalho irá tratar de três versões diferentes do pré-teste. Durante o texto, o termo “pré-teste”, desacompanhado de qualquer especificação, irá remeter à segunda versão, por nós desenvolvida neste trabalho e que foi utilizada nas aplicações cujos resultados utilizamos

como dados. Para organização textual, assim chamaremos as versões, quando houver necessidade de fazer distinção entre elas:

- pré-teste 1 (PT1) é a versão original desenvolvida por Agüena (2016);
- pré-teste 2 (PT2) é a versão inicialmente desenvolvida neste trabalho, que foi analisada mais extensivamente, pois com ela realizamos nossa coleta de dados;
- e pré-teste 3 (PT3) é a proposta feita ao final desta pesquisa, como encaminhamento, com base nas limitações encontradas durante a análise do PT2.

## **JUSTIFICATIVA E RELEVÂNCIA DO TEMA**

Nossa pesquisa se posiciona, conforme mencionado anteriormente, em um cenário de estudos sobre avaliação de proficiência linguística, tendo como público específico professores de LI em formação pré-serviço e em serviço no Brasil. É sabido que as avaliações têm poder de gerar o chamado “efeito retroativo”, um conceito complexo, que diz respeito, basicamente, “ao impacto que uma avaliação tem no ensino e aprendizagem feitos em sua preparação”<sup>1</sup> (GREEN, 2013, p. 40, tradução nossa). O termo já foi bastante discutido e ampliado, e aqui consideramos efeito retroativo não só as consequências que uma avaliação tem nas práticas de instrução, mas também na administração de uma instituição de ensino, nas atitudes e hábitos de professores e alunos e, mais amplamente, na sociedade em geral.

De acordo com Consolo (2017), o EPPL foi elaborado tendo entre suas motivações a geração de um efeito retroativo positivo e redirecionador na formação de professores, oferecendo à classe parâmetros cientificamente embasados do que se espera de um bom professor de LI. Assim, embora não seja possível prever o efeito retroativo de um teste com segurança, acreditamos que as aplicações do exame poderiam colaborar com o rompimento de um círculo vicioso de formação precária de professores nas universidades, e posterior ensino insuficiente nas escolas, que traria para a universidade estudantes com baixa proficiência, completando-se o círculo (CONSOLO, 2017, p. 64). O exame pode ser considerado de alta relevância no sentido de ter abrangência nacional e ser respeitado entre pesquisadores e professores. No entanto, ainda não foi possível divulgá-lo extensivamente, em todos os estados do país, ou, indo mais além, institucionalizá-lo como certificação da classe de

---

<sup>1</sup> No original: “(...) the impact that a test has on the teaching and learning done in preparation for it”.

professores de LI. Assim sendo, o efeito retroativo e o impacto do EPPLÉ podem ainda ser aumentados consideravelmente.

O pré-teste acrescenta ao EPPLÉ uma série de vantagens, dentre elas o oferecimento de um “pacote” contendo o exame principal (EPPLÉ), de alta relevância, e um teste de baixa relevância, fácil acesso e pequena duração vinculado ao principal. Além disso, estudar a validade do pré-teste em relação ao EPPLÉ contribui para a melhoria de ambos os testes e suas faixas, dado que pressupõe um detalhamento e definição cada vez mais precisos de seus construtos. Outra vantagem seria aumentar a visibilidade do EPPLÉ, devido à acessibilidade do pré-teste e facilidade de sua divulgação, proporcionando ao público um indicador dos conteúdos avaliados no EPPLÉ.

Assim, este trabalho vem somar-se aos estudos do EPPLÉ, tornando-o mais completo, e, portanto, contribuindo para o desenvolvimento de avaliações que possam vir a gerar mudanças sociais nos cursos de formação de professores de LI no Brasil (modificações nos currículos, nas práticas de formadores, nos hábitos e atitudes dos professores em formação, dentre outras) e na sociedade (por exemplo, aumentar o respeito pela profissão por meio de certificação e uma identificação com a excelência).

## **OBJETIVOS E PERGUNTAS DE PESQUISA**

### **Objetivo geral**

O objetivo geral deste trabalho é investigar a possibilidade de elaboração e oferecimento de um pré-teste para o EPPLÉ que possua validade de critério e que seja compatível com dispositivos tecnológicos móveis. Assim, em primeiro lugar, buscamos desenvolver e operacionalizar o pré-teste para prever os resultados do EPPLÉ, e analisar sua validade para este propósito. Além disso, pretendemos buscar alternativas e/ou soluções tecnológicas aos problemas encontrados em trabalhos anteriores que envolveram o pré-teste (AGUENA, 2016; CONSOLO & AGUENA, 2017). Por fim, a utilização da tecnologia está também focada na inovação, uma vez que desenvolvemos um pré-teste acessível através de computadores e de DTMs, procurando analisar o impacto do uso dessas tecnologias na validade do pré-teste.

## **Objetivos específicos**

Os objetivos específicos desse trabalho podem ser assim descritos:

- 1) buscar alternativas, por leitura exploratória de artigos na área, para a disponibilização do pré-teste em dispositivos tecnológicos móveis;
- 2) identificar, com a utilização de questionários aplicados após o pré-teste, aspectos levantados pelos candidatos de sua experiência com estas avaliações (especialmente o próprio uso da tecnologia), que servirão também como evidência de (falta de) validade;
- 3) analisar os resultados de aplicações do pré-teste e do EPPLÉ em busca de evidências que sustentem (ou não) sua validade.

## **Perguntas de pesquisa**

Considerando os objetivos descritos, trazemos aqui a formulação das perguntas de pesquisa. A pergunta central é: é possível o oferecimento de um pré-teste com validade de critério em relação ao EPPLÉ e disponível em dispositivos tecnológicos móveis?

As perguntas que se provêm dos objetivos específicos são as seguintes:

- 1) De que maneira é possível oferecer o pré-teste em dispositivos tecnológicos móveis?
- 2) Quais impactos da utilização das tecnologias foi possível observar nas aplicações?
- 3) Em que medida o pré-teste ao EPPLÉ é válido como instrumento de “previsão” dos resultados do exame?

## **ORGANIZAÇÃO DA DISSERTAÇÃO**

O presente trabalho encontra-se organizado em cinco grandes seções: introdução e os capítulos de metodologia, fundamentação teórica, análise e discussão dos dados e considerações finais. Na introdução apresentamos o contexto da pesquisa: o cenário brasileiro de professores de LI com baixa proficiência linguística, em que o EPPLÉ já foi proposto como ferramenta para avaliar essa proficiência e, potencialmente, exercer um efeito retroativo na formação de professores. Além disso, o objeto do estudo foi apresentado, que é a validade do pré-teste ao EPPLÉ, com ênfase especial ao uso de tecnologias em sua aplicação. Por fim, expusemos os objetivos, geral e específicos, e as perguntas de pesquisa.

No primeiro capítulo, sobre a metodologia da investigação, fornecemos informações relativas à natureza da pesquisa, os contextos de coleta de dados, os participantes envolvidos e os instrumentos e procedimentos de coleta e análise de dados. Incluímos também nessa seção as etapas anteriores à coleta de dados que se compõem da elaboração e instrumentalização do pré-teste (PT2). A ordem canônica de organização de trabalhos científicos – em que a fundamentação teórica precede a metodologia – foi aqui modificada com vistas a oferecer uma melhor visão do que se trata a pesquisa antes de partir para os textos teóricos que vieram a embasar o seu desenvolvimento. Para auxiliar a leitura, alguns termos mais específicos da área de avaliação foram definidos brevemente em rodapé.

No segundo capítulo, de fundamentação teórica, tratamos de temas que foram essenciais para o melhor entendimento do nosso objeto de estudos. Trazemos contribuições de autores que trataram de avaliações de proficiência, avaliações de diagnóstico, conceito de proficiência, validade de avaliações, desenvolvimento de instrumentos de avaliação, utilização de tecnologia em avaliação e faixas de proficiência.

A seguir, as teorias citadas no capítulo de fundamentação são utilizadas para guiar nosso olhar durante a análise e discussão dos dados. Apresentamos, inicialmente, uma análise dos resultados do pré-teste em comparação com os do EPPLE oral, tratando da validade de critério do pré-teste. Também analisamos aqui parte do questionário que trata da qualidade do pré-teste como instrumento de avaliação da proficiência do professor de LI. Em seguida, tratamos da utilização de tecnologias na aplicação, contrastando as duas utilizadas (computadores e DTMs) e buscando semelhanças e diferenças por meio da análise de relatórios de observação das aplicações e das respostas dadas pelos participantes aos questionários após a aplicação. Por fim, criamos diferentes subgrupos de participantes, buscando relacionar os aspectos tratados anteriormente em busca de padrões ou relações. Ainda nesse capítulo, trazemos uma discussão dos dados envolvendo as limitações encontradas durante o desenvolvimento do trabalho, e propomos, como encaminhamento, o que poderia ser uma terceira versão do pré-teste (PT3) a partir dos desafios encontrados em nosso percurso.

Tecemos, finalmente, considerações finais sobre a pesquisa, em que sintetizamos os resultados obtidos pela investigação, retomamos suas limitações, respondemos às perguntas de pesquisa e sugerimos encaminhamentos para pesquisas futuras.

# **METODOLOGIA**

## 1.1 NATUREZA DA PESQUISA

Além de acompanharmos o desenvolvimento do pré-teste, e a postulação de um novo modelo, este trabalho faz duas análises distintas, porém interligadas, do pré-teste ao EPPL. Propusemos um pré-teste para esta pesquisa, que é uma segunda versão do original de Agüena (2016), e nele analisamos dois aspectos principais: sua validade de critério (preditiva e paralela)<sup>2</sup> em relação ao EPPL e os desafios tecnológicos de sua implementação. Uma vez delimitados estes dois objetos (ou este objeto – o pré-teste – observado de dois ângulos), o nosso olhar sobre eles, metodologicamente falando, deve ser descrito.

Na área de avaliação, de acordo com Lazaraton (2004), tradicionalmente foram feitas muitas pesquisas de caráter quantitativo, baseadas em uma visão psicométrica dos comportamentos e em uma postura positivista de análise. Na Linguística Aplicada (LA) não foi diferente, e a área de avaliação pode inclusive ser descrita como a de orientação mais positivista dentre todas as áreas dessa disciplina, vista muitas vezes como politicamente e intelectualmente conservadora, além de ser também uma das mais envolvidas com questões de poder (MCNAMARA, 2001; SHOHAMY, 1998).

No entanto, conforme McNamara (2001), com o passar do tempo, cresceu a consciência do uso da linguagem como atividade social, e da avaliação como prática que deve ser olhada criticamente, uma vez que provém de pressuposições e julgamentos de valor subjetivos, e, potencialmente, tendenciosos. Além disso, foi tornando-se geral entre os pesquisadores um reconhecimento das limitações dos métodos estatísticos tradicionais em avaliação de línguas, especialmente no que diz respeito à validação de instrumentos de avaliação.

Os métodos qualitativos vieram, assim, apresentar um novo entendimento do **processo** de avaliação em si, para além do entendimento do produto, já fornecido por métodos quantitativos. Os avanços deste olhar qualitativo na área de avaliação em línguas são relativamente recentes e Lazaraton (2004) os considera como o mais importante desenvolvimento metodológico no campo no início dos anos 2000. A metodologia qualitativa, desde sua introdução neste campo de pesquisa, passou a ser utilizada para variados propósitos, como o desenho, a descrição e a validação de testes.

---

<sup>2</sup> Validade diz respeito ao quanto um teste é apropriado para a interpretação e uso feitos dele. Dentre as subdivisões do termo, exploradas no capítulo seguinte, a validade de critério é a que toma um instrumento já existente como referência para validar outro. Para isso, os dois testes e seus resultados são comparados. A validade de critério pode ser do tipo preditiva, com um intervalo maior de tempo entre as aplicações dos dois instrumentos, buscando ver se o instrumento a ser validado tem poder de prever o outro; ou pode ser do tipo paralela, em que as aplicações acontecem sem intervalo de tempo considerável.

Rahman (2017) traz uma discussão sobre vantagens e desvantagens de métodos qualitativos e quantitativos na avaliação de línguas. Inicialmente, o autor traz definições de vários outros estudiosos, dentre as quais destacamos a de Strauss e Corbin (1990), que diz que a pesquisa qualitativa é

Qualquer tipo de pesquisa que produz achados não alcançados através de procedimentos estatísticos ou outros meios de quantificação. Pode referir-se à pesquisa sobre as vidas das pessoas, experiências vividas, comportamentos, emoções e sentimentos assim como funcionamento organizacional, movimentos sociais, fenômenos culturais e interações entre nações.<sup>3</sup> (STRAUSS e CORBIN, 1990, p. 11 apud. RAHMAN, 2017, p. 103, tradução nossa)

Assim, a pesquisa qualitativa preocupa-se com o fator humano e social de seus objetos de estudo, e utiliza, para estudá-lo, métodos não baseados em quantificações. A pesquisa qualitativa traz à tona aspectos sociais e humanos que a pesquisa quantitativa deixa em segundo plano.

Nem todos os autores concordam com a utilização de “qualitativo” como termo guarda-chuva, em contraposição a “quantitativo”. Erickson (1986), por exemplo, escreve sobre a pesquisa na área de ensino e aprendizagem utilizando como termo maior o “interpretativismo” para se referir a abordagens de pesquisa de observação de participantes. Ele justifica essa escolha com três razões: a) é um termo inclusivo; b) não define tais abordagens como essencialmente não-quantitativas; e c) aponta uma semelhança compartilhada pela família de abordagens incluídas sob este termo, que é a centralização do interesse da pesquisa no “significado humano na vida social e em sua elucidação e exposição pelo pesquisador”<sup>4</sup> (ERICKSON, 1986, p. 119, tradução nossa).

Métodos interpretativos, ainda de acordo com Erickson (1986), são mais adequados quando se busca: a) um entendimento da estrutura de ocorrências, ao invés de sua distribuição geral; b) as significações atribuídas por atores sociais em determinados eventos; c) a localização de pontos de contraste, quando não se pode controlar influências do contexto; d) a determinação de relações causais não identificadas em pesquisas quantitativas, ou o desenvolvimento de novas teorias sobre as causas e influências nos padrões apontados por aquela.

---

<sup>3</sup> No original: “(...) any type of research that produces findings not arrived at by statistical procedures or other means of quantification. It can refer to research about persons’ lives, lived experiences, behaviours, emotions, and feelings as well as about organisational functioning, social movements, cultural phenomena, and interactions between nations”

<sup>4</sup> No original: “(...) human meaning in social life and in its elucidation and exposition by the researcher”

Dessa forma, em nosso trabalho, buscamos entender um pouco mais sobre a perspectiva dos participantes ao se depararem com os dois instrumentos avaliativos relacionados: pré-teste e EPPL oral. Coletamos essa perspectiva de três maneiras: através dos próprios resultados dos testes dos participantes; através de observações de campo durante as aplicações das avaliações; através de questionários aplicados após o pré-teste.

Utilizamos três instrumentos para a coleta de dados nos baseando no conceito de triangulação de Flick (2009). Para ele, triangulação significa trazer diferentes perspectivas para uma questão que está sendo estudada, com o objetivo não de esgotar o objeto de estudo, mas de obter uma compreensão mais completa sobre esse objeto. A triangulação é vista, desde sua proposição, como um aspecto que contribui para a qualidade da pesquisa qualitativa. Flick (2009) traz ainda o trabalho de Denzin (1989), que fez uma subdivisão dos tipos de triangulação, para mostrar como esta pode ser feita em diferentes focos da pesquisa. Assim, temos a triangulação de dados, que se trata de utilizar diferentes fontes de dados; a triangulação de investigadores, que é a comparação e inter-relação dos olhares de vários investigadores ao colocá-los para fazer observações e entrevistas; a triangulação de teorias, que diz respeito a abordar os dados por diferentes perspectivas; e a triangulação metodológica, que é feita por meio da utilização de mais de um método de coleta ou análise de dados.

Em nosso trabalho, a coleta de dados foi feita de três maneiras distintas (observações de campo, resultados das avaliações, questionários) e em dois ambientes diferentes (uma universidade pública no estado de São Paulo, e uma no estado de Minas Gerais). Dessa maneira, nossa coleta de dados configura uma triangulação metodológica. Nossos dois contextos de estudo apresentam subdivisões, sendo possível afirmar que trabalhamos também com a triangulação de dados.

Chamamos, por fim, a natureza de nossa pesquisa de “interpretativa”, levando em conta os autores discutidos nesse capítulo e suas contribuições para a definição da nossa metodologia. “Qualitativo” é o termo mais conhecido e utilizado, geralmente definido como pesquisa na qual se faz uso de poucos dados numéricos e estatísticos, ou mesmo nenhum. “Interpretativa”, por sua vez, é um termo que enfatiza o caráter do olhar do pesquisador ao tentar compreender os significados da vida social dos sujeitos estudados. Assim, por nos propormos a interpretar, analisar e criticar um instrumento de avaliação levando em conta esses significados, acreditamos que a melhor nomenclatura é de fato “interpretativa”.

Com base na análise desses dados, podemos nos voltar novamente à pergunta de pesquisa e discuti-la com mais propriedade, por meio de um viés subjetivo que permita interpretações do pesquisador como observador crítico do objeto de pesquisa. Além disso, nosso trabalho trata de instrumentos avaliativos de proficiência linguístico-comunicativo-pedagógica com vistas à sua implementação futura, que pode ter um impacto no desempenho linguístico dos potenciais candidatos, e também efeitos em sua vida profissional, social e pessoal. Assim, é de suma importância manter um olhar humano sobre o objeto.

## **1.2 O EPPLE**

Nessa seção, descreveremos brevemente o exame EPPLE, utilizado em nosso trabalho como referência para descrevermos aspectos da validade de critério do PT2.

O EPPLE é o Exame de Proficiência para Professores de Línguas Estrangeiras, que vem sendo desenvolvido no âmbito do grupo de pesquisa “Ensino e Aprendizagem de Língua Estrangeira: Crenças, Construtos e Competências” (ENAPLE-CCC). O exame foi pensado para avaliar a proficiência linguístico-comunicativo-pedagógica de professores de línguas, oferecendo-lhes um parâmetro do que é considerada uma boa proficiência linguística do professor.

Subdividido em duas seções, o teste escrito e o teste oral, o EPPLE é aplicado em dois diferentes momentos. Inicialmente, o teste oral é composto de uma entrevista, em pares, dos candidatos. São também dois os examinadores, um responsável diretamente pela condução da entrevista e o outro direciona sua atenção ao discurso dos candidatos e a anotações sobre seus pontos fortes e fracos, em relação aos descritores das faixas de proficiência. O teste oral se inicia com uma conversa pessoal sobre os próprios participantes, em que eles falam um pouco sobre sua formação, origem, e experiência como professores. A segunda parte do teste se dá em torno de um pequeno vídeo, assistido pelos participantes e examinadores no momento da aplicação, e discutido em questões subsequentes que permitem elucidar a compreensão oral dos participantes. Em seguida, aos participantes é mostrado um *slide* contendo duas dúvidas comuns de estudantes de língua inglesa. Cada participante fica com uma das dúvidas, negociando entre si qual, e então responde a essa dúvida como se fora o professor do aluno hipotético. Por fim, o examinador pergunta aos participantes qual foi sua parte preferida do teste, e por quê. Uma vez que a entrevista é finalizada, em, aproximadamente, 20 minutos, os examinadores discutem o posicionamento dos participantes nas faixas, e, se necessário, encaminham uma gravação do teste a um terceiro examinador.

O teste escrito é composto de três seções, ambas presentes nas duas versões do EPPLÉ disponíveis até o momento, com pequenos ajustes. A primeira seção inclui um texto relacionado à área de ensino de línguas, acompanhado de questões interpretativas. Na segunda seção são apresentados excertos ou textos de estudantes de língua inglesa, contendo desvios da língua padrão. O candidato é convidado a identificar as porções inadequadas e explicar o motivo de sua inadequação, recorrendo à metalinguagem da área. Por fim, o participante é convidado a redigir um pequeno texto, em situação contextualizada com o trabalho de professor de língua inglesa.

Além disso, o EPPLÉ pode ser aplicado em papel (como foi feito em nosso trabalho) ou em computadores, pois possui uma versão eletrônica.

### **1.3 DESENVOLVIMENTO E ALOCAÇÃO DO PRÉ-TESTE**

O pré-teste ao EPPLÉ é um pequeno teste, rápido e acessível<sup>5</sup>, que se propõe a prever o potencial resultado do aluno no EPPLÉ. No exame, conforme já mencionamos, há dois testes (um escrito e um oral), e cada um deles gera uma faixa de proficiência no relatório de desempenho do candidato. Até então, foi desenvolvido um pré-teste para prever apenas o resultado do EPPLÉ oral, mas há a intenção de produzir também um para o EPPLÉ escrito, futuramente.

Conforme propõe Bachman (1990), o caminho para o desenvolvimento de um teste se inicia pela determinação de seu construto, seguida da operacionalização deste em um instrumento de avaliação, que, por fim, deve ser analisado para a definição de regras e procedimentos de mensuração das informações coletadas através dele.

Para desenvolvermos o construto do pré-teste, nos voltamos para o construto do EPPLÉ, uma vez que desejamos alinhá-los. Publicado na tese de doutorado de Anchieta (2015, p. 255), o construto do EPPLÉ, que está anexado ao final deste trabalho, traz as seguintes características do exame:

- é um exame comunicativo, prevendo um uso contextualizado da língua, em situações reais, analisando o desempenho global do candidato, e expressando o resultado em faixas de proficiência.
- proficiência linguística é vista, segundo Stern (1983), como envolvendo (a) domínio de formas; (b) domínio de significados linguísticos, cognitivos, afetivos e

---

<sup>5</sup> A rapidez e acessibilidade foram conceitos norteadores desde a primeira elaboração do pré-teste: sua duração não ultrapassa 30 minutos e é um teste de livre acesso, agora inclusive através de DTMs.

socioculturais; (c) capacidade de uso com enfoque na comunicação; e (d) criatividade no uso da língua.

- são analisados um conjunto de competências linguísticas gerais e específicas desenvolvidas por professores, por meio de interações com colegas, alunos, dentre outros: as gerais incluem tratar de aspectos cotidianos, enquanto que as específicas abrangem o uso de metalinguagem e a adaptação discursiva do professor ao discurso de sala de aula.

O pré-teste foi criado sob limitações diversas, que delimitaram sua filiação aos aspectos citados acima: duração e extensão devem ser pequenos, o teste deve ser adaptável a celular e computador, limitações da própria plataforma (*Forms*, que não permite, por exemplo, gravação de áudio da parte do candidato), limitações das aplicações que fazem uso de dispositivos móveis (teclado, autocorretor). Ainda assim, o pré-teste é comunicativo na medida em que possui itens contextualizados e baseados em situações reais de uso da língua; desenvolve-se progressivamente em temas (inicialmente, perguntas pessoais; uma pergunta apresentando uma situação de ensino e, por fim, um texto oral com itens de compreensão); avalia o desempenho global do candidato; e tem seu resultado expresso em faixas de proficiência. Assim, o pré-teste possui características de um teste comunicativo e também de um teste diagnóstico, como veremos em seção posterior.

A operacionalização do construto do pré-teste se deu também tomando como base o PT1 (AGUENA, 2016), que já tinha sido desenvolvido em direção semelhante, mas disponível apenas em computadores. Assim, a migração de plataforma para possibilitar o oferecimento do PT2 em DTMs resultou em adaptações. A plataforma onde estava hospedado o PT1, *Lingt Classroom*, apresentava dois problemas. O primeiro problema, com base no estudo anterior de Consolo e Agüena (2017), é que a plataforma requeria a instalação em navegadores de internet do *Adobe Flash Player*, um software já obsoleto e incompatível com a maioria dos DTMs atuais. Algumas alternativas foram estudadas e gestores de plataformas web de teste foram contatados buscando suprir nossas necessidades e propósitos, porém foram descartadas a princípio por serem dispendiosas, com custos de produção e hospedagem do teste. No entanto, durante a escrita do projeto desta pesquisa, esse primeiro problema foi resolvido pelo próprio *Lingt Classroom*, que sofreu uma repaginação e atualização, sendo agora acessível via DTMs, e agora se chama apenas *Lingt*. Porém, tivemos um segundo motivo para não utilizá-lo: durante essa repaginação, houve a diminuição das possibilidades

de oferecimento de um teste gratuito, limitando o número possível de usuários, as possibilidades de questões, dentre outros, que seriam disponibilizados apenas para usuários *premium*.

Várias plataformas foram pesquisadas como alternativas ao *Lingt*, mas todas apresentaram limitações, sendo que, dentre as gratuitas, o *Google Forms* foi definitivamente o mais adequado aos nossos propósitos. Sua seleção decorre de algumas de suas características, como ser gratuito, oferecer o teste a um número de usuários ilimitado, ser uma ferramenta bastante intuitiva a desenvolvedores e usuários, permitir a elaboração de itens variados, oferecer automaticamente uma tabela com as respostas dos candidatos, e ser acessível de qualquer dispositivo conectado à internet (computadores e DTMs).

No entanto, como dito anteriormente, todas as plataformas pesquisadas apresentaram limitações, e com o *Forms* não foi diferente. A maior modificação durante a migração do pré-teste para o *Forms* foi a exclusão mandatória da gravação de áudio por parte do aluno, devido à sua indisponibilidade na plataforma. O pré-teste propõe-se a avaliar a proficiência oral em LI e, por isso, foi pensado inicialmente como um teste em que as tarefas envolveriam compreensão e produção oral. Entretanto, foi impossível prosseguir com esse projeto e tivemos que abrir mão da produção oral para garantir que o pré-teste fosse hospedado *online*, gratuitamente e disponível em DTMs. Desse modo, o pré-teste manteve as questões em áudio (compreensão oral), mas o formato das respostas passou a ser escrito, no computador ou celular. Portanto, a utilização do *Forms*, com as adaptações decorrentes dessa utilização, acarretou mudanças no construto original e nas faixas de proficiência do pré-teste.

Houve outras modificações durante a migração do pré-teste, desta vez na quantidade e conteúdo das questões. Mantivemos um número semelhante de itens, diminuindo apenas dois, passando de dez para oito questões. Tal decisão foi tomada visando uma manutenção na duração proposta da avaliação, que é de trinta minutos. No PT2, como as respostas são escritas, pressupõe-se que o candidato leve mais tempo redigindo-as do que levaria anteriormente no PT1, quando suas respostas eram gravadas em áudio. Além disso, mantivemos as três primeiras questões do pré-teste com tema mais geral, incluindo perguntas de cunho mais pessoal que permitem uma elaboração mais livre da parte do candidato. No entanto, modificamos a distribuição das cinco questões restantes: três passaram a ser de compreensão de um texto oral mais longo, e duas questões metalinguísticas, uma contextualizada como dúvida de um aluno e outra que requer a explicação de um uso linguístico no texto de compreensão oral. Uma vez que todas as questões foram gravadas em

áudio e o candidato não as lê, apenas escuta-as, todo o teste já avalia compreensão oral por meio da produção de respostas condizentes com as instruções de cada questão, que são apenas ouvidas, e não lidas. O *script* do PT2 encontra-se nos anexos deste trabalho (p. 130), e a seguir apresentamos o quadro 1, comparando PT1 e PT2.

**Quadro 1 – Comparação entre PT1 e PT2**

	<b>PT1</b>	<b>PT2</b>
Plataforma	<i>Lingt Classroom</i>	<i>Google Forms</i>
Número de questões	10	8
Compreensão/Produção	Compreensão oral/Produção oral	Compreensão oral/Produção escrita
Duração	30'	30'
Compatibilidade com DTMs	Não	Sim
Descritores das faixas de proficiência	Variação, precisão, fluência, interação, coerência	Variação, precisão, coerência/coesão, compreensão oral, metalinguagem

Neste quadro, já adiantamos aspectos que serão tratados nas próximas subseções, como as faixas de proficiência do pré-teste, com o objetivo de já oferecer ao leitor uma visão geral das semelhanças e diferenças entre essas duas versões do pré-teste.

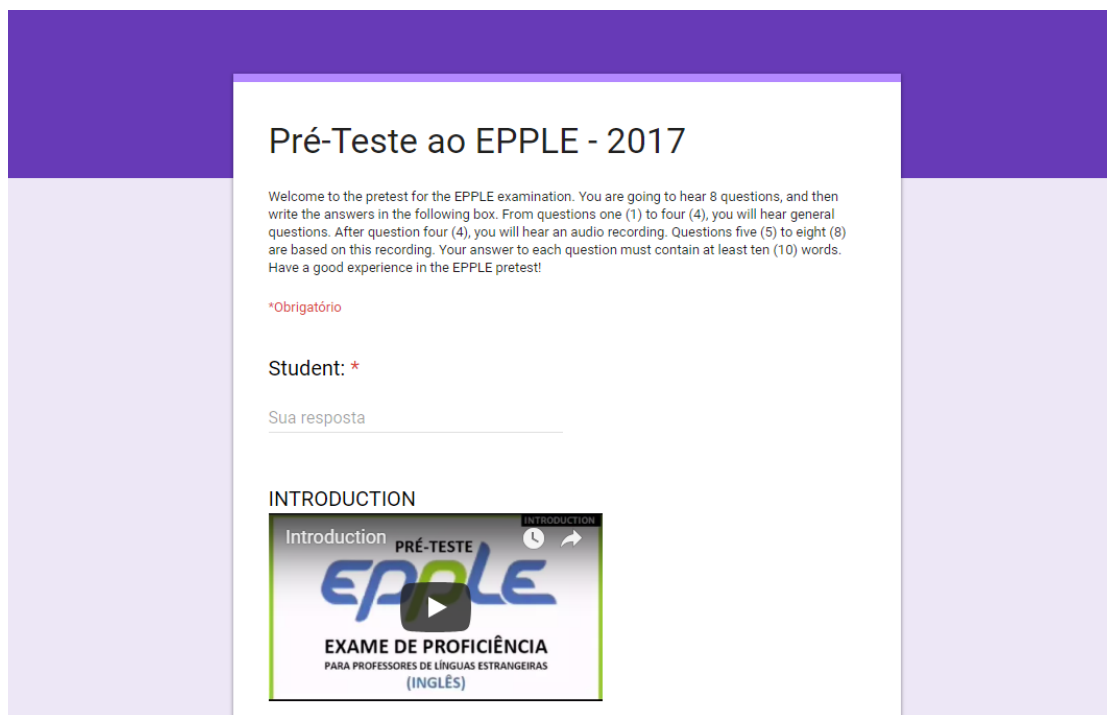
Trazemos também abaixo algumas telas do PT1 e do PT2 para efeito de comparação e visualização inicial dos dois instrumentos.

**Figura 1 – Tela do PT1**



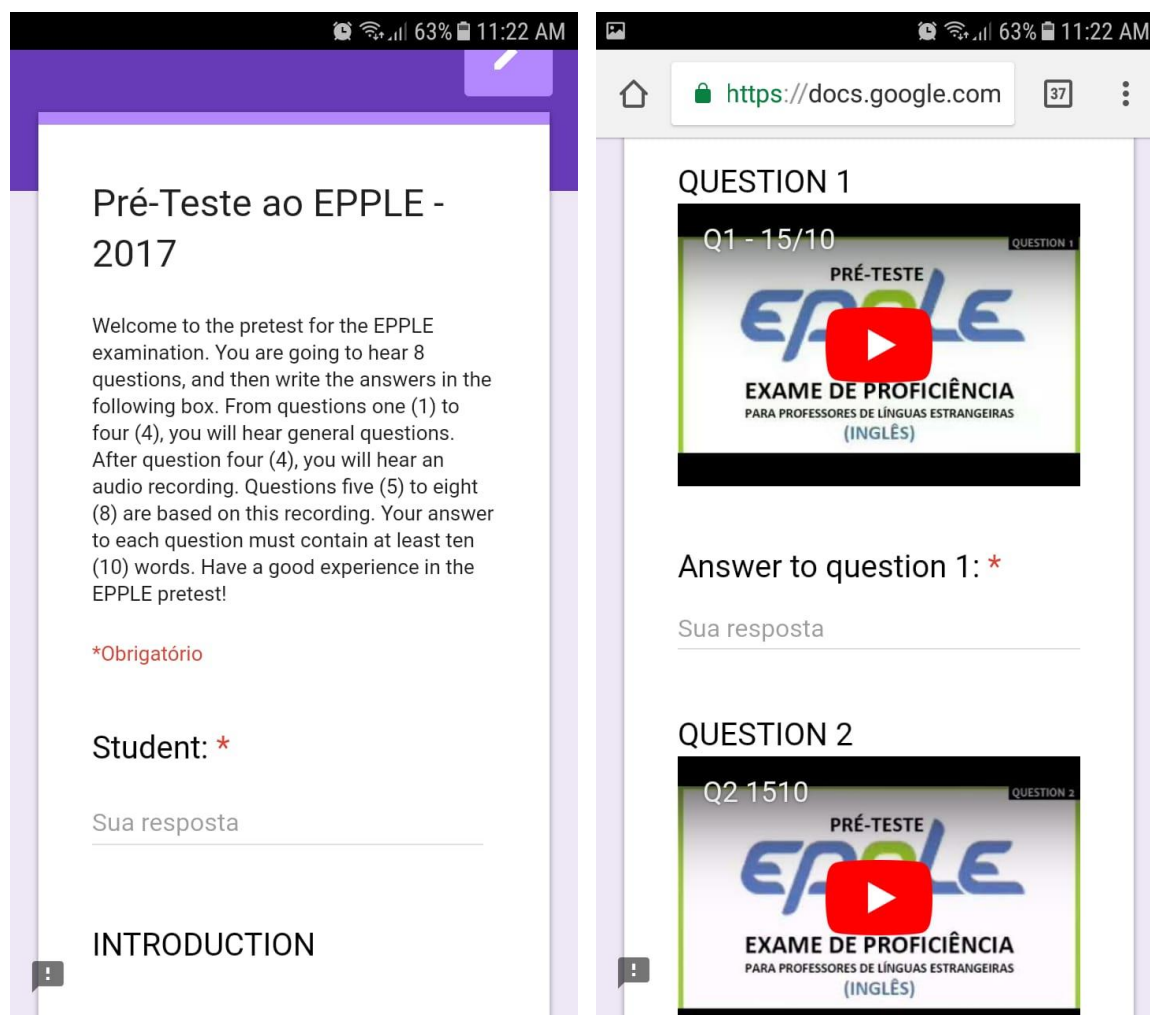
Fonte: CONSOLO; AGUENA, 2017.

**Figura 2 – Tela do PT2 em computador**



Fonte: dados da pesquisadora.

**Figura 3 – Telas do PT2 em celular**



Fonte: dados da pesquisadora.

Teceremos agora alguns comentários sobre a metodologia utilizada durante a disponibilização do PT2 na plataforma *Google Forms*. A primeira parte do desenvolvimento do PT2, já contemplada, diz respeito a pensar seu construto e operacionalizá-lo através de tarefas, processo do qual resultou um script. A partir daí pesquisamos sites de leitura de voz em inglês, buscando um áudio para as questões que fosse claro e inteligível. Nesse momento, nos deparamos com um obstáculo: os sites disponíveis gratuitamente possuíam muitas limitações, como número ou duração máxima de faixas de áudio. Além disso, que são obstáculos contornáveis por meio da divisão dos áudios em partes, com posterior junção, acreditamos que as leituras dos itens fornecidas por esses sites ficaram, de fato, extremamente robóticas. A entoação ascendente, própria de questionamentos, foi o ponto mais problemático, pois a presença do ponto de interrogação não fazia com que a frase toda ganhasse a entoação

adequada, mas, no máximo, a última palavra. Diante desse problema, decidimos gravar as questões por conta própria. Utilizamos programas de gravação gratuitos, como aplicativos de smartphone, e contamos com a ajuda de uma colega norte-americana. A maioria das questões foi lida e gravada por ela, enquanto que à pesquisadora coube introduzir algumas das questões, ou dizer partes que necessitavam de uma segunda voz, de um outro falante. O resultado ficou claro e inteligível, embora possa ser bastante melhorado com melhor equipamento, para diminuir o ruído externo – que é muito baixo, mas está lá – e também para melhor padronizar o volume.

Com os áudios preparados, passamos à tarefa de carregá-los no *Youtube* para, então, utilizar o link no pré-teste em desenvolvimento no *Google Forms*. Os vídeos foram montados com o áudio das questões e a imagem padrão do pré-teste do EPPLE, sem nenhuma pista visual. Se fosse possível fazer upload do áudio sem imagens, este seria o caminho tomado, mas o *Youtube* não permite que seja feito dessa maneira. Para nossa surpresa, essa que parecia ser a parte mais simples e rápida veio a se tornar uma das mais cansativas. O *Youtube* acusou nossa conta de *spamming* – postar o mesmo vídeo diversas vezes – embora fossem vídeos distintos. O site chegou a bloquear a postagem de vários dos vídeos, de modo que fomos obrigados a utilizar várias contas, e uma delas teve uma penalidade de não postar vídeos por três meses. Por fim, conseguimos postar todos os arquivos, em quatro contas diferentes, na categoria não-listado, apenas acessível através do link, e pudemos montar nossa versão do pré-teste.

Conforme mencionamos anteriormente, a migração do pré-teste para o *Google Forms* acarretou na não utilização da produção oral no teste, sendo substituída pela produção escrita. O pré-teste ainda visa avaliar a proficiência oral do candidato, de modo que precisamos de bastante cautela na análise e correção das respostas para evitar que sofressem demasiada interferência da proficiência escrita. Sendo assim, a produção escrita nas respostas é analisada com ênfase em aspectos linguísticos gerais e de interpretação do texto que foi ouvido pelo candidato. O tratamento das respostas enfatiza pontos como tópico (se o candidato respondeu à pergunta corretamente) e escolha lexical e gramatical; diferentemente, grafia e pontuação são olhados apenas em casos em que impeçam a compreensão, pois aí interferem na capacidade de avaliarmos o conteúdo da resposta do candidato.

## 1.4 FAIXAS DE PROFICIÊNCIA DO PRÉ-TESTE

Descrevemos as faixas de proficiência para o PT2 (chamadas provisoriamente de 1, 2, 3, 4), cujos descritores foram elaborados tendo por base o construto tratado no pré-teste (proficiência linguístico-comunicativo-pedagógica do professor de LI). São apenas quatro faixas, e elas abrangem níveis mais básicos de proficiência, uma vez que o pré-teste é um teste de curta duração e com poucas tarefas, que, a princípio, não contemplam níveis mais avançados de proficiência para que se possa fazer uma distinção entre faixas mais altas. Este tampouco é o propósito do pré-teste, uma vez que pretende, em última instância, agrupar os candidatos em categorias um pouco mais gerais de sucesso no EPPLÉ, e o exame se encarregaria de distinguir entre faixas mais específicas de sua proficiência e oferecer certificação.

Chamamos a atenção, fazendo uma pequena digressão, de que a nomenclatura das faixas poderia (e deveria) ser alterada com o propósito de serem uma interpretação o mais clara e direta possível para os candidatos ao exame EPPLÉ. Por exemplo, se descobrimos que a faixa que chamamos 4 está prevendo a faixa C do EPPLÉ oral, porque não chamá-la também de C? Além disso, se as faixas 1 e 2, juntas, estão prevendo notas acima de B, por que não uni-las e chamá-las de B? Acreditamos que, uma vez que a configuração das relações entre pré-teste e EPPLÉ oral siga uma hierarquia razoável, é possível renomearmos as faixas do pré-teste para promover uma interpretação mais direta de seu significado.

As faixas propostas inicialmente por Consolo e Agüena (2017) para o PT1 baseavam-se nos descritores do Quadro Comum Europeu de Referência (QCER) e seguiam sua nomenclatura (A1, A2, B1, B2). Embora as faixas fossem baseadas no QCER, isso não significaria dizer que um candidato classificado nas faixas em questão por meio de outro teste deva se basear nesse resultado para decidir fazer ou não o EPPLÉ, já que o exame abrange também conhecimentos específicos da profissão de professor, não avaliados pela maioria dos outros exames e testes atualmente disponíveis. Também não significava dizer que o pré-teste certificaria alunos com as faixas do QCER, já que, primeiramente, o pré-teste não foi desenhado para oferecer certificação de proficiência, apenas um diagnóstico inicial, e, além disso, não foi produzido com o mesmo construto que as faixas do QCER descrevem (proficiência linguístico-comunicativo-pedagógica, e não proficiência geral).

Assim, em nosso caso, fizemos mudanças no conteúdo dos descritores e também na nomenclatura para as faixas do PT2, a fim de dar conta das mudanças no próprio teste e de

esclarecer que ele não é vinculado ao QCER. Estamos abertos também, conforme mencionamos, a uma nova mudança de nomenclatura para benefício dos usuários do pré-teste.

O PT1 teve suas faixas elaboradas com base nas faixas de proficiência oral intituladas *Qualitative aspects of spoken language use* do Quadro Comum Europeu de Referência (COUNCIL OF EUROPE, 2001, p. 28). Os descritores, nessa etapa, eram variação, precisão, fluência, interação e coerência, exatamente os mesmos propostos no QCER. Para o PT2, essas faixas foram adaptadas, uma vez que seu construto inclui aspectos específicos da profissão de professor de línguas, que devem constar nas faixas, além do fato de que as faixas do QCER avaliam a interação do falante, e não apenas a produção ou a compreensão oral separadamente. Assim, as faixas do PT2 mantiveram os descritores “variação” e “precisão”, e ao descritor “coerência” foi adicionada a palavra “coesão” (uma vez que as respostas do pré-teste são por escrito, e a coesão é, *grosso modo*, a coerência na superfície textual). Além disso, o descritor “interação” foi substituído pelo de “compreensão oral”, e foi adicionado um descritor de “metalinguagem” para cobrir o aspecto específico da proficiência que queremos avaliar, que diz respeito à capacidade de utilizar a língua para o ensino.

A seguir, apresentamos o quadro 2 (também disponível em anexo) com as faixas de proficiência do PT2, elaboradas durante este trabalho.

### Quadro 2 – Faixas de proficiência do PT2

1	<p><b>Varição:</b> Tem uma gama suficiente de linguagem para produzir descrições claras, expressar pontos de vista sobre temas gerais, usando algumas formas de sentenças complexas para fazê-lo.</p> <p><b>Precisão:</b> Mostra um grau relativamente elevado de controle gramatical e não comete erros que causam mal-entendidos. Pode cometer erros de ortografia, causados ou não pelo autocorretor, desde que não alterem a inteligibilidade da mensagem. Marcas de oralidade também não devem ser penalizadas.</p> <p><b>Coerência/Coesão:</b> Pode usar um número limitado de ferramentas de coesão para conectar suas declarações contribuindo para um discurso claro e coerente.</p> <p><b>Compreensão oral:</b> Compreende bem um excerto de áudio e é capaz de selecionar informações para responder adequadamente a questões sobre o excerto.</p> <p><b>Metalinguagem:</b> Conhece bem as regras linguísticas, sendo capaz de explicá-las de modo claro.</p>
2	<p><b>Varição:</b> Tem conhecimento suficiente da língua para sobrevivência, com vocabulário para expressar-se sobre temas como família, hobbies e interesses, trabalho, viagens e eventos atuais.</p> <p><b>Precisão:</b> Utiliza padrões estruturais associados a situações de uso com precisão</p>

	<p>linguística razoável. Pode cometer erros de ortografia, causados ou não pelo autocorretor, desde que não alterem a inteligibilidade da mensagem. Marcas de oralidade também não devem ser penalizadas.</p> <p><b>Coerência/Coesão:</b> Pode relacionar uma série de elementos de conteúdo, curtos e simples, em uma sequência coerente e linear.</p> <p><b>Compreensão oral:</b> Compreende um excerto de áudio e é capaz de selecionar informações para responder a questões sobre o excerto, com algumas limitações.</p> <p><b>Metalinguagem:</b> Conhece as regras da língua, mas apresenta dificuldade em explicá-las de modo claro ou em selecionar informações relevantes para esclarecer dúvidas de alunos, fazendo uso limitado de terminologia específica de metalinguagem.</p>
3	<p><b>Variação:</b> Utiliza padrões básicos de sentenças e grupos de poucas palavras e fórmulas, a fim de comunicar informações limitadas em situações cotidianas simples.</p> <p><b>Precisão:</b> Utiliza algumas estruturas simples corretamente, mas ainda comete erros básicos esporadicamente. Pode cometer erros de ortografia, causados ou não pelo autocorretor, desde que não alterem a inteligibilidade da mensagem. Marcas de oralidade também não devem ser penalizadas.</p> <p><b>Coerência/Coesão:</b> Pode relacionar grupos de palavras com conectores simples tais como “and”, “but” e “because”.</p> <p><b>Compreensão oral:</b> Compreende um excerto de áudio com certa dificuldade, e não consegue responder completamente às questões sobre o excerto.</p> <p><b>Metalinguagem:</b> Conhece as regras de uso da língua parcialmente, mas não consegue explicá-las de modo claro.</p>
4	<p><b>Variação:</b> Tem um repertório básico de palavras e sintagmas simples relacionados a dados pessoais e situações concretas específicas.</p> <p><b>Precisão:</b> Mostra apenas um domínio limitado de algumas estruturas gramaticais e padrões de sentenças simples. Pode cometer erros de ortografia, causados ou não pelo autocorretor, desde que não alterem a inteligibilidade da mensagem. Marcas de oralidade também não devem ser penalizadas.</p> <p><b>Coerência/Coesão:</b> Pode relacionar palavras ou grupos de palavras com conectores lineares muito básicos como “and” ou “so”.</p> <p><b>Compreensão oral:</b> Apresenta dificuldade na compreensão de um excerto de áudio e não consegue responder a questões de compreensão sobre este excerto.</p> <p><b>Metalinguagem:</b> Pouco ou nenhum conhecimento sobre regras da língua, não conseguindo explicá-las.</p>
X	<b>Não respondeu à questão.</b>

Fonte: elaborado pela autora.

Como é possível observar, os aspectos levados em consideração para a classificação do candidato dizem mais respeito à sua produção como um todo (perspectiva global) do que a especificidades. No descritor de variação, analisamos a gama de estruturas à qual o candidato

recorre na construção de seus textos. Para o descritor de precisão, levamos em conta aspectos mais gerais de inteligibilidade, desconsiderando, por exemplo, desvios ortográficos e de pontuação por si só, uma vez que o teste pretende avaliar a proficiência oral do candidato. No entanto, depois de se levar em conta com cautela a inteligibilidade do texto, outras ocorrências do mesmo termo, e possibilidades de autocorreção automática do celular (inclusive transformando a palavra em uma parecida em língua portuguesa), alguns desvios na escrita foram considerados na verdade falhas estruturais, que teriam ocorrido também na produção oral do candidato. Esse descritor ficou um pouco mais longo que os demais e inclui uma negativa, direcionada ao examinador. Normalmente, é recomendado que descritores possuam afirmações positivas para que funcionem como objetivos para estudantes (COUNCIL OF EUROPE, 2001). Em nosso caso, as afirmações são direcionadas apenas ao examinador, não sendo redirecionadas ao candidato quando seu desempenho é relatado. Coerência/coesão é um descritor que analisa a lógica interna do texto, especialmente no que tange à utilização apropriada de conectores textuais. A compreensão oral é avaliada com base na adequação da resposta dada pelo candidato ao conteúdo da pergunta e dos textos orais apresentados a ele. Por fim, o descritor de metalinguagem trata do domínio que o candidato possui ou não de regras gramaticais e de uso das estruturas em LI, além da terminologia específica para tratar desse assunto.

Além disso, é importante chamarmos a atenção para um aspecto que será visto nos resultados: as faixas do pré-teste são apenas quatro (1, 2, 3, 4), sendo 1 a mais alta do intervalo e 4, a mais baixa. Estas faixas serão comparadas às faixas de proficiência do EPPLE, que são de outra natureza (diferentes descritores e incluem produção oral). O EPPLE possui cinco faixas de proficiência, A, B, C, D e E, sendo que a faixa A é a mais alta e a faixa E é a mais baixa. A relação entre elas será discutida brevemente ainda neste capítulo, e mais profundamente durante a análise de dados e discussão.

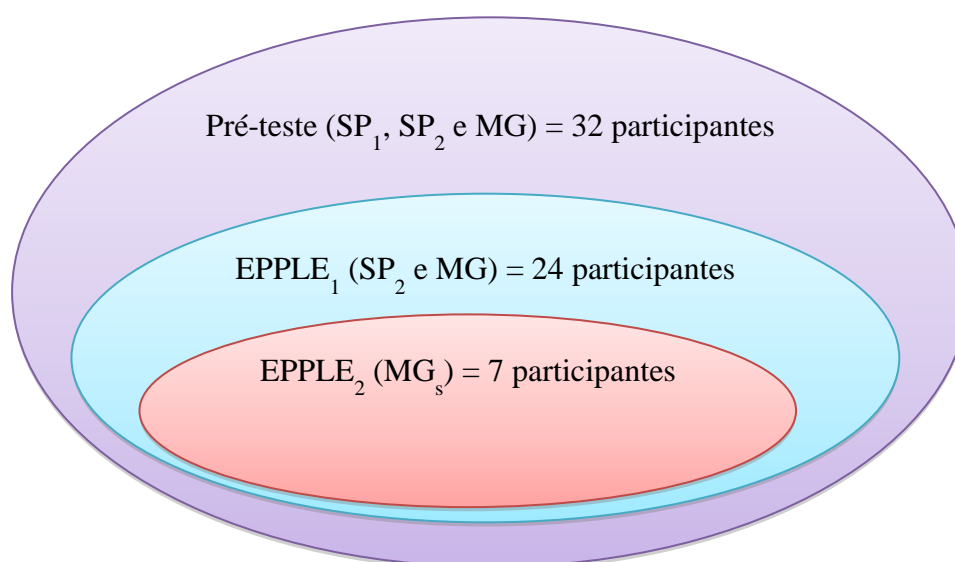
## **1.5 PERFIL DOS PARTICIPANTES**

Faremos aqui uma descrição mais detalhada dos participantes, com base em suas respostas ao questionário entregue logo após a aplicação do pré-teste. Reiteramos que o grupo MG, composto de 13 participantes, realizou o pré-teste e EPPLE<sub>1</sub>, e seu subgrupo MG<sub>s</sub>, composto por 7 dos 13 participantes iniciais, realizou o EPPLE<sub>2</sub>. O grupo SP<sub>1</sub>, composto por 8 participantes, realizou apenas o pré-teste. Ao grupo SP<sub>2</sub>, por sua vez, composto por 11 participantes, foram aplicados o pré-teste e o EPPLE<sub>1</sub>. Durante o trabalho, em alguns

momentos, ao tratarmos especificamente do pré-teste, desejamos fazer apenas a distinção entre os participantes que realizaram o pré-teste no computador (MG) e no celular (SP). Nesses casos, não há necessidade de subdividir os grupos, e, portanto, acreditamos que a utilização da sigla mais simples – apenas MG e SP – é suficientemente intuitiva para manter a organização do trabalho. Nesta seção do trabalho, entretanto, realizamos a descrição detalhada de cada subgrupo, uma vez que as aplicações foram em datas e locais distintos para cada um deles.

Antes de prosseguirmos com cada grupo, apresentamos a figura 1 com um esquema geral do número de candidatos em cada grupo, assim como as avaliações realizadas por eles.

**Figura 4 – Participantes da pesquisa**



Fonte: elaborado pela autora.

Nessa figura, é possível visualizar que todos os participantes da pesquisa realizaram o pré-teste. Dentre eles, os grupos  $SP_2$  e MG realizaram também o EPPLE, com intervalo de tempo pouco significativo. Ainda dentre estes, o subgrupo  $MG_s$  realizou uma segunda versão do EPPLE, com um mês de intervalo da aplicação do pré-teste.

### 1.5.1 Grupo MG

O grupo mais heterogêneo dentre os três, uma vez que foi composto por voluntários a prestar o exame. São treze participantes, dos quais sete se voluntariaram para fazer também o EPPLE<sub>2</sub>. Neste grupo, doze participantes são alunos de graduação, com ingresso na universidade dentro do intervalo de 2011 a 2017, e uma participante da pós-graduação. As

idades variam de 20 a 50 anos, com uma concentração maior de 20 a 29, sendo que apenas um candidato estava na idade de 50 anos. Somente dois participantes não possuem qualquer certificação de proficiência, sendo que os outros onze fizeram o TOEFL-iTP e conseguiram resultados C1 (2), B2 (6) e B1 (3). Todos os candidatos eram professores de LI na época da coleta dos dados, e eram ou já haviam sido professores do curso de extensão da universidade em questão. Alguns também trabalhavam ou haviam trabalhado em outros contextos, como aulas particulares, aulas em escolas regulares públicas e particulares, e aulas em cursos privados. Sobre o máximo nível de atingido em curso de línguas, nove participantes afirmaram ter alcançado o avançado, um chegou ao intermediário superior, dois, ao intermediário, e um, ao pré-intermediário.

### **1.5.2 Grupo SP<sub>1</sub>**

Composto por oito participantes, todos formandos de uma classe noturna de Letras – Português/Inglês. Do total, seis alunos ingressaram no curso em 2013, um ingressou em 2012 e um em 2011. As idades variam de 22 a 31 anos. Apenas quatro participantes possuem certificação de proficiência, todos pelo TOEFL-iTP, sendo que um não se lembra do resultado, outro aluno ficou com a faixa B1 na avaliação, e os dois restantes ficaram na faixa B2. Apenas três candidatos alegaram que eram ou já haviam sido professores de LI na época da coleta dos dados, em contextos diversos, como aulas particulares, cursos de extensão universitários e cursinhos privados. Três dos alunos afirmaram já terem atingido o nível intermediário superior em curso de escola de línguas, três outros participantes afirmaram ter alcançado o nível avançado, apenas um disse ter chegado ao nível básico, e um dos participantes não respondeu essa questão.

### **1.5.3 Grupo SP<sub>2</sub>**

Grupo de onze estudantes, também formandos em Letras – Português/Inglês, classe diurna. Nove participantes ingressaram no curso em 2014, e dois em 2012. Sua faixa etária variava de 21 a 35 anos. Quatro dos candidatos possuem certificação de proficiência, todos pelo TOEFL-iTP, e um também pelo TOEIC, com duas faixas B1, uma B2 e uma C1 no TOEFL, e 935 no TOEIC mencionado, que é a faixa denominada “*gold*” e corresponde à faixa B2 do QCER, mas bem próximo do seu limite com a C1. Seis dos participantes atuavam como professores de LI na época da coleta dos dados, em contextos como curso de extensão universitário, cursos privados, escolas regulares particulares e públicas e aulas particulares,

enquanto que os cinco restantes nunca exerceram a profissão. Um dos participantes, alegou ter atingido nível intermediário em curso de línguas, dois afirmaram ter alcançado o intermediário superior, e outros dois, o avançado. Os seis participantes restantes não sabiam ou se lembravam, não fizeram curso ou não responderam à questão.

O quadro 3 a seguir resume as características principais de cada grupo.

### **Quadro 3 – Características gerais dos participantes da pesquisa**

	Faixa etária	Experiência docente (passada ou presente)	Experiência de realização de exame de proficiência	Nível avançado em curso de inglês
MG	20 a 50 anos	100%	85%	69%
SP <sub>1</sub>	22 a 31 anos	37%	50%	37%
SP <sub>2</sub>	21 a 35 anos	54%	36%	18%

Fonte: elaborado pela autora.

Observando esse quadro, podemos ter uma ideia de que os grupos diferem, por exemplo, em experiência docente. O grupo MG é consideravelmente mais forte nesse aspecto, uma vez que 100% dos participantes já trabalharam como professor de LI. Quanto a experiência prévia com exames de proficiência, o grupo MG também apresenta uma maior porcentagem, que pode ser explicada pelo envolvimento dos estudantes da instituição no programa “Idiomas sem Fronteiras” e suas aplicações do TOEFL-iTP feitas nas universidades. Por fim, quanto a terem alcançado um nível avançado em curso de inglês particular, é importante ressaltar que nem todos os alunos fizeram curso e, sobre os que fizeram, não é possível afirmar muito sobre quem atingiu níveis menores do que o avançado, pois poderiam ter interrompido o curso antes de alcançar esse patamar por qualquer motivo logístico.

## **1.6 COLETA DE DADOS**

### **1.6.1 Observações de campo**

As avaliações (EPPL e pré-teste) foram aplicadas sempre com a presença da pesquisadora, que pôde observar seu andamento e registrar observações relevantes à medida

que aconteciam. Foi bastante importante presenciar especialmente os problemas de ordem técnica e como foram resolvidos no momento dos testes. As informações coletadas foram, então, organizadas, primeiramente, em diferentes contextos de aplicações (grupo MG em computadores, e grupos SP em celulares), e, em segundo lugar, foram agrupadas as semelhanças e diferenças entre esses contextos.

### 1.6.2 Aplicações e questionários

Para o desenvolvimento deste trabalho, foram feitas aplicações do pré-teste e do EPPLE em duas universidades públicas brasileiras, de dois estados da região sudeste, Minas Gerais e São Paulo. O quadro 4, a seguir, resume as aplicações com as datas e número de participantes.

**Quadro 4 – Aplicações do pré-teste e do EPPLE**

	<b>Nº de participantes</b>	<b>Pré-teste</b>	<b>EPPLE<sub>1</sub></b>	<b>EPPLE<sub>2</sub></b>
<b>MG</b>	13	20/10/2017 (computador)	20/10/2017	
<b>MG<sub>s</sub></b>	7 (dos 13 de MG)			24/11/2017
<b>SP<sub>1</sub></b>	8	31/10/2017 (celular)		
<b>SP<sub>2</sub></b>	11	28/11/2017 (celular)	30/11/2017 a 07/12/2017	
<b>TOTAL</b>	32	32	24	7

Fonte: elaborado pela autora.

As aplicações foram direcionadas a graduandos em Letras – Inglês/Português de ambas as universidades. Na universidade pública em São Paulo, duas turmas de estudantes do último período de Letras – Inglês/Português participaram do pré-teste, e uma destas turmas fez também o EPPLE, em data subsequente próxima à do pré-teste. Chamaremos de SP<sub>1</sub> e SP<sub>2</sub> as duas turmas, uma turma noturna e uma turma diurna, em que, respectivamente, oito (8) alunos fizeram apenas o pré-teste e onze (11) alunos fizeram o pré-teste e o EPPLE<sub>1</sub>, totalizando dezenove (19) pré-testes e onze (11) exames do EPPLE<sub>1</sub>. Na universidade de

Minas Gerais, os alunos se voluntariaram para realizar as avaliações, sendo que a maioria dos participantes se compõe de estudantes de graduação de diversos períodos, com exceção de uma professora já licenciada e estudante de pós-graduação. Nesse contexto, foram aplicados, em uma primeira data, o pré-teste e o EPPLE<sub>1</sub> a treze (13) participantes e, em uma segunda data, o EPPLE<sub>2</sub> (uma segunda versão do EPPLE) a sete (7) dos treze (13) participantes. Chamaremos o grupo de alunos da primeira aplicação de MG, e seu subgrupo, que fez o EPPLE<sub>2</sub>, de MG<sub>s</sub>.

De um total de trinta e dois (32) participantes, todos fizeram o pré-teste. Destes trinta e dois (32), vinte e quatro fizeram também a primeira versão do EPPLE (EPPLE<sub>1</sub>). Destes vinte e quatro (24), apenas sete (7) fizeram também a segunda versão do EPPLE (EPPLE<sub>2</sub>).

As aplicações do pré-teste e EPPLE<sub>1</sub> para o grupo MG foram feitas na mesma data: os candidatos realizaram o pré-teste e o teste escrito do EPPLE em sequência durante a manhã, e tiveram o teste oral do EPPLE marcado para diferentes horários durante a tarde, de acordo com sua disponibilidade. Por razões logísticas, o grupo SP<sub>2</sub> não pôde fazer o pré-teste e EPPLE no mesmo dia, já que as avaliações foram aplicadas durante suas aulas de LI, sendo necessárias quatro aulas para todas as avaliações (pré-teste, EPPLE<sub>1</sub> oral e escrito). O grupo SP<sub>1</sub> realizou apenas o pré-teste, e, portanto, foi necessária somente parte de uma aula.

Quanto aos equipamentos utilizados, para as aplicações do pré-teste em Minas Gerais, foi reservada uma sala com computadores com conexão à internet, mas sem fones de ouvido, de modo que os próprios participantes e examinadores trouxeram seus fones pessoais. Em São Paulo, utilizamos, conforme mencionado, a própria sala de aula dos participantes e seus celulares e fones de ouvido. Já para as aplicações do EPPLE, tanto a primeira quanto a segunda versão, foi utilizada uma sala com mesas e cadeiras para os participantes realizarem o teste escrito, já que utilizamos a versão em papel, e uma sala com computador e acesso à internet para a realização do teste oral, feito por meio de entrevista com dois examinadores. Em São Paulo, a sala disponibilizada para o EPPLE foi também a própria sala de aula dos candidatos, que contava com projetor e sistema de som integrado, além do computador conectado à internet.

Após as aplicações do pré-teste, todos os participantes responderam um questionário, que se encontra no anexo deste trabalho, que teve a intenção de coletar informações pessoais gerais sobre os participantes, além de suas impressões sobre o pré-teste e sua realização em computador ou celular. Alguns itens do questionário foram adaptados e/ou expandidos para o grupo que o realizou no aparelho celular, uma vez que esse é um dos principais interesses

desta pesquisa. Analisaremos mais profundamente as contribuições dos participantes no próximo capítulo deste trabalho.

Houve vários desafios enfrentados durante as aplicações. Considerando primeiramente o pré-teste, em Minas Gerais, o primeiro obstáculo foi encontrar uma sala multimídia na universidade em questão, pois não havia nenhum laboratório completo com computadores e fones disponíveis, apenas com computadores. Assim, a alternativa foi pedir aos participantes que trouxessem fones, e utilizar também fones conseguidos pelos pesquisadores em outros contextos. A utilização de equipamento que não pertence à sala foi, não surpreendentemente, outro problema. Alguns fones não possuíam fio suficientemente longo, tornando impossível sua conexão confortável da entrada do computador até o ouvido do candidato, sendo descartados para uso na aplicação do pré-teste. Outros fones não funcionaram bem para a reprodução do áudio, emitindo volume muito baixo para que o participante sentisse segurança na realização da avaliação. Por fim, alguns computadores tiveram problemas de conexão, atualizando a página do pré-teste, o que fez com que o aluno perdesse todo o seu progresso até então. Todos esses problemas foram resolvidos durante a aplicação do teste, através da troca de equipamentos e da disponibilização de mais tempo para os candidatos.

Nas aplicações do pré-teste em São Paulo, que foram feitas através dos próprios celulares dos participantes, houve também desafios. A conexão com a internet fornecida pela universidade era instável, com variações de sinal dentro do ambiente de sala de aula, assim como o próprio programa de dados que alguns participantes optaram por utilizar. A conexão falha resultava muitas vezes, assim como em Minas Gerais, na atualização da página do pré-teste com perda do progresso do candidato até então, o que levou uma candidata a desistir de usar o próprio celular e esperar pelo de um colega, com menos flutuações de conexão. Além disso, uma candidata não possuía smartphone, sendo necessário tomar emprestado o celular da pesquisadora, e outra candidata tinha um modelo um pouco mais antigo de smartphone, cujo sistema operacional não processava html5, o código utilizado pelo *Youtube* (onde estão hospedadas as questões do pré-teste), impedindo o acesso aos áudios. Assim, esta participante foi forçada a esperar que um colega terminasse o teste para utilizar seu celular. Outros problemas relacionados ao conteúdo e formato do teste apontados pelos participantes serão discutidos no próximo capítulo. Da mesma maneira que nas aplicações em Minas Gerais, aos participantes com problemas foi oferecido maior tempo após a troca de equipamentos.

Por fim, nas aplicações do EPPLE, foram encontrados problemas logísticos apenas no teste oral, em que também havia uma inicial dependência de conexão com a internet. Durante

a quarta entrevista realizada em Minas Gerais, houve uma interrupção da conexão com a internet, o que fez com que os examinadores tivessem que recorrer ao uso do próprio celular, com internet de dados e conectado a caixas de som, para reprodução de um vídeo. Em decorrência desse episódio, o vídeo foi baixado do *Youtube* e passou a ser reproduzido sem a necessidade de conexão com a internet. Quanto ao teste escrito, não houve problemas de ordem prática, porém uma candidata do grupo MG<sub>s</sub>, que já havia realizado o teste oral do EPPLÉ<sub>2</sub>, desistiu do teste escrito com apenas 30 minutos de sua duração. A participante alegou estar se sentindo indisposta e preocupada com outros assuntos, e afirmou que não gostaria de entregar uma prova ruim, portanto preferia se retirar. No entanto, tal evento impactou apenas seu teste escrito, e não o oral, utilizado como dado neste trabalho.

Os resultados dos candidatos nas avaliações foram coletados e contrastados de modo a se identificarem disparidades que possam invalidar o pré-teste em relação ao EPPLÉ. Os questionários também foram analisados, buscando compreender a perspectiva dos participantes do pré-teste e do EPPLÉ. Uma vez identificadas estas dissonâncias, procuraremos utilizá-las na elaboração de encaminhamentos para um novo modelo do pré-teste, o pré-teste 3 (PT3), também disponível em DTMs.

### **1.6.3 Resultados das avaliações**

Os resultados das exames EPPLÉ foram corrigidos por um time de três pesquisadores do grupo de pesquisa “Ensino e Aprendizagem de Língua Estrangeira: Crenças, Construtos e Competências” (ENAPLE-CCC). Houve algumas disparidades durante a correção devido à não-concordância dos pesquisadores com relação a alguns resultados, que foram negociados rumo a uma decisão final. Já os resultados do pré-teste foram corrigidos por dois outros pesquisadores, integrantes do mesmo grupo, e geraram bem menos disparidades.

A comparação entre os resultados já se mostra desafiadora antes mesmo de ser iniciada: as faixas do pré-teste e do EPPLÉ não foram elaboradas com uma correspondência prévia. Para o EPPLÉ, existem cinco faixas (A, B, C, D e E), sendo A a faixa mais alta e E, a mais baixa. As correções das questões são feitas com base nas faixas, mas há uma numeração correspondente a elas, que vai de 1 a 10. A faixa E abrange notas de 0 a 5,9, sendo a mais ampla de todas, abarcando candidatos com proficiência considerada insuficiente para a atividade docente. As outras faixas, em ordem crescente, cobrem as notas de 6 a 6,9, de 7 a 7,9, de 8 a 8,9 e de 9 a 10. Há duas escalas distintas para o teste oral e para o teste escrito. Como nosso trabalho está relacionado ao teste oral, informações maiores sobre essa escala

específica fazem-se necessárias. Nela, há seis descritores numerados, que dizem respeito aos mesmos aspectos em cada uma das faixas, sendo:

- 1 – fluência e ritmo de fala;
- 2 – pronúncia;
- 3 – precisão;
- 4 – capacidade de narração, descrição e reflexão sobre situações de ensino e aprendizagem;
- 5 – conhecimento metalinguístico;
- 6 – compreensão oral.

Em adendo, o descritor de número 4 nos parece questionável, uma vez que não se relaciona a uma habilidade propriamente linguística, ou relativa ao uso da LI na profissão: entendemos que o professor deve ser capaz de refletir sobre situações de ensino e aprendizagem, mas fazê-lo não faz parte do construto do EPPL e do pré-teste, que visam avaliar a proficiência linguística do profissional docente de línguas. Anchieta (2015) chama a atenção para o mesmo caso na escala de proficiência do teste escrito, que também parece contemplar alguns aspectos que extrapolam a proficiência linguística.

Antes de partir para a análise dos resultados, fizemos um estudo prévio das faixas de proficiência e seus descritores buscando uma correspondência esperada entre a escala do EPPL oral e do PT2. A interrelação entre as faixas e descritores torna-se mais desafiadora na medida em que o PT2 não avalia a produção oral (não possuindo descritores de fluência e pronúncia) e pretende avaliar níveis mais baixos de proficiência, já que é um teste simples e de curta duração. Baseando-nos dos aspectos textuais de cada escala, montamos o quadro 5 a seguir, com nossas expectativas teóricas. A coloração pretende tornar mais fácil a visualização das faixas que estamos estudando em relação às do EPPL oral (já estabelecidas).

**Quadro 5 – Correspondência esperada entre escalas de proficiência do PT2 e EPPLE**

PT2		EPPLE oral
1		A
1		B
2		C
2	3	D
3	4	E

É possível observar que esperamos que a faixa 1, a mais alta do PT2, abarque alunos proficientes de mais de um nível no EPPLE oral, ou seja, o PT2 não faz distinção entre os alunos de proficiência mais alta, pois sua avaliação alcança somente um determinado nível, não sendo um teste com grande especificidade em seus resultados. Candidatos 1 podem vir a ser muito bons ou excelentes no EPPLE oral, e ainda estarão todos agrupados na faixa 1 do PT2. Já a faixa 2, embora se espere que abranja examinandos com nível de proficiência bom, pode se relacionar a uma faixa relativamente baixa no EPPLE oral, considerando que os aspectos de produção oral não foram avaliados e que as faixas A a D do EPPLE oral cobrem um intervalo menor de variação, enquanto a E abarca todo o espaço de pontuação de 0 a 5,9, conforme explicamos anteriormente.

No capítulo de análise de dados, nossas expectativas de resultado serão comparadas com os resultados reais que encontramos durante as aplicações. Além disso, faremos um estudo mais profundo dos descritores dos dois instrumentos de avaliação (EPPLE oral e PT2), que basearam esse quadro, buscando encontrar razões pelas quais os resultados confirmaram ou não nossas expectativas.

## **1.7 PROCEDIMENTOS DE ANÁLISE**

As correções do pré-teste foram feitas por dois pesquisadores, e se basearam nas faixas de proficiência desenvolvidas para este fim. Algumas questões geraram dúvidas por lidarem com aspectos que precisam ser mais explorados em pesquisas com avaliação em DTMs, como, por exemplo, a presença do autocorretor e seu impacto na redação final dos participantes.

As correções do EPPLE foram feitas por três pesquisadores, e geraram, também, algumas discrepâncias. Acreditamos que estas apontam para a necessidade de mais trabalhos sobre as faixas de proficiência do EPPLE e o processo de correção do exame.

Os resultados finais foram agrupados de duas maneiras, uma tendo como referência os resultados do pré-teste, e outra tendo como referência os resultados do EPPL oral. Foram observadas associações com relação às faixas de proficiência, foi feita uma análise mais detalhada dos itens das avaliações e suas relações com a correção e o escore final dos participantes com resultados mais discrepantes.

Em relação aos questionários, estes foram analisados partindo de uma enumeração de respostas e posterior categorização, agrupando as semelhantes e contabilizando sua frequência nos diferentes grupos (MG, SP<sub>1</sub> e SP<sub>2</sub>). Seus dados foram utilizados para fazermos inferências tanto sobre a validade do pré-teste como instrumento de medida da proficiência linguístico-comunicativo-pedagógica (questões fechadas na primeira seção do questionário), quanto sobre os aspectos tecnológicos das aplicações (questões abertas).

Os dados das questões abertas dos questionários foram categorizados, assim como as anotações de campo, de acordo com a proposta de Gillham (2000). Assim, foram feitas várias leituras buscando elementos que se repetissem ou se agrupassem em uma determinada categoria para análise.

A seguir, prosseguimos para o capítulo de fundamentação teórica, em que discorreremos sobre conceitos essenciais para o desenvolvimento dessa pesquisa, com base na produção científica até então.

**FUNDAMENTAÇÃO  
TEÓRICA**

## 2.1 PROFICIÊNCIA LINGUÍSTICA E AVALIAÇÃO

O termo “proficiência” na área da linguagem já foi bastante discutido por vários estudiosos (por exemplo, em SCARAMUCCI, 2000; LLURDA, 2000; CONSOLO, 2006b), uma vez que sua definição foi, por vezes, entendida de maneiras diferentes, e em alguns aspectos sobreposta a outros conceitos. Nesse sentido, acreditamos ser positivo fazer uma diferenciação entre os termos “proficiência” e “competência” à luz de seu desenvolvimento histórico na literatura da área.

Em seu texto, Llurda (2000) inicia a discussão com o termo “competência”, que foi inicialmente definido na área da linguagem por Chomsky (1965/1975), em sua teoria gerativa. Chomsky foi bastante criticado na medida em que excluiu do âmbito da competência as variações sociais e situacionais da linguagem, assim como qualquer de seus aspectos observáveis ou aplicados, banindo-os para o domínio do desempenho. Assim, Chomsky define competência de maneira estática, como um conhecimento completo da língua por um falante ideal, habitante de uma comunidade homogênea, e imune às intempéries da mente humana em funcionamento real, como falhas na memória, distrações etc. Hymes (1972) teve grande importância ao questionar a exclusão da habilidade de uso do conceito de competência, introduzindo “competência comunicativa” como um termo mais amplo, que engloba, além da competência proposta por Chomsky, chamada por Hymes de gramatical, também os aspectos psicolinguístico (o que é factível em termos de processamento humano de informação), sociocultural (valor ou significado social dos enunciados) e probabilístico (enunciados que de fato ocorrem).

Outros trabalhos notáveis trataram de competência comunicativa, e desenvolveram sua definição também no âmbito da segunda língua ou língua estrangeira. Além do problema já apontado da exclusão dos aspectos sociais e situacionais na competência chomskyana, para os aprendizes de segunda língua havia disparidades maiores: como pensar competência de maneira estática ao observar alunos de segunda língua progredirem de modo diferente de crianças aprendendo língua materna; como explicar seus padrões de interlíngua; e como entender o acesso à Gramática Universal desses aprendizes? Tais perguntas foram discutidas por vários autores, dentre eles destacamos Canale e Swain (1980), que elaboraram uma discussão entre o que havia sido proposto até então e sintetizaram a proposta de Hymes (1972) para ensino e avaliação de línguas estrangeiras. Os autores se utilizam de três das quatro categorias propostas pelo linguista, deixando de lado a categoria psicolinguística por

incluir fatores que não são específicos da competência comunicativa e devem, portanto, ser tratados como parte do desempenho.

Stern (1983) traz sua contribuição para a discussão do termo “proficiência”, e o utiliza intercambiavelmente com “competência”, dando prioridade ao uso do primeiro para descrever as habilidades de aprendizes de segunda língua. O autor levanta aspectos de conhecimento de uma língua, ou seja, o que significa passar de um estado de menor para maior proficiência em uma língua estrangeira, já compreendendo o conceito como um *continuum*, que é a visão técnica atual (SCARAMUCCI, 2000). Além disso, nega que a completa competência, ou proficiência, seja atingível para um aprendiz estrangeiro e, por esse motivo, não deve ser tampouco desejável. De acordo com o autor,

Proficiência pode ser entendida como uma meta e assim definida em termos de objetivos ou padrões. Estes podem servir de critérios pelos quais se avalia proficiência como um **fato** empírico, isto é, o desempenho real de um dado aprendiz ou grupo de aprendizes<sup>6</sup> (STERN, 1983, p. 341, tradução nossa).

Atualmente, o termo proficiência é entendido como uma habilidade direcionada ao uso, enquanto que o termo competência é definido, no âmbito teórico do gerativismo, em oposição ao uso (desempenho). Nosso foco é a proficiência linguística, definida por Scaramucci (2000) como

o resultado da aprendizagem, uma meta, definida em termos de objetivos ou padrões e, portanto, de interesse de professores, administradores, elaboradores de currículos, construtores de testes, pesquisadores, pais e alunos (SCARAMUCCI, 2000, p. 12)

Embora seja resultado de uma aprendizagem, não necessariamente em situação escolar, a proficiência, quando se trata de avaliação, está desvinculada do contexto de instrução. Falamos de avaliação de proficiência linguística como a medida em que um candidato tem habilidades para uso futuro da língua em determinada/s situação/ões, em oposição à avaliação de rendimento, que pretende estimar o quanto de um curso ou unidade de instrução foi aprendido por um determinado indivíduo.

Scaramucci (2000) diferencia ainda, em seu texto, conforme mencionamos anteriormente, as visões técnica e não-técnica de proficiência, sendo que a que nos interessa aqui é a primeira, entendida como um *continuum*, em contraste com a visão não-técnica de “tudo ou nada”. Assim, proficiência pressupõe uma habilidade de uso futuro, medida em graus maiores ou menores, e direcionada a um propósito ou situação de uso.

---

<sup>6</sup> No original: “Proficiency can be looked at as a goal and thus be defined in terms of objectives or standards. These can then serve as criteria by which to assess proficiency as an empirical *fact*, that is the actual performance of given individual learners or groups of learners”

O supracitado uso futuro da língua em situação real é chamado de **critério**. Ao elaborar e aplicar uma avaliação, os examinadores nunca têm acesso ao critério, uma vez que para isso seria necessário estarem presentes nas situações posteriores à avaliação, nas quais o candidato teria que desempenhar a função avaliada. Ao invés disso, os examinadores fazem inferências sobre o critério baseando-se no desempenho do candidato apenas durante a avaliação (McNAMARA, 2000). Em testes de proficiência, devemos proceder a uma análise de necessidades (*needs analysis*) do público-alvo com relação ao uso futuro da língua (SCARAMUCCI, 2000) para que possamos desenvolver um instrumento que o avalie apropriadamente, especialmente levando em conta que testes de proficiência podem ter importantes funções sociais, como permitir que um indivíduo conquiste metas em sua educação ou carreira.

Uma vez que entendemos a proficiência, segundo Scaramucci (2000), como um *continuum*, é natural falarmos em testes de proficiência que classifiquem os candidatos em faixas de proficiência. Além disso, tratando-se especificamente da proficiência do professor de LI, podemos falar que o Pré-Teste e o EPPL se aproximam da modalidade de avaliação de ESP (*English for Specific Purposes* – Inglês para Fins Específicos), assim como proposta por teóricos como Douglas (2013). Nas avaliações mencionadas, são avaliados aspectos gerais de conhecimento linguístico assim como habilidades metalinguísticas requeridas no trabalho com o ensino de línguas, que são, portanto, específicas da profissão. Futuramente, os resultados do EPPL pretendem servir de base para certificar professores que alcancem uma faixa de nível de proficiência linguística considerada como nível mínimo para que desempenhem bem suas funções docentes.

### **2.1.1 Proficiência linguístico-comunicativo-pedagógica**

Ainda no tópico de proficiência, Consolo e Silva (2014) apresentam o construto do EPPL como a proficiência linguístico-comunicativo-pedagógica de professores. Tal construto leva em consideração o conceito de competência de uso (*competence for use*) de Hymes, que diz respeito às regras de uso de uma língua, sob as quais alinhamos todos os aspectos sintáticos, morfológicos e fonológicos. São incluídas, portanto, nas habilidades avaliadas no exame, aspectos de proficiência linguística geral, e também aspectos específicos à profissão de professor de línguas, como a capacidade de utilizar metalinguagem, linguagem para se organizar um tópico progressivamente e explicá-lo com clareza, dentre outros.

Conforme Freeman et al (2009), é importante pensar que há dois conteúdos a serem avaliados para que um professor seja apto ao ensino de línguas: um deles é a língua em si, e o outro é o conhecimento sobre a língua e seus usos (metalinguístico). É possível ensinar sobre uma língua utilizando como código uma outra língua, por exemplo. Em se tratando de testagem, importa não confundir os construtos ou inferir um através do outro: um candidato que tem bastante conhecimento linguístico não necessariamente tem conhecimento metalinguístico, ou sabe articular seu conhecimento linguístico com clareza para dar instruções, ou adequar seu discurso a níveis variados de alunos. Da mesma forma, um candidato com grande conhecimento metalinguístico, que saiba normas gramaticais do inglês, não necessariamente possui fluência em sua fala ao explicar conteúdos utilizando a própria língua estrangeira como meio.

Os dois conhecimentos são por nós considerados de igual importância para a profissão de professor de LI, uma vez que ensinar a língua estrangeira através da materna não oferece ao estudante a máxima exposição à língua estrangeira que ele teria se a aula fosse dada por meio dessa língua. Vários autores já desenvolveram trabalhos sobre a importância da interação, especialmente a oral, na língua estrangeira para que os alunos tenham oportunidade de testar hipóteses, negociar significados e mesmo praticar rotinas e automatizá-las (LONG, PORTER, 1985; CONSOLO, 2006a, por exemplo).

Portanto, ressaltamos a proficiência linguístico-comunicativo-pedagógica como construto do EPPL e, por conseguinte, também de seu pré-teste. Embora haja diferenças entre o EPPL e o pré-teste, como o formato do teste, a duração, o número de itens, dentre outras, há também a semelhança de se buscar avaliar um construto semelhante, que leve a uma interrelação confiável entre os resultados de ambos os instrumentos.

## **2.2 PRÉ-TESTAGEM E TESTES DIAGNÓSTICOS**

Conforme descrito por Consolo e Agüena (2017), o pré-teste ao EPPL pretende proporcionar uma experiência preliminar ao exame que sirva de indicação ao candidato de suas chances de sucesso no EPPL. É relevante mencionar que utilizamos aqui o termo pré-testagem não como sinônimo de pilotagem de um teste, processo no qual um teste é aplicado a um pequeno grupo anteriormente à sua implementação, com o objetivo de coletar dados para aprimoramento do próprio teste e dos procedimentos de sua realização. Em nosso caso, embora estejamos também coletando dados de uma aplicação preliminar, o nome “pré-teste”

diz respeito não a essa aplicação, mas ao fato de o teste estar disponível como experiência anterior ao EPPE, indicando ao candidato sua potencial faixa de proficiência no exame.

O pré-teste assemelha-se, em relação ao seu propósito, a testes como o *Preliminary Test* do ECPE (*Examination for the Certificate of Proficiency in English*), da Universidade de Michigan, na medida em que pretende pré-selecionar candidatos com possibilidade de aprovação no EPPE, informando-os para que possam tomar uma decisão esclarecida sobre se inscrever ou não para o exame, evitando um comprometimento financeiro e logístico que poderia não trazer os resultados esperados. Além disso, levando em conta seu formato, o pré-teste também tem semelhanças com testes como os disponíveis no projeto DIALANG (ALDERSON & HUHTA, 2005), que reúne um conjunto de testes diagnósticos em 14 línguas hospedados *online*, de acesso livre e ilimitado, de baixa relevância (chega a ser chamado por Alderson (2007) de “*no-stakes*” – nenhuma relevância). O pré-teste também é um teste de acesso livre, a diferença aqui está no fato de que ele é vinculado a um exame de maior impacto.

Conforme já mencionamos anteriormente, o pré-teste tem natureza comunicativa, é um teste de proficiência, mas possui algumas propriedades de testes diagnósticos. Assim, sentimos necessidade de explorar um pouco a temática de testes diagnósticos em nosso trabalho.

De acordo com Alderson (2007), o objetivo principal de testes diagnósticos é explicitar pontos fortes e fracos em estudantes, direcionando-os à aprendizagem e ao desenvolvimento, ou seja, comumente testes diagnósticos estão relacionados à avaliação formativa. Embora o diagnóstico seja citado por estudiosos da área como um dos grandes propósitos da avaliação (proficiência, rendimento, progresso, diagnóstico, nivelamento, aptidão), testes de diagnóstico ainda são um campo bastante negligenciado, como se observa pelo fato de que é comum entre pesquisadores a afirmação de que testes diagnósticos podem ser usados com propósito de nivelamento, e vice-versa.

No caso do pré-teste, propomos no final de nosso texto como encaminhamento maneiras para que ele possa se tornar mais parecido com um teste diagnóstico, com o fornecimento de um *feedback* mais completo, e voltado ao desenvolvimento dos professores em serviço ou em formação que se propuserem a realizá-lo anteriormente ao EPPE.

No texto de Alderson & Huhta (2005) em que os autores exploram as especificidades dos instrumentos de avaliação do projeto DIALANG, é explicado também um outro aspecto importante desses testes diagnósticos, aspecto este compartilhado pelo pré-teste: eles não

podem ser utilizados para certificação. Para que um teste ofereça certificação de forma válida, é necessário tomar uma série de medidas quanto à confiabilidade de seu processo avaliativo e de seus resultados, especialmente, nos casos do DIALANG e do pré-teste, levando em conta que qualquer pessoa pode acessá-los de qualquer lugar, e que não há como comprovar suas identidades, ou se estão realizando consultas a outros materiais, por exemplo. Esse modelo de teste, em que a responsabilidade pelo processo avaliativo se encontra nas mãos do candidato, só funciona porque estamos tratando de testes com propósito exclusivo de diagnosticar, ou seja, testes que servem apenas para informar o próprio participante de sua proficiência. No caso do DIALANG, há informação também de aspectos que podem ser melhorados, e, no caso do pré-teste, das chances de sucesso no EPPLÉ. Assim, o candidato não “ganharia nada” por burlar normas comuns de avaliações; pelo contrário, perderia a chance de receber um resultado que lhe poderia vir a ser útil.

Embora o objetivo do pré-teste seja oferecer essa proficiência preliminar aos participantes, em sua primeira versão, no trabalho de Agüena (2016), a autora afirma que o pré-teste ao EPPLÉ não pôde ser utilizado para as previsões propostas inicialmente, sendo suas previsões reconsideradas para abrangerem grupos maiores. Em suas aplicações, Agüena (2016) percebeu que os resultados do pré-teste não estavam direcionando a uma faixa no EPPLÉ oral. Assim, sua solução foi agrupar as faixas, transformando as quatro faixas de seu pré-teste em duas. Segundo Agüena (2016, p. 3), “[...] a recomendação ao EPPLÉ de acordo com os níveis do Pré-Teste [...] foi repensada para não limitar o EPPLÉ em função dos resultados obtidos no Pré-Teste”. A pesquisadora continua seu texto explicando que o motivo foi a não coincidência do nível de proficiência obtido no PT1 e no EPPLÉ. Para resolver provisoriamente tal impasse, Agüena (2016) agrupou os participantes em dois grupos: A1 e A2 provavelmente não obteriam êxito no exame; e B1 e B2, provavelmente sim.

Partindo destas experiências anteriores, desenvolvemos uma segunda versão do pré-teste, buscando uma interrelação mais estreita entre seus resultados e os resultados do teste oral do EPPLÉ. A partir da análise dos dados coletados nessa pesquisa, pretendemos também propor novas alterações, maiores, se necessário, ao projeto original do pré-teste.

### **2.2.1 Associação entre avaliações de línguas**

De acordo com North (2000), a literatura em avaliação reconhece cinco maneiras de associar um teste a outro: equivalência, calibragem, moderação estatística, previsão e

moderação social<sup>7</sup>. Em seu artigo, o autor define todos eles, explicando sua metodologia, e elicitando seu rigor científico através de seus pontos fracos e fortes. Faremos, a seguir, uma breve síntese de tais informações, relacionando-as com a metodologia do nosso trabalho.

A associação de avaliações é uma das questões mais “espinhosas” na avaliação de línguas, conforme o autor. São geralmente feitas apenas por instituições grandes, como a *University of Cambridge Local Examinations Syndicate* (UCLES) ou *Educational Testing Service* (ETS), que se apoiam na expertise de estudiosos da área. Professores comuns normalmente fazem essa associação entre testes intuitivamente, sem utilizar métodos científicos. Nos parágrafos a seguir, exploramos as cinco maneiras de associar avaliações, nos baseando no texto de North (2000).

A **equivalência** é aceita, em geral, como o método mais rigoroso dentre os citados pelo autor. Consiste em igualar dois instrumentos de avaliação em um nível que ambos se tornar intercambiáveis. É o que deve ser feito, por exemplo, com diferentes versões do mesmo teste. Em teoria clássica de testes, são calculados médias e desvios padrão dos instrumentos a serem equiparados até que estejam bastante próximos, e com os resultados dos dois testes em alta correlação. Nesse momento, os testes são declarados equivalentes. Em abordagens mais recentes, processos de equivalência não precisam incluir estatística.

Quando são relacionados itens, amostras de desempenho ou testes a graus de dificuldade dentro de uma mesma escala falamos em **calibragem**. Esse processo assume que dois testes medem o mesmo construto, mas em níveis diferentes e com diferentes graus de confiabilidade. Para que seja bem-sucedida, a calibragem deve ocorrer entre testes com a mesma abrangência de conteúdo e condições de administração.

Enquanto que as metodologias de calibragem permitem que seja feita uma associação prévia entre testes, a **moderação estatística** lida com contenção de danos, sendo, de certa maneira, sua complementar. Assim, enquanto a calibragem molda as avaliações, a moderação estatística atenua possíveis problemas posteriores. Ela é feita basicamente pela correspondência entre médias e desvios padrão de dois testes, mas os resultados não podem ser considerados equivalentes, nesse caso, uma vez que foram moderados estatisticamente para o estabelecimento de uma associação.

A **previsão** é considerada a forma mais fraca de associação estatística. Carey assim a define:

---

<sup>7</sup> No original, os termos são, respectivamente, *equating*, *calibrating*, *statistical moderation*, *predicting* (ou *benchmarking*) e *social moderation* (ou *standards-oriented assessment*).

Previsões podem ser feitas enquanto houver alguma relação entre o desempenho em uma avaliação e o desempenho em outra. A precisão da previsão dependerá da força das relações, e a previsão por si só é sensível ao contexto, grupo e momento<sup>8</sup> (CAREY, 1996, p. 9, tradução nossa)

Dessa forma, são estabelecidas relações estatísticas entre dois instrumentos de avaliação, mas tais relações podem ser alteradas caso seja modificado, por exemplo, o contexto de aplicação.

Ainda sobre previsão, North (2000) aponta contextos em que comumente se utiliza essa técnica: *mocks*, *benchmarking* e *cutting corners*. Todos se assemelham em alguma medida com o contexto do pré-teste: uma seleção prévia de indivíduos para algum propósito. *Mocks* é quando se aplicam testes antes de um exame para decidir quem deverá ser aceito em um programa linguístico, e pode acompanhar um aspecto diagnóstico. *Benchmarking* é a utilização da avaliação (também menos formal) para orientar indivíduos sobre quais exames devem visar. Por fim, *cutting corners* é, ainda segundo o autor, “um teste simples, mecanizado, que pode ser utilizado como um dispositivo de *screening* para prever/adivinhar qual resultado um estudante pode atingir em um teste mais abrangente”<sup>9</sup> (NORTH, 2000, p. 566, tradução nossa). Tal método é utilizado também para diminuir gastos com entrevistas ou outros métodos de avaliação, já eliminando candidatos que estão abaixo do padrão requerido para certa oportunidade de emprego ou de estudo.

Finalmente, a **moderação social** envolve a construção de um quadro comum de referência por meio de discussão e análise de amostras de trabalho. Estatísticos tradicionalmente consideram essa técnica como a menos rigorosa, uma vez que não há análise estatística. North (2000), no entanto, argumenta em seu texto que essa poderia ser a metodologia mais eficaz quando alia os *insights* subjetivos da moderação social a uma análise mais objetiva oferecida por alguma das outras técnicas.

Observando as técnicas de associação entre avaliações, é inegável que o trabalho feito com o pré-teste em relação ao EPPLE oral se aproxima da técnica de previsão, especialmente de suas subcategorias *benchmarking* (em que a avaliação orienta o que candidato sobre que exames visar) e *cutting corners* (avaliação que prevê resultados de testes mais abrangentes). Embora seja considerada a maneira menos precisa, dentre as estatísticas, a previsão ainda é

---

<sup>8</sup> No original: “Predictions can be made as long as there is some relation between the performance on one assessment and performance on another. The precision of the prediction will depend on the strength of the relations, and the prediction itself is sensitive to context, group and time”

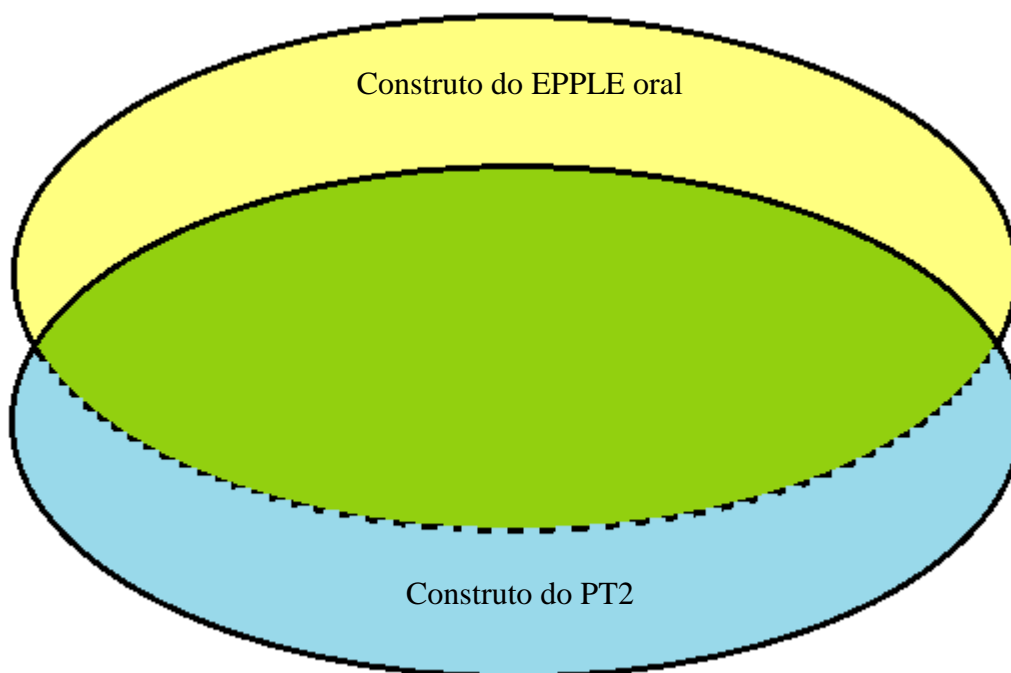
<sup>9</sup> No original: “A simple, mechanised test can be used as a screening device to predict/guess what result a learner would achieve on a more comprehensive test.”

válida como um trabalho inicial na medida em que aponta para tendências gerais a serem mais desenvolvidas em estudos posteriores.

### 2.2.2 Associação entre avaliações de construtos distintos

Como observamos nas últimas subseções, os construtos do EPPLE oral e do PT2 não são exatamente os mesmos, embora haja sobreposição. O EPPLE oral avalia a proficiência linguístico-comunicativo-pedagógica oral (compreensão e produção), enquanto que o PT2 avalia proficiência linguístico comunicativo-pedagógica oral (compreensão oral) através de respostas escritas pelo candidato. A seguir apresentamos a figura 2, que representa os construtos dos dois instrumentos.

**Figura 5 – Construtos do EPPLE oral e do PT2**



Fonte: elaborado pela pesquisadora.

Com essa figura, pretendemos demonstrar que os construtos dos dois instrumentos estão parcialmente em sobreposição. Ambos os instrumentos avaliam a proficiência linguístico-comunicativo-pedagógica do professor de línguas no que tange à compreensão oral (verde). O EPPLE oral avalia também a produção oral dos candidatos (amarelo). Já o PT2 avalia também a produção escrita dos candidatos (azul), embora tenhamos tentado minimizar

seu impacto no desempenho final de cada participante, dado que o PT2 tem como foco a proficiência oral.

Embora as respostas sejam escritas no PT2, não consideramos que ele avalie produção escrita especificamente, uma vez que para isso geralmente são levados em conta aspectos como ortografia, organização e formatação textual. Tentamos ao máximo minimizar a influência da produção escrita no resultado do candidato, de modo que os aspectos da produção escrita que são avaliados são os que seriam avaliados também na produção oral, como acuidade e variedade lexical. Nos descritores do QCER, por exemplo, quando se trata de competência comunicativa, os descritores de produção oral e escrita são os seguintes, organizados no quadro 6:

**Quadro 6 – Descritores de competência comunicativa na produção escrita e oral**

<b>Descritores de produção oral</b>	<b>Descritores de produção escrita</b>
Variedade linguística geral	Variedade linguística geral
Variedade lexical	Variedade lexical
Acuidade gramatical	Acuidade gramatical
Controle lexical	Controle lexical
Controle <i>fonológico</i>	Controle <i>ortográfico</i>
Adequação sociolinguística	Adequação sociolinguística
Desenvolvimento temático	Desenvolvimento temático
Coerência	Coerência
Precisão proposicional	Precisão proposicional

Fonte: COUNCIL OF EUROPE, 2001, p. 27-31, itálico nosso.

Neste quadro, italicizamos a única diferença entre produção oral e escrita prevista pelo QCER. No entanto, existem mais aspectos a serem levados em conta pelo fato de o EPPLE oral ser um exame em pares, que avalia não somente a produção, mas a interação entre os participantes. Nesse caso, O QCER adiciona mais descritores: flexibilidade, tomada de turno e fluência oral. Ainda assim, acreditamos ser possível que o pré-teste possa prever o EPPLE oral com suficiente propriedade, uma vez que vários descritores são comuns à produção linguística escrita e oral. Futuramente, podemos melhorar ainda mais o PT2 migrando-o para uma plataforma que permita também a produção oral.

Assim, parte do construto do pré-teste não consta no EPPLE oral, e pretendemos focar, no pré-teste, na parte que é comum aos dois instrumentos (compreensão oral). Podemos falar, então, de *muddying* (contaminação), um conceito proposto por Weir (2005). De acordo com o autor, a contaminação acontece quando não temos acesso direto a um construto, pois esse está “entremeado” a outro. No caso do pré-teste, por exemplo, durante a correção, se uma resposta não é ideal, quanto dela se deve ao fato de o participante não ter compreendido o excerto e quanto se deve a uma inabilidade de expressar-se em língua inglesa por escrito? Essa é uma pergunta cuja resposta não podemos afirmar, e que pode ser mais problemática para participantes de nível de proficiência mais baixo.

Ainda assim, é possível argumentar a favor da utilização do PT2 como instrumento de previsão. Se observamos a literatura de testes de diagnóstico, podemos perceber que testes de vocabulário, por exemplo, podem ser relacionados a resultados de proficiência geral com resultados adequados (ALDERSON & HUHTA, 2005; ALDERSON, 2007). Aliás, podemos estabelecer correlações, positivas ou negativas, entre basicamente quaisquer instrumentos de medida de habilidades, baseando-nos em seus resultados por meio do processo de previsão, conforme descrito por North (2000) na seção anterior. Além da comparação entre resultados, temos argumentos que justificam o uso das respostas por escrito, para além do motivo técnico da impossibilidade de coletar produção oral na plataforma utilizada.

Assim, além dessa justificativa técnica, temos uma segunda justificativa, de ordem teórica. Corrêa (2006) afirma que “a presença do falado no escrito não registra apenas a relação entre duas diferentes tecnologias, mas a relação entre dois modos de enunciação que se constituem mutuamente”. O autor defende ainda que a heterogeneidade não aparece **na** escrita, mas é característica constitutiva **da** escrita. Conforme Corrêa (2008), é possível reconhecer a heterogeneidade como constitutiva da escrita de três maneiras: a primeira por meio de aspectos de representação gráfica, já que o escrevente faz escolhas quanto a tentar representar as características fonológicas e fonéticas da língua ou a tentar utilizar a convenção institucionalizada de escrita; a segunda é por referência à heterogeneidade da língua em geral, que deve ser admitida tanto na fala quanto na escrita; a terceira é como o escrevente imagina ser a representação da fala pela escrita, como funciona o código escrito e qual a relação do seu texto com outros (orais e escritos).

Além disso, essa “interferência” do oral no escrito não é intrínseca ao texto, ou seja, não pode ser identificada por qualquer leitor e em qualquer contexto. De acordo com Signorini (2008), a presença do oral no escrito pode ser percebida no momento da leitura, a

depender do leitor, do papel do leitor e do redator e do contexto da comunicação. Assim, em nosso caso, podemos escolher não julgar essa presença como inadequada, mas sim como constitutiva da escrita. Ademais, a presença do oral no escrito é até mesmo uma característica positiva em um teste que visa avaliar, predominantemente, a proficiência oral.

Tratando-se especialmente da escrita feita na internet, esta é percebida ainda mais fortemente como um objeto misto, heterogeneamente constituído (SIGNORINI, 2008, p. 99). Em diversas situações de escrita através do celular e do computador, podemos observar maior uso de abreviações e menor formalidade. Alguns autores chegam a defender que, no *continuum* entre fala e escrita, a comunicação que se faz através da internet se aproxima mais da fala do que da escrita (CRYSTAL, 2004; HALLIDAY, 1996 apud MARCUSCHI, 2009, p. 75-76), uma vez que ambas tem se tornado cada vez mais próximas devido às novas possibilidades e novos discursos que surgiram com os avanços tecnológicos. Em nosso caso, o contexto de avaliação requer um reajuste dos índices de formalidade e do próprio gênero, ainda que sejam utilizados os mesmos canais de comunicação que sugerem informalidade. No momento da correção, esses aspectos foram levados em conta ao se considerar o tamanho do texto, sua organização e foram desconsiderados desvios exclusivos da escrita que não impediam a compreensão, especialmente relativos à ortografia.

Outras peculiaridades do texto digitado na web têm relação com os próprios instrumentos utilizados. No caso do celular, o próprio dispositivo apresenta aspectos que o diferenciam da escrita em papel e no computador, como a velocidade de digitação, que pode variar com maior familiaridade do usuário com o aparelho e também de acordo com o teclado disponível (teclado numérico, QWERTY virtual ou analógico, dentre outros), as autocorreções habilitadas em vários modelos atuais de celulares, a interação com vários aplicativos ao mesmo tempo, sendo que o usuário pode receber notificações enquanto estiver digitando, dentre outros. O computador também apresenta um teclado que pode ser mais ou menos familiar ao candidato do que o teclado do celular, ou mesmo a própria caneta ou lápis. Nos dias atuais, no entanto, cada vez mais pessoas possuem celulares e computadores e têm facilidade em utilizá-los para a produção de textos.

### **2.3 AVALIAÇÃO E TECNOLOGIA**

A disponibilização do pré-teste em computadores e dispositivos tecnológicos móveis se alinha com uma tendência de trabalhos sobre usos de tecnologia na educação, inclusive na área de avaliação (VIBERG, 2015; MARTINS et al., 2017). Podemos dizer que o uso de

tecnologias afetou o ensino de línguas basicamente com a mesma força com que afetou a avaliação neste contexto (CHAPELLE & VOSS, 2016, p. 116). Mais especificamente, a acessibilidade do pré-teste via DTMs, e não apenas computadores, vislumbra um cenário em que *smartphones* se tornaram bastante populares no Brasil (MANS, 2016), sendo que o maior número de acessos à internet já se dá por meio deles (GOMES, 2016).

Conforme Pathan (2012), a tecnologia possibilita várias vantagens para a aplicação de testes, como a superação de obstáculos logísticos associados com transporte de testes e de resultados, a não necessidade de os candidatos estarem todos no mesmo horário e local para realizarem a prova, uma atitude mais positiva em relação ao teste, dentre outros. Em nosso caso, com a disponibilização do pré-teste em DTMs, poderemos beneficiar professores com tais facilidades antes de se comprometerem com a inscrição para o exame EPPL, que, atualmente, requer a presença física do candidato em horário e local determinados.

Chapelle & Voss (2016) publicaram um artigo que remonta os últimos 20 anos de pesquisa na área de avaliação e tecnologia no ensino de línguas, coletando suas informações de textos publicados neste período na revista *Language Learning and Technology* (LLT). Nesse texto, os autores percebem duas grandes tendências de estudos na área de avaliação: tecnologia para a eficiência e tecnologia para a inovação. Tecnologia para a eficiência diz respeito à sua utilização para diminuir gastos e recursos com o processo de avaliação e de correção, diminuir o impacto ambiental das avaliações, flexibilizar horários e locais de aplicação, dentre outros. Já no caso da inovação, há argumentos de que a utilização de tecnologia não é apenas uma transposição do que acontece em testes em papel, mas todo um novo formato e modo de produzir e realizar testes, o que resultaria na reconceptualização do construto de habilidade em segunda língua como “habilidade de selecionar e utilizar linguagem apropriada através de tecnologias que são apropriadas para uma situação” (CHAPELLE & DOUGLAS, 2006, p. 107 apud. SUVOROV & HEGELHEIMER, 2014).

Em vídeo sobre o uso de tecnologia em avaliação de línguas, Chapelle (2010) reconhece três motivações para tal utilização: eficiência, equivalência e inovação. Uma vez que eficiência e inovação já foram tratadas no parágrafo anterior, voltamo-nos para a motivação da equivalência, que diz respeito ao intuito de reproduzir o alto padrão de uma já respeitada avaliação em papel para o formato eletrônico.

No caso do pré-teste, trabalhamos principalmente com eficiência, uma vez que estamos preocupados com a acessibilidade, a praticidade e o baixo custo durante todo o processo, desde o desenvolvimento, até a aplicação, correção, e envio do *feedback*.

Poderíamos argumentar também em direção a um traço de inovação, especialmente na escolha de aplicar o pré-teste também através de DTMs, prática ainda pouco pesquisada e desenvolvida no Brasil e no mundo.

Ainda não há muitos estudos sobre o uso de DTMs no ensino, aprendizagem e avaliação de línguas no Brasil. Recentemente, Marques-Schäfer e Rozenfeld (2018) organizaram um livro em torno do tema, *Ensino de línguas e tecnologias móveis: políticas públicas, conceitos, pesquisas e práticas em foco*. Os DTMs já tem enorme importância na vida cotidiana, e acreditamos que, com mais pesquisas, terão também crescente importância na educação. A tese de Costa (2013) é também um exemplo de trabalho com MALL (*Mobile-assisted Language Learning*) no ensino de línguas. No entanto, a legislação brasileira ainda não acompanha a integração de DTMs a práticas educativas, mas as pesquisas no Brasil e no exterior apontam para este caminho, em que há a associação de técnicas e atividades da vida exterior dos estudantes à vida escolar e à aprendizagem.

Internacionalmente, estudos como o de Viberg (2015) em que a autora analisa a utilização de DTMs na educação à distância, com estudantes suecos e chineses. Kukulska-Hume & Viberg (2018) publicaram também um artigo com o estado da arte da utilização de DTMs na aprendizagem de línguas colaborativa. Guo (2014) fez uma análise de vários aplicativos para *smartphone* focados em aprender e desenvolver a linguagem oral em LI. Na área de avaliação com DTMs não há tantas produções, mas ainda podemos ressaltar Tarighat & Khodabakhsh (2016), que realizaram um trabalho de avaliação de proficiência oral em celulares, e Laborda, Royo & Bakieva que investigaram as reações e impressões de candidatos na aplicação de um exame de proficiência universitário em *tablets*. O exame era tradicionalmente aplicado em computadores e sua transposição para DTMs foi estudada com o objetivo de diminuir os custos de aplicação.

O pré-teste é, principalmente, um teste de compreensão oral, cujas respostas são digitadas no computador ou aparelho celular (devido à impossibilidade de serem gravadas oralmente pelas plataformas gratuitas disponíveis). Olhamos, então, para algumas peculiaridades do texto digitado na internet, especialmente através do celular. Primeiramente, nesse dispositivo, alguns aspectos inerentes a ele o diferenciam da escrita em papel e no computador, como a velocidade de digitação, que pode variar com maior familiaridade do usuário com o aparelho e também de acordo com o teclado disponível (teclado numérico, QWERTY virtual ou analógico, dentre outros). Além disso, as autocorreções habilitadas em vários modelos atuais de celulares são diferentes da autocorreção em computadores, por

exemplo, pois as do celular costumam ser mais “intrusivas” na medida em que mais palavras são corrigidas sem o usuário muitas vezes perceber, e o autocorretor está ativo em quase todos os aplicativos, diferente do computador, em que, muitas vezes, o auto corretor funciona apenas no processador de texto. Outra característica do celular, que é compartilhada pelo computador em alguns pontos, é interação com vários aplicativos ao mesmo tempo, sendo que o usuário pode receber notificações enquanto estiver digitando, dentre outros.

### **2.3.1 CALT – *Computer-Assisted Language Testing***

A facilidade com que muitos tipos de itens de testes podem ser adaptados a uma versão eletrônica, via computador, é evidente no enorme número de testes disponíveis na internet. A área de testes eletrônicos já vem sendo estudada há anos, sendo que, em 2001, a revista *Language Learning & Technology* publicou um volume especial completo sobre a teoria e prática de avaliação eletrônica (*computer-based language testing*).

De acordo com Suvorov & Hegelheimer (2014), o que hoje chamamos de *Computer-Assisted Language Testing* (CALT), ou Avaliação de línguas assistida por computador, é, por definição “qualquer avaliação aplicada através de um computador ou dispositivo móvel”<sup>10</sup> (SUVOROV & HEGELHEIMER, 2014, p. 2, tradução nossa).. Assim, o pré-teste e o EPPL (quando aplicado em sua versão eletrônica) fazem parte dessa tendência no uso de tecnologias para desenvolvimento e aplicação de avaliações de línguas.

Exploramos agora alguns motivos para se utilizar tecnologia em avaliações. Certamente o primeiro argumento, que já foi mencionado, é a eficiência em comparação a testes em papel. A diminuição de recursos e aumento da flexibilidade são motivos bastante relevantes para se migrar testes para o meio eletrônico, ou desenvolvê-los já para esse ambiente. No entanto, ainda não é sabido com precisão o quanto a tecnologia afeta o construto de testes eletrônicos, ou o quanto indivíduos submetidos a certo teste eletrônico se beneficiam de habilidades no manuseio do computador (DOUGLAS & HEGELHEIMER, 2007). Nesse sentido, Chapelle & Douglas (2006) questionam se os custos de utilizar tecnologia justificam os ganhos em eficiência e precisão, uma vez que a migração de testes em papel para eletrônicos deve alcançar uma medida *melhor* do construto, e não apenas uma medida mais eficiente.

Além disso, Douglas (2000, p. 275) afirma que “avaliação de línguas motivada pelo uso da tecnologia, ao invés da tecnologia sendo empregada a serviço da avaliação de línguas,

---

<sup>10</sup> No original: “(...) any test delivered via a computer or a mobile device”.

provavelmente nos levará a um caminho que não deveríamos traçar”<sup>11</sup>. É preciso, portanto, que a tecnologia sirva a um propósito da avaliação de línguas. No pré-teste, sua função principal é torná-lo acessível. Dessa maneira, devemos estar bem atentos ao objetivo do uso da tecnologia na avaliação e ao seu impacto, positivo e negativo, que analisaremos na seção de discussão.

Na literatura sobre avaliação, já é conhecido o desequilíbrio entre as principais características positivas de um teste quando se trata de priorizar uma em inevitável detrimento da outra (praticidade, confiabilidade, validade, autenticidade, efeito retroativo, conforme BROWN, 2004). Assim, um aumento, por exemplo, na validade de um teste, pode resultar em uma diminuição na praticidade. No caso do pré-teste, por ser um teste de baixa relevância, a preocupação principal com a acessibilidade e flexibilidade (que podem ser abarcadas pelo conceito de “praticidade”) pode vir a gerar uma redução de outros fatores, conforme discutiremos também no próximo capítulo do texto.

Sobre a tipologia e características de testes eletrônicos, Suvorov & Hegelheimer (2014) trazem, em seu artigo, uma lista de nove atributos para descrevê-los. No quadro 7 abaixo, trazemos uma caracterização do pré-teste baseada nos fatores propostos pelos autores.

#### **Quadro 7 – Atributos do pré-teste em relação à tecnologia**

<b>Atributo</b>	<b>Categoria a que pertence o pré-teste</b>
Direcionalidade	Linear
Formato de distribuição	<i>Web-based</i>
Densidade de mídia	Única mídia
Habilidade alvo	Compreensão oral
Mecanismo de correção	Correção humana
Relevância	Baixa
Propósito	Não-curricular (proficiência/diagnóstico)
Tipo de resposta	Construída
Tipo de tarefa	Produtiva

Fonte: SUVOROV & HEGELHEIMER, 2014, p. 2.

Nos parágrafos seguintes, teceremos breves comentários sobre essas categorias, baseando-nos no artigo de Suvorov & Hegelheimer (2014), fazendo uma descrição de cada aspecto e estabelecendo sua relação com o presente estudo. Quanto à **direcionalidade**

<sup>11</sup> No original: “(...) language testing that is driven by technology, rather than technology being employed in the service of language testing, is likely to lead us down a road best not traveled”

(*directionality*), um CALT pode ser linear, semi-linear ou adaptativo. Linearidade significa que todos os itens são apresentados na mesma ordem a todos os participantes da avaliação. Já a adaptatividade se tornou possível através da utilização de computadores, que podem selecionar o próximo item de uma avaliação dentre os itens de um banco, com base na resposta do candidato a um item anterior. Assim, uma resposta correta leva o computador a sugerir um item de maior dificuldade, enquanto que uma incorreta leva ao caminho oposto. Devido a algumas limitações de testes adaptativos (como o alto custo e problemas com algoritmos de seleção de itens), foram propostos testes em que partes maiores do teste (e não itens) são submetidos à adaptação individual. Isso pode acontecer também com a avaliação toda, caso em que um pré-teste é aplicado para determinar o nível da avaliação principal. Tais avaliações são categorizadas como semi-adaptativas. O pré-teste é claramente um exemplo de teste linear, uma vez que todas as suas questões são propostas a todos os candidatos, na mesma ordem, sem adaptações individuais.

Tratando-se do **formato de distribuição** (*delivery format*), um teste eletrônico pode ser *computer-based* ou *web-based*. Testes do tipo *computer-based* são CDs ou DVDs contendo um *software* que deve ser instalado individualmente em cada máquina a ser utilizada em sua aplicação. Em contrapartida, testes *web-based* estão disponíveis *online* e podem ser acessados através de qualquer máquina. O pré-teste é *web-based*, pois está disponível em qualquer dispositivo com acesso à internet. Por ser uma característica importante do pré-teste, exploraremos mais as peculiaridades deste tipo de teste na próxima seção do texto.

A **densidade de mídia** (*media density*) do pré-teste é única, em oposição a multimídia, uma vez que seus itens são áudios, sem acréscimo de figuras ou vídeos. Por ser um teste de compreensão oral, e por haver debate sobre a utilização de auxílio visual na avaliação dessa habilidade, preferimos optar pelo modelo mais tradicional apenas de escuta.

A compreensão oral é, como mencionamos, a **habilidade alvo** (*target skill*). Embora o pré-teste vise prever a proficiência oral no geral (a ser confirmada no EPPL oral, que avalia tanto a produção como a compreensão), apenas a compreensão oral é de fato foco dos itens do pré-teste, uma vez que há impossibilidade de gravar a produção do aluno na plataforma *Google Forms*, que foi por nós utilizada.

É fato que já há pesquisas com correção automática de itens abertos, muitas vezes revisados por corretores humanos, embora tal tecnologia ainda seja dispendiosa (DEANE, 2013). No caso do pré-teste, tal alternativa não pôde ser sequer considerada, já que avaliamos a habilidade de compreensão oral indiretamente. Se nosso objetivo fosse avaliar a escrita dos

participantes, poderíamos pensar nessa alternativa, entretanto o objetivo é avaliar a compreensão oral, da qual esperamos encontrar evidências nas respostas escritas pelos candidatos. As habilidades de compreensão, isoladamente, não são passíveis (ainda?) de correção automatizada. Assim, o **mecanismo de correção** (*scoring mechanism*) do pré-teste é humano.

Em outros momentos, já consideramos a **relevância** (*stakes*) do pré-teste, que é baixa, já que é um teste com o objetivo apenas de informar o candidato ao EPPLÉ de suas chances de sucesso, e não serve a nenhum propósito de certificação.

O **propósito** (*purpose*) do pré-teste, conforme mencionado acima, é fornecer essa informação prévia ao candidato sobre sua proficiência. É um teste com características de teste de proficiência e de teste de diagnóstico, na medida em que avalia a proficiência do candidato de maneira comunicativa, mas não a certifica, e sim oferece um diagnóstico, ou *feedback*, da probabilidade de ser bem classificado no EPPLÉ. Além disso, o *feedback* pode servir um propósito direcionador de futuros estudos do candidato em prol da melhoria de sua proficiência. O propósito se classifica, assim, em não-curricular.

Embora não avaliemos a performance do candidato no pré-teste, uma vez que não há produção, e sim compreensão oral, o **tipo de resposta** (*response type*) é construída. Os participantes devem construir um pequeno texto atendendo ao que as perguntas orais requerem, com base também nos textos orais disponibilizados. Reiteramos mais uma vez que essas respostas não são analisadas em termos do mérito da habilidade escrita, uma vez que este não é o construto do pré-teste.

O **tipo de tarefa** (*task type*) é, por fim, produtivo, uma vez que aos participantes é solicitada uma resposta curta de produção escrita.

Encerrada a classificação do pré-teste dentro dos atributos propostos por Suvorov & Hegelheimer (2014), voltamo-nos ao tema de desenvolvimento de testes eletrônicos. Atualmente, há uma variedade de plataformas disponíveis a educadores e desenvolvedores para o desenho de testes de baixa ou média relevância (em casos de alta relevância, há problemas de segurança relacionados a se hospedar um teste diretamente na web). Dentre as possibilidades, há, por exemplo, ambientes virtuais de aprendizagem ou aplicativos específicos ao fim de desenvolvimento de avaliações. Podemos chamar tais plataformas de **ferramentas de autoria** (*authoring tools*). Abaixo a definição de Kessler (2013):

Ferramentas de autoria, que incluem uma gama de programas e sites (*software*) desenvolvidos para a criação de material para computadores ou internet, são onipresentes no mundo de desenvolvimento da Web, incluindo o ensino de línguas. Tais ferramentas, elaboradas especificamente para a avaliação de línguas, permitem que professores, desenvolvedores de material e administradores de departamentos de línguas criem uma variedade de avaliações eletrônicas e interativas”<sup>12</sup> (KESSLER, 2013, p. 302, tradução nossa)

Chapelle & Douglas (2006) reconhecem que, embora as ferramentas de autoria sejam bastante flexíveis em alguns pontos, elas geralmente apresentam limitações inerentes ao seu design e funcionalidade. Assim, é necessário um alinhamento entre tais ferramentas e as necessidades da avaliação e recursos disponíveis. Dentre as plataformas gratuitas oferecidas, merecem destaque o *Moodle* e o *Google Forms*, sobre os quais discorreremos brevemente, chamando a atenção para seus pontos fortes e fracos.

O *Moodle* (<http://moodle.org/>) é um ambiente virtual de ensino e aprendizagem que permite gerenciamento de cursos (CMS – *Course Management System*) e ferramentas para avaliação. Dentre as possibilidades, destacamos a variedade na criação de itens (verdadeiro ou falso, múltipla escolha, *cloze*, respostas abertas curtas e longas) e na disponibilização de *feedback* (modo interativo, *feedback* imediato, dentre outros). O *Moodle* oferece também integração com algumas outras plataformas, como o *TaskChain*. Para gravação e reprodução de áudio, o *Moodle* não oferece ferramenta interna ao seu sistema, embora Kessler (2013) argumente que inserção do módulo *Gong* no ambiente virtual abre para essa possibilidade. Outro problema do *Moodle* é o design complexo e pouco intuitivo a novos usuários.

Utilizado neste trabalho, o *Google Forms* impressiona inicialmente pela sua simplicidade e *design* intuitivo a usuários e desenvolvedores. Associado ao *Google Docs*, que é um processador *online* de documentos, o *site* permite a criação e publicação de questionários e testes na rede. Há uma gama de tipos de questões disponíveis (múltipla escolha, respostas abertas curtas e longas, correspondência entre colunas), e há também a possibilidade de modificar o *layout* da página, para que fique mais atraente ao tipo de audiência. O *Forms* gera um relatório com as respostas dos participantes e um gráfico que as sintetiza. Dentre os problemas do *Forms*, estão a impossibilidade de coletar áudio dos

---

<sup>12</sup> No original: “Authoring tools, which include an array of software designed to create material for computers and the Internet, are ubiquitous throughout the Web development world, including language education. Such tools, designed specifically for language assessment, allow language teachers, material developers, and language department administrators to create a variety of interactive computer-based assessments” (KESSLER, 2013, p. 302)

participantes, a não disponibilização de um relógio dentro do teste e a não possibilidade de muitas mudanças de formato, como, por exemplo, apresentar uma questão de cada vez.

Voltamos a argumentar, com Chappelle & Douglas (2006), que a escolha de uma plataforma deve ser feita levando em conta o propósito do teste em relação às limitações de cada ferramenta. O *Forms* foi escolhido por nos oferecer acessibilidade, nenhum custo de hospedagem, número ilimitado de usuários, facilidade de manuseio e total compatibilidade com DTMs, apesar de suas limitações, como os tipos de itens que podem ser criados.

### 2.3.2 Avaliação *web-based*

Avaliações ditas *web-based*, em oposição a *computer-based*, são aquelas hospedadas diretamente *online*, sem a necessidade de instalação de um programa específico que processe o teste. Podem ser acessadas através de qualquer máquina com conexão à internet. *Computer-based*, por sua vez, são avaliações que exigem a instalação de um programa em cada máquina, e não estão disponíveis na Web.

De acordo com Roever,

Definidos com simplicidade, os *web-based language tests* (WBT) são avaliações do tipo *computer-based* disponibilizadas através da *World Wide Web* (WWW). WBTs compartilham de muitas características de *computer-based tests* (CBTs), mas a utilização da Web como o meio de serem ministradas adiciona vantagens específicas ao mesmo tempo em que complica alguns aspectos (ROEVER, 2001, p. 84, tradução e grifos nossos)<sup>13</sup>

Uma vantagem de se utilizar WBTs é que são fáceis de construir e manter, pois não há necessidade de contratação de um servidor, o que geralmente implica em contratação de programadores. É, portanto, uma abordagem preferível quando o contexto é de baixa escala: espera-se pouca quantidade de dados, não há necessidade de adaptatividade, bancos de itens são pequenos e os desenvolvedores tem interesse em se manter independentes de profissionais da computação.

Sobre WBTs, Roever (2001) afirma também que são mais apropriados para testes de baixa relevância, como é o caso de checagem de aprendizagem, preparação para outra avaliação, ou atividades de pesquisa, ou seja, situações em que o teste está a serviço da aprendizagem ou do fornecimento de *feedback* para que os candidatos saibam quão próximos estão de chegarem a um objetivo predeterminado. Em todos os contextos citados, todas as

---

<sup>13</sup> No original: "Simply defined, a Web-based language test (WBT) is a computer-based language test which is delivered via the World Wide Web (WWW). WBTs share many characteristics of more traditional computer-based tests (CBTs), but using the Web as their delivery medium adds specific advantages while their delivery medium complicates matters" (ROEVER, 2001, p. 84).

vantagens de WBTs são mantidas: os participantes têm pouco ou nenhum motivo para “colar”, seriam beneficiados pela possibilidade de fazer o teste sem as restrições logísticas de testes em papel (quando quiserem, onde quiserem, e no seu próprio ritmo) e os custos para os desenvolvedores são baixos ou inexistentes.

Ainda com Roever (2001), WBTs são excelentes quando usados como testes de pré-nivelamento para testes de alta relevância, de modo que os candidatos possam decidir de maneira informada se são necessários mais preparação e estudos para que seu desempenho na avaliação posterior seja satisfatório. Assim, a escolha de uma plataforma hospedada *online* para o pré-teste se justifica totalmente na perspectiva desse autor: é um teste de baixa relevância, voltado para o diagnóstico da proficiência do candidato, cujo *feedback* pode ajudar no direcionamento dos estudos, e que informa sobre as chances de sucesso em outro exame. Não há motivos para trapacear, uma vez que o teste não oferece certificação, e um diagnóstico correto é o que pode ser oferecido de interesse ao participante.

O autor ainda resume as vantagens de se utilizar WBTs de tecnologia mais simples: o princípio da assincronia; produção em nível local; e produção não onerosa. Respectivamente, assincronia diz respeito à flexibilidade do teste, que pode ser feito a qualquer hora e em qualquer lugar, e é apontada como um dos pontos mais fortes desse tipo de teste. Produção em nível local foi nossa tradução do inglês, nas palavras de Roever (2001), de “*testing goes grassroots*”. Esse aspecto diz respeito à independência dos desenvolvedores de profissionais de programação, o que não acontece no caso dos CBTs. Por fim, a produção não onerosa: em comparação a testes de papel, já exploramos o assunto, e em comparação a CBTs também há uma economia de recursos, uma vez que não há necessidade de programadores e, muitas vezes, as aplicações podem ser feitas de forma mais simples, sem, por exemplo, a necessidade de mídias para instalação de programas nos computadores.

Em contrapartida a essas vantagens, alguns desafios no desenvolvimento e aplicação de WBTs incluem a inserção de variação irrelevante ao construto – especialmente devido ao nível distinto de familiaridade com computadores dos participantes do teste – e a possibilidade de falhas repentinas nas máquinas em que as aplicações são realizadas. Além disso, talvez o maior problema de WBTs para Roever (2001) seja a impossibilidade, a um baixo custo, de gravação da produção oral do participante. É essa também a maior limitação que encontramos, mesmo mais de quinze anos após a publicação desse artigo.

Da mesma forma que no caso das vantagens, o autor cita também uma lista de desvantagens de se utilizar WBTs. Dentre elas, destacamos: fraude e exposição de itens;

armazenamento de dados; falhas no servidor; e incompatibilidade com diferentes navegadores. É uma grande desvantagem de WBTs a falta de segurança contra “cola” e a exposição das tarefas, uma vez que é impossível assegurar de quem está fazendo o teste está sozinho, ou é realmente a pessoa que diz ser. Além disso, nada impede que o candidato faça consultas, ou leia as questões, estude, e depois volte para realizar o teste. Quanto ao armazenamento de dados, Roever (2001) faz menção à possibilidade de uma falha no sistema durante o teste, fazendo com que o candidato perca as informações fornecidas até então. Isso acontece porque o teste é feito com a utilização de um navegador, que é um programa sem permissão para escrever no disco rígido do computador em questão. Como solução, o autor discute algumas possibilidades, em itens menores, de se utilizar os *cookies* de alguns navegadores como mecanismo de armazenamento. A falha no servidor significa que o *site* onde o teste está hospedado não está funcionando, o que, nos dias de hoje, é cada vez mais incomum. Já a incompatibilidade com diferentes navegadores, ou mesmo versões mais ou menos atuais do mesmo navegador, ainda é um desafio presente, para o qual chamamos a atenção. Pode acontecer quando o navegador não processa o teste, ou partes do teste, como vídeos, ou desconfigura o *layout* da página, ou faz com que funções da página não respondam apropriadamente.

Ao fim de seu texto, o autor argumenta a favor da decisão de se utilizar WBTs de acordo com a relevância (*stakes*) de um instrumento de avaliação. Testes de alta relevância são os que possuem maior impacto na vida dos candidatos, oferecendo oportunidades para que sejam admitidos em um emprego ou programa de estudos. Tais testes, para Roever (2001), seriam melhor aplicados em contextos mais fechados do que a Web, como uma rede segura interna. Em se tratando de avaliações de média relevância, essas incluem exames finais de cursos, testes de nivelamento para estudantes estrangeiros, e outras situações que afetam a vida de quem participa do teste, mas não de maneira tão impactante como no caso dos testes de alta relevância. Para esses casos, Roever (2001) prossegue dizendo que poderiam ser usados, como no caso de cursos a distância, ou de estudantes que poderiam fazer o teste antes de se mudar para uma instituição de ensino em outro país. Outra possibilidade seria realizar os testes com supervisão. Por fim, os testes de baixa relevância são qualquer instrumento focado em dar *feedback* ao desempenho dos candidatos, especialmente no que diz respeito a direcioná-los a aprendizagem ou a informá-los sobre quão próximos estão de atingir um determinado objetivo. Roever (2001) inclui também aqui a possibilidade de se utilizar WBTs para a pesquisa com testes de línguas, já que os instrumentos são bastante flexíveis, e que

vários tipos de dados podem ser coletados da interface *online* (como, por exemplo, quanto tempo em média os indivíduos demoram para responder certa questão). Para o pesquisador, portanto, testes de baixa relevância são ideais para serem desenvolvidos como WBTs, uma vez que não trapacear é de interesse dos participantes, e o teste pode usufruir de todas as vantagens já citadas de ser hospedado na internet.

O pré-teste ao EPPL, por pertencer a esse grupo de testes de baixa relevância e ter como propósito apenas a previsão do resultado em um exame, se alinha à descrição de Roever de que

instrumentos na web podem ser usados para a preparação para testes, tanto para exames grandes e padronizados quanto para testes de pré-nivelamento de estudantes se preparando para ingressar em um programa de línguas estrangeiras ou uma universidade em outro país. Tal pré-nivelamento informará os candidatos uma noção geral sobre como será seu desempenho na avaliação em questão para que possam decidir se é necessária uma preparação adicional<sup>14</sup> (ROEVER, 2001, p. 91, tradução nossa)

Discorreremos mais sobre aspectos específicos ao pré-teste em nossa análise de dados, voltando a algumas das questões descritas aqui, explorando como se deram as aplicações do pré-teste nos contextos estudados e sua relação com o fato de o teste ter sido hospedado *online*.

Por fim, acreditamos que é válido lembrar que a tecnologia, qualquer que seja, não faz um bom teste, não importa o quão sofisticado é o *script* ou, o quanto o site é chamativo, ou quão bem feitas são as animações. No entanto, é inegável que o uso de equipamentos tecnológicos na área da avaliação expandiu a exploração de maneiras diferentes de se desenvolver e aplicar testes, com várias vantagens, e abrindo caminho para que seja um dia, possivelmente, o meio mais comum de se realizar uma avaliação.

## 2.4 VALIDADE

Validade é um conceito central na área de avaliação, pois trata-se, *grosso modo*, da medida em que um teste avalia aquilo que se propôs a avaliar. Ademais, Henning (1987) afirma que um teste é válido para um determinado propósito, o que significa que um teste deve ser julgado, com base em evidências, como mais ou menos válido, de acordo com sua finalidade. Assim, consideramos validade não como característica intrínseca ao teste, mas sim

---

<sup>14</sup> No original: “(...) instruments on the Web can be used for test preparation, either for large standardized tests or as pre-placement tests for students preparing to enroll in a foreign language program or a university in a foreign country. Such pre-placement will give test takers a general notion about how the students will perform on the test in question so that they can decide whether additional preparation is needed”

ao seu uso para tomar certas decisões, baseado em que construto decidiu-se que seria adequado medir, e no quanto o teste realmente avalia esse construto.

Portanto, tradicionalmente, validade refere-se “à adequação de um teste ou de algum de seus componentes como uma medida do que esse teste deve medir” (HENNING, 1987, p. 89 apud. ALDERSON; CLAPHAM; WALL, 1995, p. 170). Esta visão tradicional passou por reelaborações no decorrer dos anos, e o conceito deixou de ser uma característica do próprio teste, para vir a se tornar um argumento relativo à sua interpretação e uso, e como estes podem ser justificados.

Ademais, a confiabilidade, entendida originalmente como um aspecto distinto da validade, passou a ser entendida como evidência desta (CHAPELLE, 1999 apud SCARAMUCCI, 2011). A confiabilidade de uma avaliação diz respeito à sua consistência, ou seja, se um candidato fizer o mesmo teste espera-se que ele obtenha o mesmo resultado (excetuando-se o caso em que ele tenha melhorado sua habilidade e seu resultado diferente seja consequência disso). Para conseguir que um teste seja confiável, é preciso que se diminua a presença de erro, isto é, de mudanças assistemáticas no ambiente, nos avaliadores, dentre outros (ALDERSON; CLAPHAM; WALL, 1995, p. 87). Quanto menos esses fatores externos interferirem na avaliação, mais confiável ela é, uma vez que medirá apenas o que pretende medir: em nosso estudo, a proficiência linguístico-comunicativo-pedagógica oral. Assim sendo, é essencial que um teste seja confiável para que seja considerado válido, pois a confiabilidade funciona como um dos indicativos de validade, uma vez que “[u]m teste não pode ser válido sem antes ser confiável (consistente e estável)” (SCARAMUCCI, 2011, p. 105).

Acreditamos ser adequado fazer um breve percurso histórico do conceito de validade, já que, como mencionamos, este sofreu várias mudanças no decorrer do tempo. Faremos isso nos parágrafos seguintes, nos baseando majoritariamente em Kane (2001).

Na década de 1950, acreditava-se que as variáveis a serem medidas por um teste tinham um valor definido para cada pessoa, e o papel da avaliação era simplesmente estimar esse valor com a maior precisão possível. Desse modo, era comum a utilização de uma medida que forneceria esse valor “real”, ou pelo menos uma aproximação dele, como **critério** contra o qual se compararia o teste a ser validado. A pergunta feita era: quanto os resultados do teste a ser validado preveem os resultados do critério? Esse tipo de validade foi chamado posteriormente de validade de critério, sendo subdividido em validade preditiva e validade paralela. A primeira acontece quando o teste a ser validado é aplicado anteriormente ao

critério, enquanto que a segunda presume que os resultados do critério e do teste a ser validado são coletados simultaneamente. Esta validade pode ser bastante útil quando uma medida adequada de critério está disponível (KANE, 2001), e será explorada mais detalhadamente na próxima subseção, visto que é o foco deste trabalho.

Muitas vezes, no entanto, é difícil estabelecer uma medida de critério bem definida ou, mesmo no caso de possuí-la, às vezes há dúvidas sobre a sua própria validade. Face a esses desafios, estudiosos passaram a avaliar também o **conteúdo** dos testes, se o teste abrangia a totalidade do conteúdo pretendido, e se era relevante para avaliá-lo (BACHMAN, 1990, p. 244). A decisão era tomada, nesse caso, com base em julgamento de experts: pessoas com experiência e reconhecimento na área examinavam as avaliações e emitiam sua opinião sobre elas. É desnecessário dizer que tal método era altamente subjetivo e de alta “tendência confirmatória”, uma vez que os experts muitas vezes eram os próprios desenvolvedores dos testes, ou seus colegas (KANE, 2001).

Para prosseguirmos para a validade de **construto**, precisamos, primeiramente, definir este termo:

Quando se classificam situações, pessoas ou reações, são utilizados **construtos**. O termo **conceitos** poderia ser usado em seu lugar, mas **construto** enfatiza a ideia de que as categorias são criações deliberadas para organizar a experiência em afirmações gerais ('leis'). Alguns construtos são muito próximos da matéria da experiência comum – vermelho, por exemplo, ou ser humano.... Alguns construtos (por exemplo, daltonismo) estão integrados em teorias bem articuladas e bem substanciadas<sup>15</sup> (CRONBACH, 1971, p. 462 apud BACHMAN, 1990, p. 255, tradução nossa, grifo no original)

Portanto, um construto é um conceito articulado dentro de uma teoria, conectado a outros com os quais se interrelaciona. Em Kane (2001), percebemos que, historicamente, recorrer à teoria se apresentou como uma alternativa quando acreditava-se não haver uma medida de critério satisfatória disponível. Fazia-se uma relação entre uma teoria sobre um traço de natureza inobservável e associava-se o traço a atributos observáveis e mensuráveis. Esperaria-se que, por exemplo, um teste de introversão proporcionaria resultados similares a um teste de timidez, sendo este um caso de evidência de validade convergente, uma vez que os construtos teóricos são relacionados positivamente. Ao contrário, se esperaria que um teste de introversão forneceria resultados opostos aos de um teste de sociabilidade, pois são

---

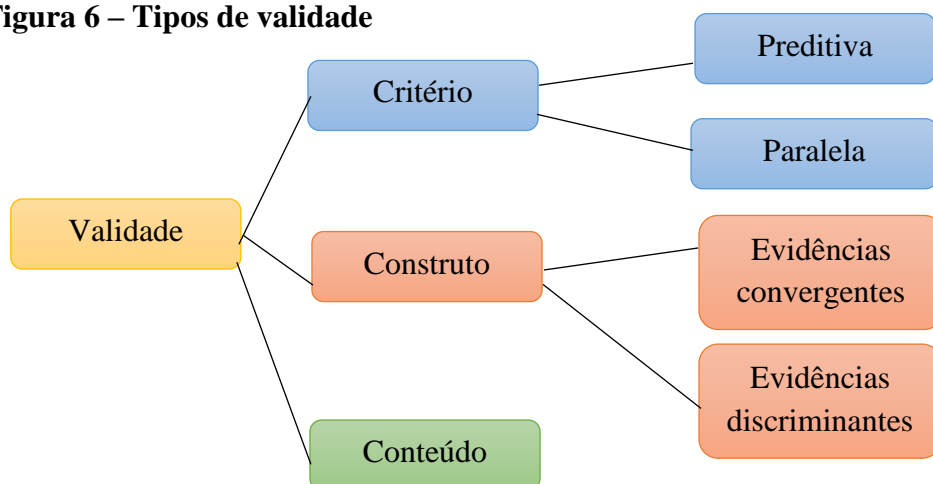
<sup>15</sup> No original: “Whenever one classifies situations, persons, or responses, he uses **constructs**. The term **concepts** might be used rather than **constructs**, but the latter term emphasizes that categories are deliberate creations to organize experience into general law-like statements. Some constructs are very close to the stuff of ordinary experience – red, for example, or human being... Some constructs (e.g., red-green color blindness) are embedded in well-articulated, well-substantiated theories”.

construtos teóricos opostos, sendo, assim, evidência de validade discriminante. Cronbach (1971 apud KANE, 2001) afirma que qualquer medida de processos individuais internos (ansiedade, por exemplo) invariavelmente requer validação de construto, que é a medida de quanto os resultados encontrados no teste satisfazem a teoria.

Kane (2001) descreve o entendimento dos tipos de validade no período de 1950 a 1970 como uma *toolbox* (caixa de ferramentas), na medida em que os pesquisadores lançavam mão dos vários tipos de validade de acordo com seus trabalhos e bom senso. Em tal abordagem, havia potencial para escolhas altamente oportunistas, ou, no mínimo, dúbias, de validades que, ao estudioso, se mostravam mais adequadas, simples ou promissoras para resultados positivos em seus trabalhos.

No esquema abaixo, podemos visualizar os tipos e subtipos de validade.

**Figura 6 – Tipos de validade**



Fonte: elaborado pela autora.

Com o tempo, alguns princípios de validação emergiram como universais:

1. necessidade de análise extensa no processo de validação, incluindo o desenvolvimento de uma teoria, de procedimentos de medida de seus construtos, de hipóteses baseadas na teoria, e a comparação destas com observações;
2. necessidade de uma declaração explícita da interpretação proposta, uma vez que a variável deve ser explicada, ela não preexiste pronta para ser medida;
3. necessidade de desafiar as interpretações propostas e considerar alternativas.

Assim, o que antes era entendido como um conjunto de tipos de validade, dentre as quais se poderia escolher, agora é unificado em torno de um conceito central, sendo que os

então tipos de validade passaram a funcionar como possíveis evidências para que se valide um teste para certo propósito.

Desenvolvimentos mais recentes no conceito trouxeram discussões sobre a inclusão ou não de consequências como parte da validade de um teste, além de exporem mais minuciosamente possíveis trajetórias para a validação de testes (BACHMAN, 2005; XI, 2010). Em nosso caso, trataremos mais especificamente dos dois tipos de validade, ou de evidências, que buscamos examinar neste trabalho: as validades de critério, subdivididas em preditiva e paralela.

#### 2.4.1 Validades de critério

A literatura de subclassificações para o conceito de validade é vasta, e aqui nos voltamos para as categorias internas à chamada validade de critério (validade externa, para ALDERSON; CLAPHAM; WALL, 1995), que são validade preditiva e validade paralela. Conforme já mencionamos, a validade de critério está ligada à utilização de uma medida externa como referência para se validar uma nova medida. Reiteramos, também, que não propomos neste trabalho um processo de validação, mas apenas a análise de um aspecto de (in)validade do pré-teste.

Algumas vezes, os limites entre as duas tipologias de validade de critério não são tão claramente definidos. Em geral, fala-se em **validade preditiva** quando a medida que se quer validar prevê um resultado na medida que é referência (critério). Já a **validade paralela** pressupõe que duas medidas se alinhem no mesmo momento. Portanto, podemos dizer que a diferença entre elas está no intervalo de tempo da administração da medida referência e da medida que se quer validar.

Para organizarmos uma visão geral dos dois conceitos, contrastando-os, apresentamos contribuições de alguns autores que se dedicaram ao tema da tipologia de validade. Para Alderson, Clapham e Wall (1995, p. 177), falamos em validade paralela ao compararmos “resultados do teste com outra medida dos mesmos candidatos coletada aproximadamente no mesmo período do teste”<sup>16</sup>, sendo que essa outra medida, utilizada como referência, pode ser um outro teste, uma outra versão do mesmo teste, auto-avaliações, observação de professores, dentre outros. Hughes (1989) chama a atenção para o mesmo aspecto simultâneo da validade paralela. Bachman (1990, p. 248) acredita que a análise de validade paralela se dá, comumente, por meio de uma das duas maneiras: “(1) análise das diferenças de desempenho

---

<sup>16</sup> No original: “test scores with some other measure for the same candidates taken at roughly the same time as the test”

num teste entre grupos de indivíduos com diferentes níveis de habilidade linguística, ou (2) análise das correlações entre várias medidas de uma dada habilidade”<sup>17</sup>. No caso desse autor, nossa pesquisa se enquadra na segunda abordagem, pois medimos no pré-teste e no EPPLE oral o construto de proficiência oral linguístico-comunicativo-pedagógica. Ebel & Frisbie (1991, p. 106) acrescentam que a utilidade de evidência paralela seria para “mostrar que os alunos aparecem em um mesmo ranque em duas medidas”<sup>18</sup>.

Em contrapartida, ao analisar o que os mesmos autores declaram sobre a validade preditiva, Alderson, Clapham e Wall (1995, p. 180) acreditam que sua diferença em relação à paralela está no fato de que “as medidas externas somente serão coletadas algum tempo depois da aplicação do teste”<sup>19</sup>. Essa diferença de tempo é enfatizada também pelos outros autores, sendo que Bachman (1990, p. 250) dá ênfase ao comportamento futuro, ao invés de falar em “medida externa” futura. Ebel & Frisbie (1991, p. 106) e Hughes (1989, p. 25) também enfocam esse comportamento futuro, mas utilizam a palavra “desempenho”. Os autores falam ainda em um contexto de uso da validade preditiva, como quando selecionando indivíduos para admissão, emprego, oportunidades educacionais, dentre outros, que configuram situações de seleção para desempenho de uma função. É possível dizer que essa previsão lembra muito o que se faz em testes de proficiência: prever o desempenho futuro de um participante.

A partir da perspectiva desses autores, entendemos, então, que tratamos de validade de critério paralela quando o teste e a medida externa são aplicados aproximadamente ao mesmo tempo, e que a validade preditiva pressupõe um intervalo de tempo maior. Para analisarmos esses dois tipos de validade de critério, portanto, precisaríamos de uma variação no tempo da medida do critério (EPPLE oral). Assim, nossa metodologia, discutida no capítulo anterior, incluiu aplicações do pré-teste acompanhadas do EPPLE em curto espaço de tempo (mesma data, intervalo de poucos dias) como dados para a análise da validade paralela. Já para a análise da validade preditiva, contamos com a aplicação de uma segunda versão do EPPLE a uma parcela do grupo MG (os que se voluntariaram, chamados MG<sub>s</sub>) que já havia feito o pré-teste e uma primeira versão do EPPLE um mês antes.

De acordo com Alderson, Clapham e Wall (1995), as pesquisas com validade preditiva apresentam alguns problemas. O primeiro deles é o problema da amostra truncada, pois como

---

<sup>17</sup> No original: “(1) examining differences in test performance among groups of individuals at different levels of language ability, or (2) examining correlations among various measures of a given ability”

<sup>18</sup> No original: “to show that students appear in the same relative rank order on the two measures”

<sup>19</sup> No original: “the external measures will only be gathered some time after the test has been given”

são aplicados muitas vezes de modo a pré-selecionar indivíduos para situações profissionais e estudantis, apenas a parcela de candidatos aprovada no teste 1 chega a fazer o teste 2, que conta, então, apenas com resultados dos melhores candidatos. Em segundo lugar, o problema de que, durante o intervalo entre o teste 1 e o 2, alguns participantes – ou todos – podem melhorar suas habilidades, e isso está fora do controle do pesquisador. Por fim, a dificuldade de se conseguir uma medida de referência da habilidade do candidato.

Procuramos evitar, em nosso trabalho, estes problemas. O primeiro na medida em que não selecionamos os candidatos que fizeram o pré-teste e o EPPL: todos foram convidados e incentivados a participar de todas as aplicações, uma vez que o contexto é de pesquisa, e não de seleção. Quanto ao segundo problema, o intervalo entre as aplicações para a análise da validade preditiva foi de um mês, que consideramos não ser tão longo que pudesse causar mudanças significativas na proficiência dos candidatos, mas longo o bastante para que configure uma previsão de desempenho futuro. Por fim, Alderson, Clapham e Wall (1995, p. 181) fazem uma ressalva quanto à terceira questão: “é improvável que estaria disponível uma medida externa adequada da habilidade dos estudantes no uso da língua inglesa no contexto da pesquisa, a não ser que fosse uma outra versão da avaliação inicial”<sup>20</sup>. Em nosso estudo, o pré-teste não é uma outra versão do EPPL, mas é um teste desenvolvido com o objetivo de se aproximar de seu construto e de seus resultados.

Assim, antes de prosseguirmos para o próximo capítulo, acreditamos que cabe aqui uma retomada dos conceitos apresentados nesse capítulo e sua aplicação dentro da pesquisa. As definições de proficiência e de proficiência linguístico-comunicativo-pedagógica, por exemplo, são fundamentais para o entendimento do construto do EPPL e do pré-teste. Além disso, características de pré-testagem e de testes diagnósticos foram analisadas em busca de uma melhor definição do formato do pré-teste e de como relacioná-lo com o EPPL, observando as características de cada instrumento. Também foram discutidos aspectos de tecnologia na área de avaliação de línguas, um tema central a este trabalho e que lançou luz sobre a nossa análise da comparação entre as duas tecnologias utilizadas, além de ter sido uma orientação para a etapa de desenvolvimento de nossa versão do pré-teste (PT2). O último tópico foram aspectos teóricos do conceito de validade, que, por sua complexidade, não seríamos capazes de abranger de maneira completa neste trabalho. Estes nos ajudaram a melhor compreender as evidências de validade, ou não, encontradas no pré-teste. Com essas considerações em mente, prosseguimos para a análise e discussão dos dados.

---

<sup>20</sup> No original: “(...) it is unlikely that a suitable external measure of the students’ ability to use English in the study setting would be available, unless it were another version of the original test”

# **ANÁLISE DE DADOS E DISCUSSÃO**

Dividiremos os resultados desta investigação em dois momentos principais, antes de, ao final, sintetizá-los em uma discussão sobre os dados e os desafios encontrados neste estudo. Primeiramente, analisamos os resultados do pré-teste em comparação com os do EPPL oral, tratando da medida em que há validade de critério entre os dois instrumentos. Também analisamos os questionários realizados pelos participantes, que dizem respeito à qualidade do pré-teste como instrumento de avaliação da proficiência linguístico-comunicativo-pedagógica. A seguir, trazemos uma exploração da utilização de tecnologia nas aplicações, tanto do ponto de vista das observações de campo feitas pela pesquisadora, quanto de dados coletados através dos questionários. Buscamos, também, contrastar os grupos que realizaram o pré-teste através do celular e do computador. Por fim, fazemos uma discussão sintetizando os resultados e descrevendo as limitações do estudo, e do próprio pré-teste, já vislumbrando possíveis encaminhamentos de pesquisa.

### **3.1 O PRÉ-TESTE E SEUS RESULTADOS**

O pré-teste foi aplicado, conforme já mencionamos, para dois grupos de estudantes em uma universidade pública no estado de São Paulo (grupos SP<sub>1</sub> e SP<sub>2</sub>) e em uma universidade pública no estado de Minas Gerais (MG). Em alguns pontos da análise, focamos apenas na diferença entre quem realizou o pré-teste através do celular e quem o realizou através do computador, o que resultou em dois grandes grupos, que convencionamos simplesmente chamar de, respectivamente, SP e MG. O grupo MG conta com 13 participantes e o grupo SP com um total de 19 integrantes.

#### **3.1.1 Impressões sobre o teste**

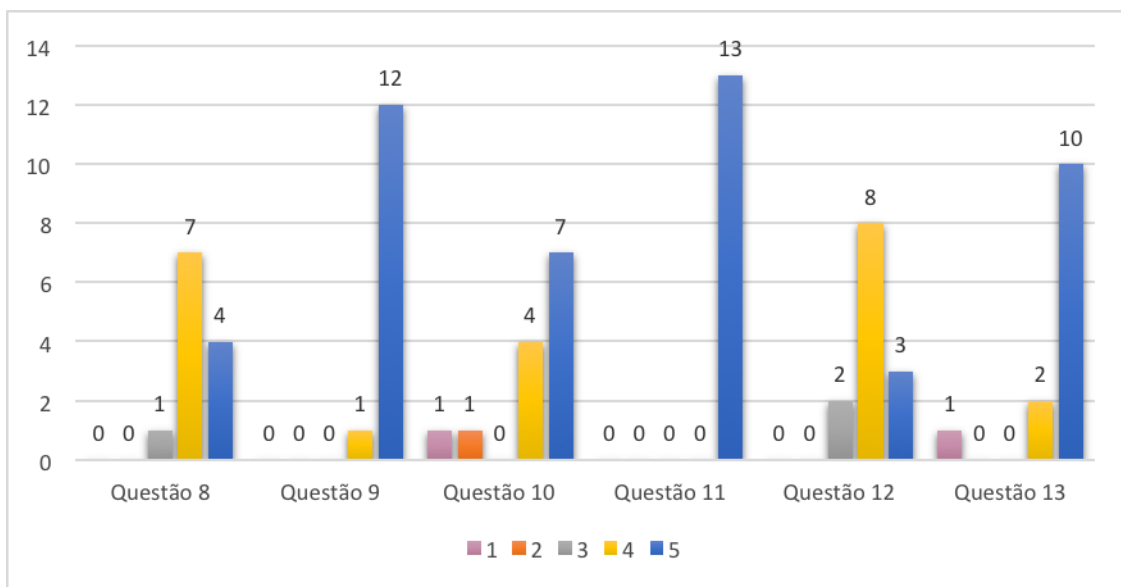
No final deste trabalho, em anexo, é possível encontrar o questionário aplicado aos participantes do estudo (p. 126). A primeira seção coleta informações sobre o perfil dos participantes e foi utilizada para montarmos a descrição dos sujeitos da pesquisa. A segunda seção é a de que trataremos nesse momento, pois apresenta questões sobre o pré-teste e sua qualidade em escala *Likert*. Por fim, o questionário traz questões abertas sobre o pré-teste, especialmente no que tange à tecnologia utilizada nas aplicações, sendo que os dados coletados nessa parte serão discutidos posteriormente, quando comentamos os aspectos tecnológicos.

Trazemos aqui a análise dos dados coletados na segunda seção do questionário, em que os participantes devem se posicionar diante das seguintes afirmações:

- 8) O pré-teste avaliou satisfatoriamente a minha proficiência em língua inglesa.
- 9) As orientações para a execução do pré-teste são claras.
- 10) O tempo disponibilizado para responder as questões do pré-teste é suficiente.
- 11) A plataforma do pré-teste (*Google Forms*) é de fácil utilização.
- 12) O pré-teste do EPPLÉ avalia adequadamente a proficiência linguística de (futuros) professores de língua inglesa.
- 13) A qualidade do áudio do teste permite boa compreensão de seu conteúdo.

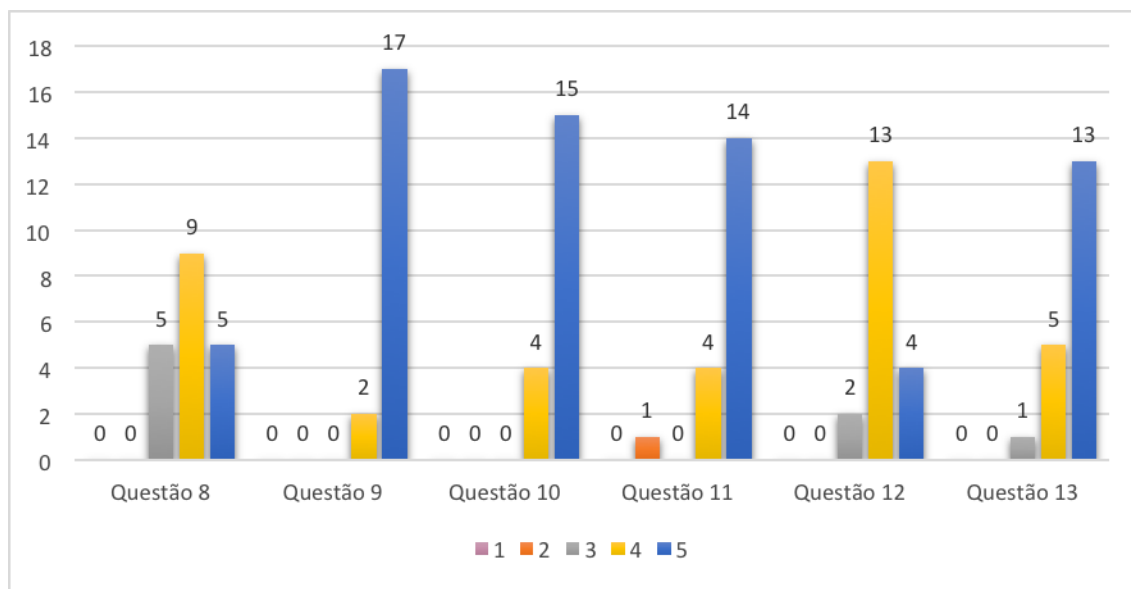
As possibilidades de resposta dos participantes são: discordo totalmente (1), discordo parcialmente (2), não concordo nem discordo (3), concordo parcialmente (4) e concordo totalmente (5). Os números entre parênteses servem de guia para as legendas dos gráficos 1 e 2 que apresentamos a seguir. Assim, quem concorda totalmente com alguma questão está representado na coluna azul (rotulada como “5”), quem concorda parcialmente está agrupado na coluna amarela (4), quem não concorda e nem discorda está na coluna cinza (3), quem discorda parcialmente está na coluna vermelha (2) e quem discorda totalmente está na coluna lilás (1).

**Gráfico 1 – Impressões sobre o pré-teste – MG**



Fonte: dados coletados por meio de questionário aplicado aos participantes.

**Gráfico 2 – Impressões sobre o pré-teste – SP**



Fonte: dados coletados por meio de questionário aplicado aos participantes.

Sobre a **questão 8**, em que o candidato avalia o quanto o pré-teste é um bom instrumento de medida de sua proficiência, a maioria das respostas, tanto do grupo SP quanto do MG, foi a “concordo parcialmente”. Acreditamos que isso provavelmente se deve ao reconhecimento de que o pré-teste seja uma boa avaliação, mas com suas limitações, como, por exemplo, o fato de ser muito curto ou de não exigir desempenho oral dos participantes. No grupo MG, nessa questão, um participante não respondeu esse item específico. Analisando a porcentagem num total de 12 respostas, portanto, temos que 33% concordaram totalmente, 58% concordaram parcialmente e 9% não concordaram nem discordaram. Já no grupo SP, 26% concordaram totalmente, 47% concordaram parcialmente e 26% não concordaram nem discordaram. Vemos que, no todo, o grupo MG se posicionou mais favorável a essa afirmação (total de 91% concordaram – total ou parcialmente – contra total de 73% no grupo SP).

Na **questão 9**, a maioria (aproximadamente 90%) dos participantes concordou totalmente com o fato de que as instruções do pré-teste foram claras. Com exceção de algumas instruções logísticas fornecidas pelo aplicador do exame, como a duração do teste, todas as outras foram fornecidas pelo próprio áudio de introdução do teste, e não houve problemas nesse quesito. Tal resultado aponta para uma flexibilidade do pré-teste para utilização autônoma de professores (em formação) de língua inglesa.

O tempo disponibilizado de 30 minutos foi informado aos participantes pelo aplicador no momento da aplicação do pré-teste. Com a **questão 10**, gostaríamos de saber se esse é um tempo razoável para a realização das questões. Nos surpreendeu que embora, a princípio,

poderíamos imaginar que o grupo do celular precisaria de mais tempo, uma vez que parece ser mais laborioso digitar no aparelho, 79% do grupo SP concordou totalmente que o tempo era suficiente e 21% concordou parcialmente. Assim, há apenas uma pequena parcela desse grupo que chegou a considerar uma maior quantidade de tempo para a realização do teste. Em contrapartida, o grupo MG, que realizou o pré-teste no computador, ficou mais dividido: 54% concordaram totalmente que o tempo é suficiente, 31% concordaram parcialmente, 7,5% discordaram parcialmente e 7,5% discordaram totalmente. Durante a aplicação para esse grupo, uma das participantes de fato comentou em voz alta sobre o tempo curto, dizendo que geralmente precisava de um período maior para realizar uma avaliação. Acreditamos que isso não desfaz o resultado que esperávamos. Gostaríamos que o pré-teste fosse curto, e, por isso, a depender desse resultado, poderíamos diminuir seu número de questões para que conseguíssemos manter sua brevidade. Observando os resultados, vemos que a maioria dos alunos concorda, mesmo que parcialmente (quase 94%), o que nos leva a crer que participantes que precisam de mais tempo são a exceção, e não a regra. Para eles, todavia, o pré-teste pode ser feito em um tempo maior. Por ser um teste de baixa escala e realizado sob responsabilidade do próprio candidato, a instrução de tempo é uma sugestão, e não uma restrição, podendo ou não ser seguida.

A plataforma *Google Forms* já foi aqui descrita como bastante simples, limpa e intuitiva. Os participantes da pesquisa corroboram essa afirmação, uma vez que, na **questão 11**, 100% do grupo MG se posicionou em total concordância, e quase 74% do grupo SP fez o mesmo. O restante do grupo SP se dividiu entre a concordância parcial (21%) e a discordância parcial (5%). Acreditamos que o posicionamento desfavorável do grupo SP se deve a alguns problemas ocorridos durante a aplicação e que podem ter sido atribuídos à plataforma: o problema de desconexão com a internet, que resultava na atualização da página e perda das respostas digitadas até então, e a incompatibilidade do aparelho de uma participante com o *Youtube*, onde estão hospedados os áudios das questões. Os dois problemas estão apenas indiretamente relacionados à plataforma, sendo que uma migração não resolveria o problema da desconexão, por exemplo, a não ser que a plataforma apresentasse cada questão de uma vez e conseguisse “guardar” a resposta, relacionando-a a um usuário, o que exigiria também um *login* dos participantes. A migração do áudio, acreditamos, seria contraproducente, uma vez que o *Youtube* é uma das plataformas mais universalmente aceitas e compatíveis com variados dispositivos, além de sua qualidade de reprodução e velocidade para carregar os arquivos.

A **questão 12** é uma especificação da 8, já que chama a atenção para o foco do pré-teste e do EPPLE, que é avaliar a proficiência específica à profissão de professor. Para o grupo MG, 23% concordaram totalmente que é uma boa avaliação da proficiência do professor, enquanto que 62% concordaram parcialmente e 15% não concordaram nem discordaram. Já no grupo SP, 21% concordaram totalmente, 68% concordaram parcialmente e 11% não concordaram nem discordaram. A distribuição entre os dois grupos é semelhante, e é interessante compararmos os resultados aos da questão 8. Vemos que o grupo MG concorda (total ou parcialmente) com o pré-teste ser bom instrumento de avaliação de proficiência e de proficiência do professor na mesma medida: 11 participantes (85%). No grupo SP, vemos outra relação: mais participantes disseram acreditar que o pré-teste era um bom instrumento de avaliação da proficiência do professor do que da proficiência em geral – 17 participantes (89%) contra 14 (74%). A avaliação da proficiência do professor ainda é pouco comum em contextos brasileiros, e acreditamos que isso possa fazer com que os participantes da pesquisa tenham observado mais atentamente o tipo de questão que é incomum às provas de proficiência geral.

Por fim, a última afirmação diz respeito à qualidade do áudio. Conforme descrevemos na metodologia, gostaríamos que o áudio pudesse ter sido de um site de leitura de voz, ou gravado em estúdio específico para este propósito, para a diminuição de ruído externo. No entanto, não foi possível que fizéssemos tal gravação no momento de desenvolvimento do pré-teste por falta de acesso livre a um estúdio dessa natureza. Para essa **questão 13**, 77% dos participantes de MG concordaram totalmente que o áudio tinha qualidade que permitisse boa compreensão, 16% concordaram parcialmente e 7% discordaram totalmente. Observando o grupo SP, 69% deles concordaram totalmente, 26% concordaram parcialmente e 5% não concordaram nem discordaram. Aqui, além da qualidade do arquivo de áudio, precisamos levar em conta o equipamento que foi utilizado, pois é nele que o áudio foi de fato reproduzido, maximizando ou minimizando as possíveis dificuldades. No grupo MG, os estudantes fizeram o pré-teste em um laboratório de informática que não possuía fones, de modo que os fones utilizados foram trazidos pelos pesquisadores e participantes, e apresentaram uma série de problemas, como cabo muito curto ou flutuações de volume. Os participantes do grupo SP, por sua vez, fizeram o pré-teste em seu próprio aparelho celular, trazendo seus fones. Os pesquisadores levaram fones também, mas estes não foram necessários. Imaginamos que grande parte desses participantes levou o fone que já acompanha

o próprio celular no ato da compra, resultando, possivelmente, em uma reprodução melhor do que a do caso dos participantes do grupo MG.

Assim, de modo geral, a maioria dos participantes confirmou as afirmações feitas sobre o pré-teste como um instrumento válido para medir a proficiência do professor de língua inglesa. Há ressalvas e limitações no teste, como comentamos em cada item, mas a impressão sobre o pré-teste como um todo é positiva.

### 3.1.2 Resultados e validade de critério

Voltando-nos agora para os resultados das avaliações, e sua relação entre si, fizemos a divisão levando em consideração os subgrupos, dado que um deles, o MG<sub>s</sub>, foi o único a realizar o EPPLE<sub>2</sub>, cujos dados nos serviram para a análise da validade preditiva do pré-teste.

Em nossa metodologia, fizemos uma previsão do que seriam os resultados esperados entre o pré-teste e EPPLE oral, que resultou no quadro 5 (p. 42), relacionando a faixa 1 do pré-teste às faixas A e B do EPPLE oral, a faixa 2 do pré-teste às faixas C e D do EPPLE oral, a faixa 3 do pré-teste às faixas D e E do EPPLE oral, e a faixa 4 do pré-teste à faixa E do EPPLE oral. Essa relação foi baseada nos descritores das faixas de proficiência de cada exame, que guiam o processo de correção das avaliações. Analisando os aspectos textuais de cada escala, montamos o quadro 8 a seguir, com algumas relações entre os descritores, que guiaram nossas expectativas teóricas.

**Quadro 8 – Relação entre descritores das faixas de proficiência do pré-teste e do EPPLE oral<sup>21</sup>**

PT2		EPPLE oral	
Descritor	Faixa	Faixa	Descritor
“gama suficiente de linguagem”	1	A, B	“utiliza estruturas linguísticas (...) adequadas e específicas”
“grau relativamente alto de controle gramatical e não comete erros que causam mal-entendidos”	1	A, B	“atinge plenamente os objetivos de comunicação verbal”
“compreende bem um excerto de áudio”	1	A, B	“não apresenta dificuldades de compreensão”

<sup>21</sup> A transcrição completa das faixas do pré-teste e do EPPLE oral se encontra em anexo, respectivamente, nas páginas 137-138 e 139-141.

“conhece bem as regras linguísticas”	1	A	“conhecimento sobre regras de uso da língua”
“conhecimento suficiente da língua para sobrevivência”	2	C, D	“estruturas linguísticas simples e pouca variedade lexical”
“compreende um excerto de áudio (...), com algumas limitações”	2	C	“pode apresentar dificuldades de compreensão”
“conhece as regras da língua, mas apresenta dificuldade em explicá-las”	2	B	“conhecimento sobre regras de uso da língua, mas apresenta dificuldade em explicá-las”
“padrões básicos de sentenças”	3	C, D	“estruturas linguísticas simples e pouca variedade lexical”
“estruturas simples corretamente, mas ainda comete erros básicos”	3	E	“estruturas linguísticas simples e pouca variedade lexical, (...) erros estruturais”
“compreende um excerto de áudio com certa dificuldade”	3	D, E	“apresenta dificuldades de compreensão”
“conhece as regras de uso da língua parcialmente, mas não consegue explicá-las”	3	D	“conhecimento limitado sobre regras de uso da língua, (...) não consegue explicar as regras de modo claro”
“repertório básico de palavras e sintagmas simples”	4	E	“estruturas linguísticas simples e pouca variedade lexical”
“domínio limitado de algumas estruturas gramaticais”	4	E	“erros estruturais”
“dificuldade na compreensão de um excerto de áudio”	4	D, E	“apresenta dificuldades de compreensão”
“pouco ou nenhum conhecimento sobre regras da língua, não conseguindo explicá-las”	4	E	“pouco ou nenhum conhecimento sobre regras de uso da língua”

Com o quadro 8, pretendemos relacionar aspectos dos descritores das duas escalas de proficiência. Reiteramos que, em alguns casos, não há como estabelecer uma relação de correspondência entre um instrumento e outro. No caso do EPPLE oral, por exemplo, aspectos como a pronúncia inteligível são avaliados, enquanto que no pré-teste, por não haver coleta da produção oral do aluno, tal característica não pode ser avaliada ou considerada nos descritores das faixas de proficiência. Esse tipo de “desencontro” entre os dois instrumentos,

em que o pré-teste deixa de avaliar um aspecto que é avaliado no EPPLE oral, pode estar por trás de disparidades entre os resultados dos dois testes.

Assim, o texto das duas escalas não está completamente alinhado de modo a tornar simples e direta a sua correspondência. Aspectos como a variação, na escala do pré-teste, são tratados na escala do EPPLE oral em meio a outros aspectos, como fluência e precisão. Em alguns casos houve também a sobreposição de duas faixas do EPPLE que se adequariam à descrição da faixa do pré-teste.

A correspondência entre as duas escalas foi feita através de uma análise de relações majoritárias. Nenhuma faixa correspondeu perfeitamente a outra na escala do outro instrumento, mas houve aproximações suficientes para se determinar sua associação.

Na seção a seguir, vemos a associação real dos resultados para podermos contrastá-la às nossas expectativas.

### 3.1.3 Validade paralela

A validade paralela, como discutido anteriormente com base em Alderson, Clapham e Wall (1995), Bachman (1990), Ebel & Frisbie (1991) e Hughes (1989), prevê que o teste e a medida de critério sejam aplicados simultaneamente. Os participantes que realizaram o pré-teste e o EPPLE (excluindo-se SP<sub>1</sub>, que realizou apenas o pré-teste) totalizam vinte e quatro (24). Seus resultados são passíveis de análise de validade de critério, uma vez que temos sua medida através dos dois instrumentos. O intervalo de aplicação entre as avaliações para o grupo MG (13 participantes) foi de apenas algumas horas, enquanto que para o grupo SP<sub>2</sub> foi de dois dias para o teste escrito do EPPLE (não tratado nessa pesquisa) e de 7 a 9 dias para o teste oral. Construímos o quadro 9 a seguir com as relações entre os resultados obtidos.

**Quadro 9 – Resultados – Validade paralela – Pré-teste como referência**

REFERÊNCIA - PRÉ-TESTE	EPPLE <sub>1</sub> ORAL	
<b>1</b>          <b>7 PARTICIPANTES</b>	<b>A</b>	
	0	0%
	<b>B</b>	
	5	<b>71%</b>
	<b>C</b>	
	2	28%
	<b>D</b>	
	0	0%
	<b>E</b>	
	0	0%

<b>2</b>  15 PARTICIPANTES	<b>A</b>	
	0	0%
	<b>B</b>	
	9	<b>60%</b>
	<b>C</b>	
	6	40%
	<b>D</b>	
	0	0%
	<b>E</b>	
0	0%	
<b>3</b>  2 PARTICIPANTES	<b>A</b>	
	0	0%
	<b>B</b>	
	0	0%
	<b>C</b>	
	2	<b>100%</b>
	<b>D</b>	
	0	0%
	<b>E</b>	
0	0%	

Fonte: elaborado pela autora.

O quadro foi organizado de modo a relacionar os resultados do pré-teste (primeira coluna) com os do EPPLE oral aplicado logo em seguida. Dentro da coluna do EPPLE oral, a primeira subcoluna representa o número de alunos daquela faixa (dentro os da faixa de referência) e a segunda subcoluna contém a porcentagem que esse número representa dentro do total de alunos que estavam na faixa de referência. Nenhum participante teve a faixa 4 como resultado no pré-teste, por isso ela não consta no quadro. Assim, por exemplo, tomando por base a faixa mais alta do pré-teste, 1, vemos que 7 dos 24 participantes a atingiram. Passando à coluna do teste oral do EPPLE<sub>1</sub>, vemos que, destes 7, a maioria ficou na faixa B do EPPLE<sub>1</sub> oral (5 alunos, ou 71% dos 7 alunos de faixa 1 no pré-teste).

Na mesma linha, os resultados das outras faixas do pré-teste em que obtivemos resultados – 2 e 3 – mostram uma correlação maior com as faixas de proficiência do EPPLE<sub>1</sub> oral. A única ressalva seria que, no caso das faixas 1 e 2 no pré-teste, ambas parecem apontar para a mesma faixa B do EPPLE<sub>1</sub> oral. Tal coincidência pode vir a ser problemática, uma vez que as faixas 1 e 2 parecem estar abrangendo alunos de mesmo nível no EPPLE, que inicialmente não esperávamos encontrar em duas faixas distintas no pré-teste.

Comparando-se os resultados às nossas expectativas baseadas nos descritores das faixas de proficiência, temos que, para a faixa 1, houve um cumprimento parcial do que

esperávamos, na medida em que a faixa B do EPPLÉ oral apareceu como majoritária, como prevíamos, mas houve também dois estudantes na faixa C. A faixa C estava abaixo das nossas expectativas para os estudantes que se classificassem como 1 no pré-teste. Para a faixa 2, cuja expectativa era faixa C ou D no EPPLÉ oral, temos a maioria dos estudantes na faixa B. Por fim, os resultados da faixa 3 no pré-teste resultaram na maioria dos participantes classificados na faixa C do EPPLÉ oral. Observamos, com exceção dos participantes que se enquadraram na faixa 1, uma tendência de os resultados no EPPLÉ estarem mais elevados do que esperávamos na nossa análise teórica (quadro 5, página 46, em que relacionamos as faixas antes da análise). Isso pode decorrer de o construto do EPPLÉ oral ser mais amplo, incluindo produção e compreensão oral, enquanto que o do pré-teste envolve apenas compreensão oral. Esperaríamos, inicialmente, que os resultados do pré-teste levariam a resultados mais baixos no EPPLÉ oral pela comparação entre descritores, uma vez que o pré-teste abrange faixas de menor proficiência em sua escala. No entanto, os resultados acabaram sendo mais elevados, o que pode sugerir que o pré-teste possui, na verdade, itens de dificuldade um pouco mais elevada do que esperávamos, apesar de não conter a produção oral. Outra possibilidade é que essa “dificuldade elevada” seja gerada pela contaminação (“*muddying*”) do construto oral do pré-teste pelas respostas dadas pelos candidatos através da produção escrita, que pode ser uma habilidade que exija maior domínio da língua estrangeira, e, portanto, acabe por classificar os candidatos em faixas mais altas também no EPPLÉ oral.

Diante dessas inesperadas relações, estudamos também os mesmos resultados, agora tomando como referência o próprio EPPLÉ<sub>1</sub> oral. Dessa perspectiva, todos os candidatos que ficaram na faixa B do EPPLÉ<sub>1</sub> oral estão nas faixas 1 e 2 do pré-teste, sendo, respectivamente, 36% e 64% do total de faixas B no EPPLÉ<sub>1</sub> oral. Esperávamos que a faixa B fosse relacionada à faixa 1 do pré-teste, mas a maioria dos resultados aponta para a 2. Portanto, o pré-teste, da maneira que se encontra, parece indicar aos alunos das faixas 1 e 2 grandes chances de conseguirem a faixa B ou a faixa C no EPPLÉ oral. A relação da faixa B é maior com a faixa 2 do que com a 1, o que nos faz questionar nossas expectativas iniciais do quanto o teste oral do EPPLÉ pode ser considerado de fato mais “difícil” do que o pré-teste, no sentido de que menos pessoas conseguem alcançar as faixas mais altas. Aparentemente, embora o EPPLÉ oral seja um instrumento maior e mais completo que o pré-teste, os resultados entre os dois não estão tão distantes como esperávamos. Outros aspectos a serem levantados como hipóteses da aproximação entre o nível de dificuldade dos instrumentos são as diferenças no construto (produção escrita vs. produção oral) e no uso da tecnologia (EPPLÉ foi realizado

em papel, pré-teste em computador ou celular). O quadro 10, a seguir, apresenta uma organização de dados tomando como referência o EPPLÉ<sub>1</sub> oral e comparando-o aos resultados do pré-teste obtidos pelos participantes.

**Quadro 10 – Resultados – Validade Paralela – EPPLÉ<sub>1</sub> oral como referência**

REFERÊNCIA – EPPLÉ <sub>1</sub> ORAL	PRÉ-TESTE	
<b>B</b>  14 PARTICIPANTES		<b>1</b>
	5	36%
		<b>2</b>
	<b>9</b>	<b>64%</b>
		<b>3</b>
	0	0%
	<b>4</b>	
0	0%	
<b>C</b>  10 PARTICIPANTES		<b>1</b>
	2	20%
		<b>2</b>
	<b>6</b>	<b>60%</b>
		<b>3</b>
	2	20%
	<b>4</b>	
0	0%	

Fonte: elaborado pela autora.

Nossa amostra de resultados das faixas 3 e 4 do pré-teste ficou bastante limitada, dado o público de nossa aplicação, que possui, em sua maioria, alta proficiência. Apenas dois candidatos (que realizaram pré-teste e EPPLÉ<sub>1</sub>) se posicionaram na faixa 3, e ambos tiveram faixa C no EPPLÉ<sub>1</sub> oral. Tomando como referência a faixa C do EPPLÉ<sub>1</sub> oral, podemos ver que, na verdade, 60% deles ficaram na faixa 2 do pré-teste, 20% na faixa 3 e 20% na 1. Assim, embora a faixa 3 do pré-teste aparente indicar um resultado C no EPPLÉ<sub>1</sub> oral (observável no quadro 9), candidatos com notas 1 e, principalmente, 2 também conseguiram a mesma faixa C, embora em menor proporção.

Como podemos observar, todos os alunos que fizeram o EPPLÉ<sub>1</sub> oral acabaram por se posicionar entre as faixas B e C. Este não é um resultado inesperado, pelo contrário, já que lidamos com um público de alunos presumidamente considerado de alta proficiência, dado que alguns se encontravam no final de seus cursos de licenciatura, outros já haviam concluído a graduação em Letras, e outros ainda já tinham experiência como professores. Com relação

ao quadro, vemos que a maioria dos candidatos, que tirou B ou C no EPPLE<sub>1</sub> oral, obteve a faixa 2 na realização do pré-teste. No entanto, nenhum candidato que alcançou B no EPPLE<sub>1</sub> oral tirou menos que 2 no pré-teste, enquanto que os candidatos que obtiveram C no EPPLE<sub>1</sub> oral tiveram faixas variadas no pré-teste (incluindo 1, 2 e 3, sendo que as duas primeiras também estão relacionadas à faixa B). As faixas 1 e 2, portanto, não parecem conseguir distinguir bem entre B e C no EPPLE<sub>1</sub> oral, mas há uma hierarquia de proporção, já que o B foi atingido por maior porcentagem de participantes que conseguiram 1 do que 2 no pré-teste. Também confirma essa hierarquia o fato de que os participantes que obtiveram 3 no pré-teste não alcançaram a nota B no EPPLE<sub>1</sub> oral, sendo seu máximo a faixa C.

### 3.1.4 Validade preditiva

De maneira semelhante à seção anterior, apresentamos um quadro com a relação entre o pré-teste e o EPPLE<sub>2</sub>, dessa vez aplicados com pouco mais de um mês de intervalo, configurando um caso de validade preditiva (ALDERSON, CLAPHAM & WALL, 1995; BACHMAN, 1990; EBEL & FRISBIE, 1991; HUGHES, 1989). A amostra de alunos que realizaram o EPPLE<sub>2</sub> é bem menor, composta de apenas 7 candidatos que se voluntariaram a realizar a segunda versão do EPPLE (grupo MG<sub>s</sub>).

Neste segundo caso, por nossa amostra ser bem menor, os resultados parecem indicar menos variação. No entanto, por esse mesmo motivo de tamanho da amostra, devemos tomar bastante cuidado ao falar em “tendências” sem relacioná-los aos resultados obtidos na amostra maior. Abaixo se encontra o quadro 11 com as relações entre os resultados.

**Quadro 11 - Resultados – Validade Preditiva – Pré-teste como referência**

REFERÊNCIA - PRÉ-TESTE	EPPLE ORAL 2	
<b>1</b>          <b>2 PARTICIPANTES</b>	<b>A</b>	
	0	0%
	<b>B</b>	
	2	<b>100%</b>
	<b>C</b>	
	0	0%
	<b>D</b>	
	0	0%
	<b>E</b>	
	0	0%

<b>2</b>  5 PARTICIPANTES	<b>A</b>	
	0	0%
	<b>B</b>	
	2	40%
	<b>C</b>	
	3	<b>60%</b>
	<b>D</b>	
	0	0%
	<b>E</b>	
	0	0%

Fonte: elaborado pela autora.

Nenhum participante que fez o EPPLE<sub>2</sub> ficou nas faixas 3 e 4 no pré-teste. A faixa 1 no pré-teste continua sendo indicativo de faixa B no EPPLE<sub>2</sub> oral. A faixa 2 também continua se relacionando às faixas B e C no EPPLE<sub>2</sub> oral, mas, ao contrário do EPPLE<sub>1</sub> oral, aqui há maior concentração na faixa C. Baseado nas duas análises, podemos dizer que a faixa 2 prevê uma faixa B ou C no EPPLE oral, com equilíbrio semelhante. E, mais uma vez, isso contradiz nossas previsões de que a faixa 2 estaria relacionada às faixas C e D do EPPLE oral, o que nos leva novamente ao questionamento dos níveis de dificuldade dos dois instrumentos, ou, talvez, de terem sido feitas correções com diferentes graus de rigor, ou mesmo levantar hipóteses quanto à diferença da modalidade de produção linguística (escrita ou oral).

Mudando a perspectiva para partirmos dos resultados do EPPLE<sub>2</sub> oral, temos o quadro 12, a seguir.

**Quadro 12 - Resultados – Validade Preditiva – EPPLE<sub>2</sub> oral como referência**

REFERÊNCIA - EPPLE ORAL 2	PRÉ-TESTE	
<b>B</b>  4 PARTICIPANTES	<b>1</b>	
	2	<b>50%</b>
	<b>2</b>	
	2	<b>50%</b>
	<b>3</b>	
	0	0%
	<b>4</b>	
	0	0%

<b>C</b>  3 PARTICIPANTES	0	<b>1</b> 0%
	0	<b>2</b> 0%
	<b>3</b>	<b>100%</b>
	0	<b>3</b> 0%
	0	<b>4</b> 0%
	0	0%
	0	0%

Da mesma maneira que no caso da validade paralela, no EPPLE<sub>2</sub> oral todos os candidatos se posicionaram nas faixas B e C. Neste caso, a faixa B está bem semelhante à encontrada na comparação paralela: apenas quem obteve 1 ou 2 no pré-teste conseguiu a faixa B no EPPLE<sub>2</sub> oral. A faixa C, no entanto, se diferencia da encontrada na validade paralela porque a totalidade dos candidatos que se encontra nela obteve a faixa 2 no pré-teste. Isso significa dizer que todos os candidatos que obtiveram C no EPPLE<sub>2</sub> oral foram classificados como 2 no pré-teste, mas nem todos os participantes que obtiveram 2 no pré-teste foram alocados na faixa C do EPPLE<sub>2</sub> oral, dois (40% do total de candidatos 2) foram alocados na faixa B. Assim, parece-nos, novamente, que os resultados do pré-teste apontam para resultados no EPPLE oral um pouco mais altos do que esperávamos.

### 3.1.5 A associação entre os dois instrumentos

Conforme comentamos no capítulo de fundamentação teórica, existem diferentes metodologias de associação entre testes. A que foi utilizada neste trabalho é a previsão, sendo a relação feita entre os desempenhos de participantes em dois instrumentos. A força das relações nessa metodologia pode ser sensível a variações de contexto, grupo e momento. Nossos participantes são uma amostra de professores em formação e alguns já em exercício, em maioria jovens e pertencentes a um contexto semelhante (universidades públicas do sudeste do Brasil). Assim, a pesquisa seria muito engrandecida pela ampliação da amostra, com aplicações para grupos mais variados.

Essas aplicações também beneficiariam nossa análise pelo fato de nossos resultados tratarem majoritariamente de participantes de alta proficiência. A grande maioria deles obteve faixa 1 ou 2 no pré-teste, de modo que essas foram as faixas que mais pudemos observar. Já é possível dizer que a linha divisória entre elas não está muito clara, sendo necessários mais estudos e potenciais modificações nos instrumentos e faixas. Ao reanalisar as correções do pré-teste diante destes resultados, percebemos que alguns dos candidatos geraram dúvidas aos

examinadores, dúvidas estas que podem vir a ser diminuídas, por exemplo, com alterações em algumas instruções do pré-teste (maior número mínimo de palavras a serem escritas, por exemplo). Quanto a esse caso específico, retomamos as respostas de três participantes à questão 6 do pré-teste, duas avaliadas em faixa 1 (a mais alta), e uma avaliada em faixa 2. A pergunta era “*What does the speaker mean when he says ‘power means different things at different times’?*”. A resposta esperada, de acordo com o áudio, era de que, em determinados momentos históricos, diferentes poderes definiam a sociedade dominante (por exemplo, poder militar, poder econômico, poder científico)<sup>22</sup>. A seguir, as respostas dos participantes A2, A30 e A32:

A2: “*Each time has different needs, thus different types of power*”

A30: “*It means that the concept of power changes depending on the socio-historical context and the influence some people have over the others.*”

A32: “*The speaker is referring to the fact that each period of time in history has a great power that is more important for society because it is the focus of such period. There was a century when the greatest power was political, and then it was the power of science and technology, after that there was an emphasis to the economic power, and the last one he mentions is the cultural power, after all those aforementioned.*”

Nas respostas A30 e A32 é possível perceber alta proficiência, mas a resposta de A30, pela menor extensão, oferece menor amostra para análise. Assim, as correções foram baseadas na amostra obtida, mas seria bastante interessante aumentar o número mínimo de palavras para que a amostra de desempenho do candidato pudesse ser mais significativa. Já no caso de A2, a resposta não é somente mais curta, mas menos elaborada, de modo que os corretores questionaram se houve mesmo uma boa compreensão oral do áudio. Nesse caso, foi atribuída faixa 2, com base na resposta não demonstrar uma boa compreensão do áudio.

Ainda com relação a esses excertos, podemos comentar que a grafia dissonante “*emphasys*” não foi considerada para se posicionar o candidato em uma faixa, uma vez que o construto do pré-teste não analisa ortografia.

---

<sup>22</sup> No script do pré-teste, na página 134, é possível consultar a transcrição completa do áudio.

Expressamos preocupação anteriormente com casos de informalidade, dado que o celular é utilizado para comunicação em contextos predominantemente casuais. No entanto, não houve nenhum caso de informalidade que violasse o contexto avaliativo nas respostas dadas aos itens do pré-teste nas aplicações realizadas durante este trabalho.

Outra questão em que houve cuidado foram as palavras com erros de digitação, ou possivelmente corrigidas pelo auto-corretor. Os casos a seguir dos participantes A7 e A22, ainda na mesma questão 6, ilustram esse aspecto:

A7: *“Because in each time **be** power means something different, for example in the 16th and 17th century we have the power of science and in the 19th the economic power”*

A22: *“The speaker means that different things are important **att** different times and, because of that, different things can make you powerful in some occasions and powerless on **other**.”*

As palavras foram grifadas para melhor observação, na resposta original o texto está sem formatação. Observando as palavras em negrito, “be”, “att” e “other”, percebemos que seriam consideradas erros estruturais, a primeira possivelmente um erro de escolha lexical, a segunda um erro de ortografia e a terceira um erro de concordância. Nenhuma delas foi considerada na correção, o participante A7 foi posicionado na faixa 2 nessa questão, e o participante A22, na faixa 1. No caso do participante A7, não foi levantada a hipótese imediata de autocorretor, porque esse participante realizou o pré-teste em computador. Ainda assim, sua resposta não satisfaz os corretores em termos de compreensão oral, recebendo faixas 1 e 2 inicialmente, mas nota final de 2. Já no caso do participante A22, que realizou o pré-teste em celular, acreditamos ser muito mais provável que seja um problema de autocorretor, e o participante ficou com a faixa 1, uma vez que por meio de sua resposta é possível inferir que ele compreendeu o áudio. Além disso, o erro “att” é superficial e relativo apenas à ortografia, que não faz parte do construto do pré-teste. Sobre “other”, que é um erro estrutural, não é possível afirmarmos que foi um erro de digitação ou um erro que o participante de fato cometera em sua produção oral.

A seguir, trazemos o quadro 13 com base nas análises feitas até então, considerando a maioria (mais de 60%) dos participantes alocados nas faixas. Assim, se 60% ou mais

candidatos da faixa 1 do pré-teste estão na B do EPPLE oral, esse é o valor que foi para o quadro. Desconsideramos os valores menores para analisarmos as relações principais entre as faixas de proficiência do pré-teste e as faixas de proficiência do EPPLE.

**Quadro 13 – Relações globais de resultados do pré-teste e do EPPLE oral com base nas análises**

Faixa pré-teste	Faixa EPPLE <sub>1</sub> oral	Faixa EPPLE <sub>2</sub> oral	Relação
1	B	B	1 → B
2	B	C	2 → B ou C
3	C	-	3 → C
4	-	-	-

Acreditamos que, a partir desse quadro, podemos dar um passo adiante nas previsões, baseadas também nos descritores que analisamos e na própria hierarquia interna de cada escala. Não seria difícil propor, dessa maneira, um novo quadro de previsões, que se pareceria com o quadro 14 a seguir:

**Quadro 14 – Previsões do pré-teste ao EPPLE oral com base nos resultados analisados**

PT2		EPPLE oral
1		A
1	2	B
2	3	C
3	4	D
4		E

A proposta desse segundo quadro, no entanto, pode vir a ser problemática. De maneira geral, temos que os resultados do pré-teste previram faixas mais elevadas no EPPLE oral do que antecipamos na análise das escalas de proficiência de ambas as avaliações. Nossas expectativas teóricas estavam baseadas na comparação das faixas de proficiência e no próprio desenvolvimento dos instrumentos. O EPPLE foi criado para ser uma avaliação da proficiência de professores, o que significa dizer não só que estamos tratando de uma

proficiência específica, diferente da proficiência geral, mas que a inclui, como também estamos falando em padrões de excelência. O professor de línguas é um profissional da linguagem, e espera-se dele um alto nível de proficiência linguística para que possa executar bem suas funções. Assim, as faixas do EPPLE foram inicialmente pensadas a partir desse modelo “ideal” de professor, e do que se esperaria de um excelente profissional em termos de proficiência. O termo “ideal” aqui não é utilizado no sentido de ser algo inatingível, que habita o mundo das ideias, mas de ser um padrão de excelência. Já o pré-teste foi desenvolvido como um instrumento de pré-seleção, planejado com potencial avaliativo de faixas mais baixas de proficiência. Além disso, por ser um teste curto, e que precisa abranger uma certa amplitude de níveis de proficiência, o pré-teste não é capaz de diferenciar entre níveis de proficiência mais elevados. Sua escala foi desenvolvida, dentre outros métodos, por meio de uma consulta à escala do QCER, em especial às faixas de A1 a B2 (as quatro mais baixas, que vão de “iniciante” até “intermediário superior”). Assim, a expectativa, antes de iniciarmos até mesmo a análise dos descritores, era de que resultados na faixa mais alta do pré-teste se relacionariam, por exemplo, a resultados na primeira e segunda faixa completas do EPPLE oral, e os resultados da faixa 2 do pré-teste equivaleriam, no máximo, à terceira faixa do EPPLE oral.

Conforme observamos na análise dos dados, a realidade foi de que os resultados no EPPLE oral, se comparados aos do pré-teste, apontaram para o fato de que o pré-teste não foi um teste tão “fácil” como prevíamos. O exemplo citado logo acima, de que o resultado 2 do pré-teste equivaleria, no máximo, à C no EPPLE, não se concretizou, sendo que quem obteve 2 no pré-teste conseguiu, na maioria dos casos, faixa B no EPPLE oral.

Comparando os dois quadros (o realizado a partir da análise teórica e o realizado a partir da análise de dados), temos o quadro 15 a seguir, que os sintetiza:

**Quadro 15 – Síntese entre previsões baseadas na análise de dados e na análise teórica das faixas de proficiência**

PT2 (teoria)		PT2 (dados)		EPPLE oral
1		1		A
1		1	2	B
2		2	3	C
2	3	3	4	D
3	4	4		E

Fonte: elaborado pela autora.

A segunda coluna, que diz respeito aos dados, abrange todos os resultados obtidos na pesquisa, com exceção de dois resultados 1 que ficaram na faixa C do EPPLE oral (faixa menor que a esperada, uma vez que 1 indicaria A ou B). Esses dois resultados, dos participantes A26 e A27, foram reanalisados em busca de razões para a discrepância.

O participante A26 fez um bom pré-teste que o posicionou na faixa 1, mas podemos observar uma série de erros de ortografia e alguns estruturais (*“ascnption”*, *“is happening”* ao invés de *“has been happening”*, *“trafic”*, *“strucked”* para *“stuck”* se referindo ao trânsito, dentre outros). Os erros *“is happening”* e *“strucked”* dão a entender problemas na proficiência geral do candidato. Os corretores, nesse caso, haviam discordado, um posicionando o candidato na faixa 2 do pré-teste (que seria condizente com o resultado do EPPLE oral) e o outro na faixa 1. Ao negociarem suas visões, em uma visão geral, acreditaram que o candidato se saiu muito bem, posicionando-o na faixa 1.

O mesmo participante, ao realizar o EPPLE oral, apresentou uma quantidade significativa de problemas estruturais em sua fala. Apesar de atingir B em alguns descritores, o candidato acabou por ser posicionado na faixa C por não conseguir se enquadrar na B de maneira completa.

O participante A27, em seu pré-teste, cometeu também alguns erros de ortografia (*“wheater”*, *“limiteded”* e alguns estruturais, especialmente na utilização de preposições, como *“because Ø the structure”*, *“due Ø the values”* e *“this”* no lugar de *“these”*). Aqui, mais uma vez, os corretores discordaram, um deles atribuindo faixa 1 e o outro, faixa 2. Mais uma vez, em negociação, decidiram por posicionar o candidato na faixa 1.

Durante o EPPLÉ oral, o participante A27 cometeu alguns erros de pronúncia, alguns estruturais, mas possuía fala fluente e vocabulário relativamente amplo, embora tenha se utilizado de expressões idiomáticas em língua portuguesa decodificadas, palavra por palavra, para o inglês (“*reach a half term*”, para “encontrar um meio-termo”). O que determinou de fato a nota baixa deste participante foi sua resposta à questão metalinguística. O candidato não soube respondê-la, mesmo com a ajuda da participante com quem realizava o EPPLÉ oral. Assim, os examinadores tiveram dúvida entre posicioná-lo nas faixas C e D, e o terceiro corretor optou por faixa C.

Voltamo-nos agora para a primeira coluna do quadro, que expressa os resultados esperados com base na análise teórica dos descritores das faixas de proficiência, temos quinze resultados discrepantes no total. Dois deles são os mesmos que foram discrepantes para a primeira coluna, os dois casos  $1 \rightarrow C$  discutidos nos últimos parágrafos. Os outros casos são de onze participantes  $2 \rightarrow B$  e de dois participantes  $3 \rightarrow C$ . Esses treze casos são de resultados no EPPLÉ oral superiores à previsão do pré-teste, e não inferiores, como foi o caso dos participantes  $1 \rightarrow C$ .

Do ponto de vista do candidato, que se inscreveria para fazer o exame a partir da previsão do pré-teste, seria definitivamente um problema obter uma faixa do pré-teste que indicasse uma determinada faixa do EPPLÉ oral, e, após fazer o exame, obter uma faixa inferior àquela prevista (casos  $1 \rightarrow C$ ). No entanto, se, ao fazer o EPPLÉ oral, o candidato acaba por se posicionar em uma faixa superior à prevista (casos  $2 \rightarrow B$  e  $3 \rightarrow C$ ), o impacto desse fato não é diretamente prejudicial ao candidato e sua decisão de fazer o exame.

Observando os resultados de nossas aplicações, acreditamos que talvez seja mais sensato, por ora, manter a tabela de previsões inicial (segunda coluna do quadro 15, ou quadro 5, p. 42), para que se minimize casos em que a nota real do EPPLÉ oral seja inferior à previsão feita por meio do pré-teste. Apesar de nossos resultados apontarem para um número maior de discrepâncias ao utilizar esse quadro, consideramos que seria um quadro mais seguro quanto às situações mais inconvenientes aos candidatos (EPPLÉ oral com nota inferior à prevista pelo pré-teste). Além disso, uma amostra maior de aplicações do pré-teste e do EPPLÉ oral seria bastante útil para alinharmos melhor os resultados das duas avaliações com relação, especialmente, às faixas mais baixas, nas quais não houve expressiva quantidade de participantes posicionados.

## 3.2 USO DA TECNOLOGIA

Nesta seção, analisamos as informações coletadas pela observação das aplicações do pré-teste, incluindo interações com os participantes durante essas aplicações (quando houve algum problema ou dúvida) ou logo após. Além disso, trazemos os dados coletados através de questionários respondidos pelos participantes sobre a experiência de realizar o pré-teste. Relembramos que o grupo SP realizou o pré-teste no celular, totalizando a participação de 19 integrantes, e o grupo MG, composto de 13 participantes, realizou o pré-teste no computador. As informações discutidas aqui são as referentes ao aspecto tecnológico das aplicações.

### 3.2.1 Aplicações no computador

#### *Observações de campo*

Faremos nessa seção uma breve descrição de pontos de interesse durante as aplicações dos testes, baseados nas anotações feitas naquela ocasião.

A aplicação do pré-teste ao grupo MG se deu através de computadores em um laboratório de uma universidade pública local, no dia 20/10/2017. O laboratório reservado, assim como todos os outros disponíveis na universidade, não contava com fones de ouvido, o que requereu adaptações da parte dos aplicadores e participantes. O ideal seria que o laboratório fornecesse fones de qualidade e com isolamento do som externo, do tipo *headphones*. Diante da impossibilidade, os aplicadores trouxeram consigo fones de ouvido pequenos, do tipo intra-auricular, os mesmos que comumente acompanham aparelhos celulares, que não contam com as mesmas especificações de qualidade. Os participantes também foram convidados a levar seus próprios fones, o que resultou em uma potencial despadronização de volume e qualidade de áudio.

Outro problema com a utilização de fones não pertencentes ao laboratório foi o próprio comprimento do fio, que em alguns casos não foi suficiente para que a conexão entre o computador e o ouvido do aluno fosse confortável. Tais fones foram substituídos imediatamente por outros que proporcionavam conforto ao candidato.

## Questionários

Analisando os questionários aplicados ao grupo MG, que fez o pré-teste no computador, encontramos também muitos comentários relacionados à tecnologia. As respostas aos questionários foram agrupadas sob categorias e organizamos as críticas e sugestões no quadro 16, a seguir, com comentários da pesquisadora na coluna à direita.

**Quadro 16 - Sugestões e críticas ao pré-teste no computador**

SUGESTÃO/ CRÍTICA	RELEVÂNCIA (%)	COMENTÁRIO
Sala com fones de ouvido padrão.	31%	Procuramos por uma sala nesse molde na universidade em questão e não encontramos. Havia apenas laboratórios com computadores, o que é lamentável não apenas para a aplicação do pré-teste, mas para os próprios alunos da universidade, não só – mas especialmente – os de Letras, que poderiam ter aulas diferenciadas e desenvolver projetos com a utilização do laboratório.
Áudio muito baixo.	15%	Novamente, o fato de termos tido que improvisar os fones, trazendo-os de casa (pesquisadores e participantes) para adaptá-los aos computadores do laboratório, gerou alguns problemas, dentre eles áudios muito baixos ou comprimento do fio insuficiente para o candidato fazer a prova confortavelmente.
Mais perguntas, para melhor avaliar a proficiência.	15%	O propósito do pré-teste é ser uma amostra pequena da proficiência do candidato, um teste prático e curto que o indica a fazer o EPPLÉ, que, aí sim, é um exame maior e que coleta muito mais produção do candidato. Talvez a partir de novas aplicações, em futuros estudos, isso poderia ser revisado e o número de questões aumentado, se esse aspecto estivesse prejudicando a avaliação da proficiência.
Plataforma não permite contabilizar quantas vezes o aluno repetiu o áudio.	15%	Os candidatos são instruídos a escutar o áudio duas vezes, mas realmente não há impedimento na plataforma para que escutem mais. Porém, a própria duração do teste funciona como impedimento, uma vez que o candidato precisa de tempo para escutar, elaborar as questões e revisá-las.
Presença de cronômetro no próprio teste.	8%	A plataforma <i>Google Forms</i> não disponibiliza a opção “relógio” ou “cronômetro”, por isso nós examinadores fizemos o aviso de duração e o de 5 minutos restantes. Inclusive tentamos utilizar o add-on <i>Timify.me</i> , que acrescenta apenas uma tela inicial antes do <i>Forms</i> informando a duração. A coleta do horário que o candidato envia o formulário já é realizada automaticamente pelo <i>Forms</i> sem o add-on. Há também um relógio em cada celular e

		computador para que o candidato se oriente, e é uma boa ideia o examinador chamar a atenção para isso no início da aplicação.
Disponibilizar mais detalhes sobre a prova anteriormente para alunos poderem se preparar melhor.	8%	Um dos objetivos do grupo de pesquisa que trabalha com o EPPLÉ é criar e disponibilizar esse tipo de material, mas ainda não foram publicados trabalhos nesse tema.
Tempo curto.	8%	O tempo foi determinado com base no número de questões, na duração dos áudios e no tamanho das respostas esperadas. O pré-teste pretende ser curto, por isso mesmo sua extensão já é reduzida.
Vídeo original no texto oral.	8%	Não quisemos adicionar imagens aos áudios, apenas uma das questões, a do texto, vinha realmente de um vídeo e apresentaria essa possibilidade. Preferimos manter uma padronização, mas podemos considerar mais essa opção.

Fonte: elaborado pela autora.

Ao serem perguntados, alguns participantes responderam que não possuíam críticas ou sugestões ao pré-teste (sem críticas 69%, e sem sugestões 46%). Alguns participantes (54%) utilizaram o espaço para deixarem também elogios ao pré-teste, dizendo que os áudios eram claros e de fácil compreensão, o formato é organizado e simples, as orientações são claras, e o teste em si, rápido e prático. Trazemos abaixo tais excertos.

A12: *“Achei o pré-teste organizado e simples (em relação ao formato e à plataforma)”*

A11: *“Os áudios foram claros e de fácil compreensão”*

A6: *“Parabenizo pela organização, profissionalismo e competência que permitiram o andamento do pré-teste”*

A5: *“Não encontrei problemas ao realizar o pré-teste. Achei tudo muito claro e objetivo”*

A4: *“Achei as orientações bem claras. Não encontrei nenhuma dificuldade ao realizar o pré-teste, além de ser rápido e prático de responder”*

A1: “*Não encontrei nenhum problema ao realizar o teste. As instruções foram claras e simples. A plataforma é uma das melhores na minha opinião para trabalhar digitalmente*”

Entretanto, reconhecemos que ainda há um longo caminho a ser percorrido para melhorarmos o pré-teste ao EPPLÉ devido às limitações aqui discutidas.

### **3.2.2 Aplicações no celular**

#### *Observações de campo*

As duas aplicações aconteceram em datas distintas (31/10/2017 e 28/11/2017), mas suas observações foram agrupadas uma vez que estamos tratando do aspecto comum entre elas, que foi a tecnologia. Ambas aplicações (aos grupos SP<sub>1</sub> e SP<sub>2</sub>) foram feitas através de aparelhos celulares.

As dificuldades relatadas pelos participantes à própria pesquisadora, durante ou logo após a aplicação, foram duas: o problema da autocorreção do celular e a falta de um local para anotações quando escutando o texto oral. Uma das alunas não teve como utilizar o próprio celular e, ao pegar emprestado o da colega que havia terminado, relatou que o celular desta não possuía dicionário bilíngue, então autocorrigia todas as palavras para uma semelhante em língua portuguesa, o que tornava o exercício de responder às questões bem mais laborioso. Outros alunos relataram o problema do autocorretor também nos questionários. O segundo apontamento foi feito por uma aluna que pediu uma folha para fazer anotações do texto oral. Este texto tem duração de dois minutos e trinta e três segundos, descontando-se o tempo da introdução da questão, e também das questões subsequentes, que estão em áudios separados. Os itens subsequentes dizem respeito a aspectos gerais de interpretação e uso. Ao elaborarmos a prova, acreditamos que não seria necessário fazer anotações, e esta foi a única participante que sentiu necessidade da folha de papel. Ainda assim, seria interessante disponibilizar essa folha opcional aos candidatos que preferirem fazer anotações.

#### *Questionários*

Ao serem questionados sobre as vantagens de realizar o pré-teste em dispositivo móvel, os participantes citaram frequentemente pontos como a liberdade (possibilidade de controlar o áudio – volume, quantas vezes ouvir, possibilidade de fazer as questões na ordem preferida), a facilidade, a rapidez, e a interatividade. Já com relação às desvantagens, com

exceção de um aluno, que não observou desvantagens, todos os restantes mencionaram dificuldades tecnológicas (flutuações na conexão com a internet, incompatibilidade do dispositivo com algum tipo de arquivo ou função, interrupção da atividade do teste por chegada de mensagens ou notificações de outros aplicativos), o que mostra a importância de se pesquisar como uma avaliação pode ser aplicada com mais eficiência através de DTMs.

Em questão posterior, os participantes opinam sobre sua preferência em realizar o pré-teste por meio do celular ou de um computador. Um participante preferiu o celular pela praticidade de realizar o teste em qualquer local com rede de internet. Dois participantes alegaram indiferença, desde que haja disponibilidade de boa conexão com a internet. Quatro participantes disseram preferir fazer pelo computador, por motivos diversos, como tamanho do teclado e da tela do DTM, presença de autocorretor, digitação mais lenta que a do computador, maior chance de o computador abrir os arquivos, possibilidade de habilitá-lo para o teste com antecedência, dentre outros. O participante restante disse preferir o computador se pudesse fazer a prova em casa, mas preferir o celular se tivesse que se deslocar para realizá-la.

A parte final do questionário abre espaço para críticas e sugestões dos candidatos, que são, em sua maioria, em relação à tecnologia utilizada. O quadro 17, a seguir, mostra as informações coletadas e um comentário da pesquisadora, no mesmo molde do presente na seção anterior.

**Quadro 17 – Sugestões e críticas ao pré-teste no celular**

SUGESTÃO/ CRÍTICA	RELEVÂNCIA (%)	COMENTÁRIO
Disponibilização de rede sem fio estável para a realização do teste (problemas de conexão com a internet geraram demora no carregamento e atualizações involuntárias da página)	50% (SP <sub>1</sub> ) 9% (SP <sub>2</sub> )	Trabalhamos para aplicar o pré-teste em ambientes em que haja essa disponibilidade. No entanto, dependemos da conexão das universidades, que pode vir a sofrer falhas e flutuações. Infelizmente, nas universidades, ainda há problemas de queda da conexão ou de locais dentro do <i>campus</i> onde a conexão não se estabelece de maneira estável.
Problemas com o autocorretor	50% (SP <sub>1</sub> ) 9% (SP <sub>2</sub> )	Esse tipo de problema não é de fato do pré-teste, mas do próprio aparelho que o candidato utilizou. Pensamos sobre maneiras de instruir os candidatos quanto a isso, mas existem muitos modelos de <i>smartphone</i> , e com sistemas operacionais distintos. Acreditamos que poderia

		ser futuramente disponibilizado, nas instruções do pré-teste, orientações para configurar os modelos mais comuns de celular e sistema operacional.
Aparelho lento para reprodução de vídeo	0% (SP <sub>1</sub> ) 27% (SP <sub>2</sub> )	Pode ser um problema do aparelho em si, ou da velocidade da internet. De qualquer forma, a aplicação seria beneficiada em um ambiente de conexão estável e com equipamentos padronizados e de boa qualidade. Além disso, quanto às sugestões de as questões serem apenas áudio, e não vídeo, é precisamente o que gostaríamos, mas a plataforma do <i>Google Forms</i> não permite inserção de áudios apenas, e nem o <i>Youtube</i> permite que carreguemos áudios sem imagens que os acompanham.
Aparelho incompatível	25% (SP <sub>1</sub> ) 0% (SP <sub>2</sub> )	A incompatibilidade total aconteceu com apenas dois participantes. As plataformas utilizadas ( <i>Google Forms</i> e <i>Youtube</i> ) foram escolhidas justamente pela ampla aceitação em diferentes dispositivos. Conforme o comentário acima, a aplicação seria melhor com equipamentos padronizados.
Inclusão de itens de produção oral	25% (SP <sub>1</sub> ) 0% (SP <sub>2</sub> )	O único motivo pelo qual não incluímos itens de produção oral foi a indisponibilidade de plataformas <i>online</i> que oferecessem esse serviço.
Inclusão de itens de leitura	25% (SP <sub>1</sub> ) 0% (SP <sub>2</sub> )	Por ser um teste de proficiência oral, não acreditamos que este seria uma das tarefas mais próximas das situações de uso que gostaríamos de avaliar.
Disponibilização do teste off-line (menos distrações e menos problemas com flutuações na conexão)	12% (SP <sub>1</sub> ) 9% (SP <sub>2</sub> )	Seria bastante complicado disponibilizar o teste off-line, pois precisaríamos de um programa, ou, como sugeriu uma aluna, poderíamos enviar os arquivos aos candidatos.
Resultados expressos em níveis (iniciante, intermediário, avançado), além das notas	12% (SP <sub>1</sub> ) 0% (SP <sub>2</sub> )	Acredito que o participante em questão inferiu que seria avaliado em notas, pois é a maneira mais tradicional, embora isso não tenha sido dito em momento algum, uma vez que os resultados do pré-teste são expressos em faixas de proficiência.
Instrução para escrever as respostas em LI	12% (SP <sub>1</sub> ) 0% (SP <sub>2</sub> )	Essa sugestão foi de grande valia, uma vez que deixamos passar essa informação. No decorrer do teste, que é todo em áudio e em LI, não mencionamos que as respostas devem ser escritas também nessa língua. Já realizamos a correção.
Inclusão de questões gramaticais e de múltipla escolha	12% (SP <sub>1</sub> ) 0% (SP <sub>2</sub> )	Já há questões gramaticais no teste, talvez o candidato tenha se referido a questões de formato mais estrutural, como as mencionadas de múltipla

		escolha. Gostaríamos que o teste fosse mais comunicativo e, para isso, buscamos coletar a produção dos alunos, e criar contextos e situações de uso autênticas. No entanto, não descartamos a possibilidade de inclusão de questões de múltipla escolha devido à sua provável maior adaptabilidade a DTMs.
Inserção de imagens explicativas e/ou frases escritas nos vídeos, como complemento ao áudio	12% (SP <sub>1</sub> ) 0% (SP <sub>2</sub> )	O propósito do teste de avaliar a compreensão oral seria comprometido, uma vez o candidato poderia responder às questões por meio de sua proficiência escrita, ou da assistência visual das imagens.
Melhor padronização do volume do áudio	12% (SP <sub>1</sub> ) 0% (SP <sub>2</sub> )	Gostaríamos muito de disponibilizar um áudio mais padronizado, como discutimos na seção de metodologia, mas, para isso, precisamos de uma plataforma de voz com boa entoação, ou a disponibilidade de aparelhos (microfones) de boa qualidade, em estúdio, para a gravação.
Inclusão de itens com textos menos formais	0% (SP <sub>1</sub> ) 9% (SP <sub>2</sub> )	Os primeiros três itens do teste são de situações informais, pois se tratam de questões pessoais. Os itens posteriores tem caráter mais específico da profissão de professor de LI e, por esse motivo, apresentam linguagem mais profissional e formal.
Preferência pelo computador	0% (SP <sub>1</sub> ) 9% (SP <sub>2</sub> )	Este participante sugeriu que o pré-teste seria melhor se feito no computador, devido a dificuldades no layout de seu celular, que não respondeu como esperado ao comando de repetir o áudio. Nessa pesquisa, analisamos justamente algumas características do impacto da utilização de um DTM em uma avaliação de línguas.
Problema com pergunta metalinguística	0% (SP <sub>1</sub> ) 9% (SP <sub>2</sub> )	Este participante achou problemática a inclusão de uma questão metalinguística: <i>“Tive problemas para explicar o porque a sentença do aluno estava errada. Sabia que estava errada, mas não sabia dizer o porquê”</i> . Nesse caso, esse é exatamente o ponto da questão, o conhecimento do professor das regras de uso da língua e de como explica-las.

Fonte: elaborado pela autora.

Alguns participantes mencionaram que problemas com a conexão não são necessariamente da prova em si, mas externos a ela. No entanto, ao se aplicar uma prova, é preciso levar em conta questões externas a ela, que são essenciais para a sua boa realização. Mesmo que o problema de conexão seja externo à prova, ele ainda deve ser levado em conta no seu desenvolvimento e aplicação para que os resultados sejam confiáveis e válidos. Um

problema mais sério do que esse é a utilização do celular do próprio participante, que pode não ser compatível com o formato da avaliação, ou não processar os vídeos, ou estar com quaisquer outros problemas. Em nosso caso, não temos como ter controle sobre os aparelhos pessoais dos alunos, e isso também foi um problema no momento da aplicação dos testes. Abaixo trazemos excertos dos questionários em que os candidatos criticam aspectos externos ao instrumento.

A21: *“Somente o problema do celular, mas ao teste em si não tenho críticas”*

A19: *“Sem críticas. Os problemas que tive durante o teste estão relacionados à internet, e não à plataforma”*

Quanto aos participantes que não tiveram críticas, 50% do grupo SP<sub>1</sub> e 45% do grupo SP<sub>2</sub> afirmaram não ter críticas ao pré-teste. Tratando-se das sugestões, 25% do grupo SP<sub>1</sub> não apresentou sugestões ao pré-teste e 63% dos participantes do grupo SP<sub>2</sub> também não.

Da mesma forma que nas aplicações com computadores, alguns candidatos teceram elogios (25% do grupo SP<sub>1</sub> e 9% do grupo SP<sub>2</sub>), dizendo que gostaram das questões que levam em conta a atuação do professor e que acharam o áudio e as instruções claras. Abaixo mais excertos dos questionários.

A16: *“Gostei do exercício perguntando sobre como explicaríamos algo para um aluno. Esse tema raramente é tratado na formação de professores”*

A26: *“achei o teste adequado”*

A24: *“Achei que o teste é bastante coerente”*

A15: *“Não, o áudio é muito bom e as explicações de como realizá-lo são claras. Gostei do teste!”*

### **3.2.3 Comparação entre PT2 no computador e no celular**

O questionário aplicado ao grupo SP<sub>1</sub>, composto de oito participantes, continha além das questões presentes nos outros questionários, três itens a mais sobre o uso do aparelho celular na aplicação. Nos questionários desse grupo, os itens 14, 15 e 16 eram, respectivamente:

14. Quais foram as vantagens de realizar o teste através de um dispositivo móvel (celular)?
15. Quais foram as desvantagens de realizar o teste através de um dispositivo móvel (celular)?
16. Se pudesse escolher, você faria esse teste pelo computador ou pelo celular? Por quê?

Sobre as vantagens, os candidatos mencionaram a possibilidade de controlar o áudio de acordo com o seu tempo, inclusive fazendo pausas (50%), a liberdade para voltar ou avançar questões (37,5%), a familiaridade com o aparelho/facilidade de manuseio (37,5%), a mobilidade/praticidade (37,5%), a rapidez (25%), a facilidade para organizar as ideias (12,5%), a clareza e volume do áudio (12,5%), a interatividade (12,5%), os recursos multimídia (12,5%), e o fato de a avaliação “mais palpável” e “menos cansativa” (12,5%). Podemos notar aqui que a maioria das vantagens que poderiam ser também observadas ao se realizar o teste pelo computador (como, por exemplo, a possibilidade de controle do áudio e das questões). A única vantagem que identificamos que seria realmente específica dos DTMs seria a mobilidade.

Já quanto às desvantagens, os participantes citaram a conexão flutuante e/ou lenta com a internet (62,5%), a (possibilidade de) incompatibilidade do aparelho com os arquivos do teste (25%) e as notificações de outros aplicativos durante o teste (12,5%). As desvantagens são, também, parecidas com as que poderiam ser encontradas em um computador pessoal, com exceção, talvez, das notificações, que são menos comuns em computadores. No entanto, a grande diferença da realização do teste em computadores é, no momento, a possibilidade de preparar as máquinas com antecedência para evitar os problemas já de antemão. Para o celular pessoal, tal preparação poderia ser feita no momento do teste, já com a presença dos participantes, e com a ajuda de um examinador que tivesse conhecimentos técnicos para configurar diferentes modelos e sistemas operacionais.

Por fim, quando questionados se preferiam realizar o pré-teste através do celular ou do computador, 4 participantes (50%) afirmaram que o computador seria mais confortável. Dentre os motivos apontados por esses candidatos estão a presença de autocorretor do celular (1 menção), a digitação mais rápida no computador (1 menção), teclado e tela do celular são pequenos (1 menção), maior chance de compatibilidade com arquivos (2 menções) e conexão mais estável (1 menção).

As respostas dos 4 participantes restantes foram distintas, e não as agrupamos. Um participante afirmou que, se pudesse fazer o teste em casa, preferiria o computador, mas, se fosse para se deslocar, preferiria pelo celular. Um participante disse que, se a conexão com a internet fosse estável, qualquer dos dispositivos lhe seriam indiferentes. Outro participante disse que ambos, computador e celular, pois os dois possibilitam dinamismo e praticidade na execução. E o último participante disse que preferiria fazer pelo celular, pois poderia acessá-lo de qualquer lugar com conexão à internet.

No quadro 18 a seguir, fazemos também uma síntese dos principais problemas discutidos nas subseções anteriores.

**Quadro 18 – Síntese dos principais problemas nas aplicações via computador e via celular**

Potenciais problemas	Computador (MG, 13)	Celular (SP, 19)
Conexão com a internet	0%	26%
Problemas no áudio	15%	5%
Incompatibilidade ou equipamento	31%	11%
Autocorretor	0%	26%
Tempo curto	8%	0%

Fonte: elaborado pela autora.

Nos dados do quadro 18, é possível perceber que os participantes apontaram problemas distintos em cada equipamento. Para os candidatos que realizaram o pré-teste através do computador, o fato de os fones não serem padronizados foi mais problemático do que o fato de os *smartphones* não serem no outro grupo. Isso está relacionado, certamente, com os problemas no áudio. Além disso, os participantes acharam o tempo curto para a realização do teste, o que não foi uma reclamação entre os candidatos que o realizaram por meio do celular. Por outro lado, a conexão com a internet foi bem mais desafiadora para os participantes do grupo SP, que dependiam da disponibilidade sem fio da rede, o que é comumente mais problemático do que laboratórios de computação, que possuem cabos de conexão com a internet. O outro maior problema desse grupo foi o autocorretor do celular, que causou alguns inconvenientes aos participantes durante a realização do teste.

Como pudemos observar por meio de nossa análise, as aplicações através do celular foram proveitosas, porém ainda com aspectos a serem mais investigados e melhorados (especialmente a conexão com a internet e a presença do autocorretor). As aplicações pelo computador não foram, no entanto, livres de problemas (equipamento e áudio). Podemos afirmar que quase todos os problemas dos dois grupos podem ser resolvidos através de tecnologia de melhor qualidade: melhores equipamentos para a gravação do áudio de elaboração do teste; melhores laboratórios de computação, padronizados e com fones; melhor disponibilidade e qualidade de internet. O único problema que pode ser resolvido por meio de apenas uma melhor preparação dos examinadores é a disponibilização de instruções técnicas para habilitação do autocorretor em língua inglesa antes do teste.

Assim, ambos os equipamentos utilizados no pré-teste parecem configurar instrumentos legítimos para sua aplicação. Os problemas mais apontados pelos participantes podem ser resolvidos, principalmente, com mais investimento em tecnologia nas universidades.

### **3.3 DISCUSSÃO E ENCAMINHAMENTOS**

#### **3.3.1 Sobre a pesquisa**

Dentre as limitações encontradas no desenvolvimento desta pesquisa, vamos abordar à parte as que dizem respeito ao pré-teste em si, seu desenvolvimento e alocação. Nessa primeira seção, descreveremos desafios encontrados em nossa metodologia e durante nosso estudo.

Inicialmente, nossa amostra de participantes possui majoritariamente um nível de proficiência intermediária ou alta em língua inglesa, o que resultou em pouca variedade quando da comparação dos resultados e associação entre as faixas. Tal problema poderia ser resolvido com mais aplicações, com participantes de proficiência mais baixa. Infelizmente, durante este estudo, não houve tempo hábil para ampliarmos nossa coleta de dados. Além disso, os dados de faixas mais baixas também seriam úteis na medida em que enriqueceriam as informações sobre *feedback* aos candidatos.

Dentre as limitações com o material de pesquisa, citamos o improvisado do aparato eletrônico utilizado nas aplicações. No caso do grupo MG, não havia laboratório de informática com fones na universidade em questão, o que é surpreendente e entristecedor não só por nossa pesquisa, mas por serem aparelhos relativamente baratos e que dão acesso aos

estudantes a informações de multimídia, ampliando muito suas possibilidades de estudo. Tais limitações refletem os desafios pelos quais passam grande parte dos estudantes e profissionais do ensino superior no Brasil. Em nossa pesquisa, isso resultou em uma interferência da variável “fone” na aplicação do grupo MG, que não foi prevista quando da proposta desse projeto de pesquisa.

Além disso, outra limitação foi a utilização apenas de aparelhos celulares como representantes da categoria “DTMs”. Teria sido interessante a realização do pré-teste também em *tablets*, já que tais dispositivos possuem a tela um pouco maior e são bem mais amplamente utilizados para ensino e aprendizagem do que aparelhos celulares, embora ainda não sejam tão populares na vida cotidiana de brasileiros quanto o *smartphone*.

Prosseguindo com limitações durante as aplicações, o tempo levado para preparar a sala dos computadores para a aplicação do pré-teste foi maior do que esperávamos. Antes do início da aplicação, para abrir o pré-teste em todos os computadores, a pesquisadora teve que acessar a conta *Google*® do pré-teste em cada máquina para abrir o link do formulário e, logo depois, deslogar e logar na próxima. Após a experiência, reconsideramos uma ideia pensada anteriormente, que foi a de pedir que os alunos mesmos entrem em seus e-mails (coletados anteriormente pelo examinador no momento de sua inscrição) e que acessem o link enviado através deles. Essa opção foi inicialmente descartada pois permite que os candidatos tenham consigo o link do pré-teste nesse momento em que ainda está em pesquisa, e possam utilizá-lo e compartilhá-lo livremente. Entretanto, é possível “fechar” o formulário temporariamente, impedindo que aceite mais respostas ou acessos, e apenas reabri-lo para as aplicações subsequentes. De qualquer forma, acreditamos agora que deixar o link disponível aos participantes é um problema menor se comparado à praticidade que acrescenta ao processo de administração do teste, embora ainda seja preciso abrir o pré-teste pelo link para os computadores extras.

Ainda se tratando da administração do pré-teste, uma participante das aplicações através de DTMs indicou a falta da instrução aos candidatos de desabilitarem outros aplicativos em seus aparelhos celulares. Tal candidata ficou incomodada, e com toda razão, com o fato de que, durante o teste, recebia mensagens instantâneas em sua tela que provinham de outros aplicativos. Porém, instruir os candidatos a desabilitarem aplicativos pode ser desafiador, já que eles podem não saber como realizá-lo em seus próprios aparelhos. Para isso, precisaríamos de examinadores treinados para ajudar os participantes a configurar modelos diferentes de *smartphone*, muitas vezes com sistemas operacionais distintos, ou

mesmo outras versões do mesmo sistema. Não seria uma instrução simples, como desconectar o aparelho da internet. Tal instrução resolveria as notificações de aplicativos *online*, mas também desabilitaria o candidato de enviar suas respostas do teste, que está hospedado também na internet. Além disso, a simples desconexão com a internet não garantiria que aplicativos ainda assim enviassem notificações ou lembretes *off-line*, ou impediria o soar de despertadores previamente estabelecidos pelos candidatos.

Dentre as limitações encontradas na aplicação do EPPLÉ oral, tivemos apenas uma. O teste oral inclui um pequeno vídeo, hospedado no *Youtube*. Na primeira data das aplicações ao grupo MG, houve uma falha na conexão com a internet durante o período da tarde, que estavam sendo aplicados os testes orais do EPPLÉ. A solução encontrada nesse momento foi utilizar o *smartphone* dos pesquisadores conectado à caixa de som do computador onde o vídeo seria mostrado. Para as aplicações posteriores, no entanto, o vídeo foi baixado para se tornar disponível *off-line* e que não haver dependência do bom funcionamento da internet. Acreditamos, inclusive, que as possibilidades de não utilizar a internet devem ser sempre exploradas, ao menos como plano alternativo no caso de falha de conexão. Na aplicação dos instrumentos dessa pesquisa, o vídeo do EPPLÉ oral é a única situação em que podemos prevenir problemas na administração em caso de falha de conexão, uma vez que o pré-teste é hospedado totalmente *online*, sem possibilidades de ser realizado sem internet.

Por fim, retomamos alguns pontos que são essenciais na aplicação de qualquer teste eletrônico. Primeiramente, deve haver, sempre que possível, o teste dos materiais a serem utilizados. No caso dos celulares e fones dos próprios alunos, não tivemos acesso anterior a eles, o que gerou alguns problemas durante a aplicação. Além disso, e como consequência de problemas potenciais, é preciso sempre ter materiais extras: trouxemos muitos fones extras para as aplicações e deixamos também computadores extras já ligados na tela do pré-teste. Assim, é muito mais fácil e rápido lidar com os obstáculos que se apresentam no momento da administração do teste. Uma última consideração seria quanto à conexão com a internet. É sempre proveitoso pensar em alternativas quanto a falhas de conexão, quando o instrumento permite essas alternativas.

### **3.3.2 Sobre o pré-teste (PT2)**

Ao final de nossa análise, concluímos que alguns fatores limitaram a utilização do pré-teste como uma avaliação válida e disponível em computadores e DTMs. Faremos aqui uma breve descrição dessas limitações internas ao instrumento, apresentando também

encaminhamentos e ideias que as abordem, resultando numa potencial terceira versão do pré-teste (PT3).

Em primeiro lugar, um dos aspectos mais problemáticos é a sub-representação do construto do pré-teste. O EPPL oral é um teste de proficiência oral completo, incluindo produção e compreensão, e o pré-teste se propõe a avaliar o mesmo construto. Infelizmente, nenhuma das plataformas gratuitas encontradas possui a função de coleta da produção oral do participante, inclusive a plataforma utilizada (*Forms*), o que limita o construto do pré-teste à compreensão oral.

Durante a pesquisa, continuamos pensando em como incluir a produção oral no pré-teste, mesmo quando já não havia tempo hábil para fazê-lo dentro deste estudo. Uma das alternativas pensadas posteriormente seria a utilização do aplicativo *Whatsapp* para o pré-teste. O aplicativo é disponível em *smartphones* e pode ser utilizado também em computadores, através da função “*Whatsapp Web*”. Esse aplicativo permite gravação e reprodução de áudio com imensa facilidade, além de ser familiar e amplamente utilizado por brasileiros (LINK, 2017). Funcionaria da seguinte maneira: o EPPL possuiria um número oficial para a disponibilização do pré-teste, o candidato entraria em contato com esse número requerendo a avaliação, que lhe seria, então, enviada, com instruções de silenciamento de notificações do celular e, especialmente, da própria plataforma, para que não haja interrupções na realização do teste.

O desafio do “*muddying*” também está presente nessa versão do pré-teste, uma vez que fizemos questões de avaliação de compreensão oral em que o candidato fornece respostas por escrito. De acordo com Weir (2005), a contaminação acontece quando não temos acesso direto a um construto, o que nos faz questionar o que de fato estamos avaliando. No caso do pré-teste, se uma resposta não é ideal, quanto dela se deve ao fato de o participante não ter compreendido o excerto e quanto se deve a uma inabilidade de expressar-se em língua inglesa por escrito? Essa é uma pergunta cuja resposta não podemos afirmar. Além disso, não podemos deixar de mencionar o corretor automático do celular, que pode, ao mesmo tempo, sabotar ou auxiliar o candidato. Durante a pesquisa, houve o cuidado para minimizar essa contaminação por meio de um modelo de correção que não leva em consideração erros de escrita, exceto se impedirem a compreensão.

Outra limitação do pré-teste é a existência de um único modelo. O pré-teste não é um teste que é feito apenas uma vez, e poderia servir de parâmetro durante várias etapas de desenvolvimento e estudo dos candidatos. Assim, seria de grande valia que houvesse um

banco de questões do pré-teste, agrupado em versões diferentes do teste, para que não seja repetido o mesmo teste caso o candidato tivesse interesse de realizá-lo mais de uma vez. Assim, o pré-teste poderia ter validade de 30 dias, por exemplo – que foi o intervalo de validade preditiva com o qual trabalhamos –, e ser reaplicado sempre que o candidato quisesse, em uma nova versão. Se fosse feito no *Whatsapp* a última versão que o candidato fez estaria salva no histórico da conversa, tornando simples a tarefa de alternar entre as versões do pré-teste disponíveis.

Um ponto que havíamos considerado e que não conseguimos instrumentalizar foi a presença de um cronômetro no teste. No início do teste, avisamos o horário inicial, a duração, o horário do término, e que avisaríamos oralmente cinco minutos antes do limite do tempo. Esse aspecto da falta do cronômetro foi levantado também por um participante e nos fez determinar que, além dos avisos iniciais do teste, devemos informar que todos os candidatos possuem relógio em seu computador e/ou celular onde podem consultar o horário de término do teste, para que eles mesmos possam acompanhá-lo para além dos avisos dos examinadores. Para que isso seja bem sucedido, é preciso, antes do teste, confirmar se os relógios dos computadores ou dos celulares dos alunos estão acertados. Essa seria uma solução temporária, pois o ideal seria mesmo apresentar um cronômetro dentro do próprio *layout* do teste, um encaminhamento bastante positivo para próximas pesquisas.

Outro ponto importante interno ao pré-teste foi a limitação de número de palavras. Nas instruções do teste, pedimos que os participantes escrevessem pelo menos 10 palavras. Observando os resultados, percebemos que foi um número que gerou respostas muito próximas de 10 palavras e também muito distantes, chegando a mais de 70 palavras, por exemplo, o que dificultou a correção pela heterogeneidade das respostas disponíveis. Acreditamos, portanto, que esse limite mínimo deve ser maior. Exploramos novas opções no *Forms* e acreditamos que adicionar esse limite no próprio teste é uma boa alternativa, pois o *Forms* apresenta uma função de “regra” que permite impedir o candidato de submeter uma resposta menor que a requisitada. Podemos aplicar em cada questão um número mínimo de palavras ou caracteres diferente, sendo menor em uma questão mais direta e maior em outra da qual se espera uma resposta mais longa. Gostaríamos de poder adicionar também um limite máximo, mas o *Forms* só permite incluir uma regra de número de caracteres, e acreditamos que essa seria a mais útil, visto que a própria duração do teste limita o participante de escrever respostas muito longas.

Um último encaminhamento seria a criação de um pré-teste para o EPPLÉ escrito. O exame é composto por dois testes, o oral e o escrito, e o pré-teste desenvolvido até então pretende prever os resultados apenas do teste oral. Para um pré-teste escrito, a plataforma *Forms* oferece já muitas opções e não limita o construto, podendo ser utilizada para produção e compreensão em língua escrita.

**CONSIDERAÇÕES**

**FINAIS**

Iniciaremos nossas considerações finais retomando as subperguntas de pesquisa e respondendo-as por meio de uma síntese dos resultados obtidos durante este estudo, antes de prosseguirmos para a pergunta central do trabalho e também respondê-la. Assim, as subperguntas são as que se seguem:

1. De que maneiras é possível oferecer o pré-teste em dispositivos tecnológicos móveis?
2. Quais impactos da utilização das tecnologias foi possível observar nas aplicações?
3. Em que medida o pré-teste ao EPPLÉ é válido como instrumento de “previsão” dos resultados do exame?

Quanto à possibilidade de oferecer o pré-teste em DTMs, a resposta é sim, e há várias maneiras de fazê-lo. As possibilidades de integrar o uso de DTMs à aplicação de um teste simples e de baixa relevância como o pré-teste ao EPPLÉ são amplas. A questão de maior importância é quais são, de fato, as vantagens e limitações de se fazer uso desse tipo de dispositivo, e quais adaptações podem ser feitas para tornar o pré-teste mais válido.

A segunda questão trata justamente desse tópico. O impacto da utilização de diferentes tecnologias é amplo e de difícil apreensão, dado o enorme número de variáveis inter-relacionadas quando um teste é aplicado (disponibilidade e qualidade do equipamento utilizado, familiaridade do participante com o equipamento, conhecimento de conteúdo do participante, estado emocional do participante, instruções claras, silêncio, impressão do participante com relação à qualidade do teste, impacto do teste na vida do participante, etc.). Podemos, no entanto, visualizar algumas tendências de acordo com os resultados das avaliações e também as opiniões dos participantes da pesquisa. As aplicações feitas através de computadores parecem ter sido vistas pelos participantes como mais positivas do que as feitas através do celular. Os participantes que realizaram o teste em DTMs apontaram dificuldades com a digitação, especialmente no que tange ao auto-corretor, além de uma preferência geral pelo computador. Um ponto curioso foi o de que, em uma visão geral, as faixas dos participantes foram relativamente semelhantes: o grupo que fez no computador (MG), obteve 23% de faixa 4 e 77% de faixa 3, enquanto que o grupo do celular obteve, proporcionalmente, mais faixas 4 (32%), mas menos faixas 3 (42%), uma vez que nesse grupo alguns participantes também foram posicionados na faixa 2 (26%). Assim, considerando que os grupos são relativamente semelhantes em questão de sua proficiência alta ou intermediária,

podemos dizer que não parece haver grande diferença entre os resultados do computador e do celular.

Em relação ao assunto abordado na questão três, podemos dizer que, como instrumento independente de medida da proficiência, o pré-teste foi visto de forma positiva pelos participantes da pesquisa. Mas tratá-lo isoladamente não basta, já que a proposta inicial é que o pré-teste seja também um bom indicativo de posicionamento no EPPL oral. Na comparação dos resultados do teste com o exame, é possível perceber que os dados apontam para uma relação, embora não totalmente direta, entre os dois instrumentos. Portanto, a nossa terceira pergunta de pesquisa deve ser respondida com cautela. Os resultados estão, paralelamente, nas duas avaliações, hierarquizados, de modo que a maioria dos participantes que obteve maiores notas no pré-teste também obteve maiores notas no EPPL oral. Embora haja dificuldades relacionadas ao uso da tecnologia, acreditamos ser possível afirmar que o pré-teste, da maneira que se encontra, prevê relativamente bem os resultados do EPPL oral. Acreditamos que ainda há muitas adaptações que podem ser feitas para melhorar o teste, mas que ele já esboça resultados positivos de previsão, apresentando um grande potencial para se tornar um indicador consistente para o EPPL.

Um outro ponto importante na questão da validade foi seu dilema com o aumento da acessibilidade do pré-teste. A premissa do desenvolvimento do pré-teste é de que fosse um instrumento acessível, a qualquer hora e lugar, por candidatos que quisessem ter uma indicação prévia de sua proficiência no EPPL oral. Em nossa pesquisa, as aplicações não foram feitas “em qualquer hora e lugar”: foram todas realizadas em ambiente controlado, em silêncio, e com contador de tempo (relógio disponível e avisos do examinador). Sendo essas condições modificadas para que o pré-teste possa ser feito sem supervisão, nos perguntamos até que ponto o resultado de um pré-teste feito em ambientes não ideais (com ruído ou distrações) seria de fato válido (contando com o fato de que fosse inicialmente válido nos ambientes “ideais”). Contaríamos, nesse caso, com o bom senso do candidato em seguir as instruções para que o resultado fosse de fato uma indicação para a realização do EPPL oral. Acreditamos, assim, que essa premissa do “qualquer hora e lugar” pode vir a ser questionada, e talvez tenhamos que limitar algumas liberdades dos candidatos em prol de uma maior validade do teste. Isso só será sabido por meio de mais estudos, uma vez que fosse feita a implantação de fato do pré-teste, e pudéssemos investigar mais casos, especialmente os de disparidade, e averiguar quais suas possíveis causas. Talvez uma alternativa, por ora, poderia ser uma maior afirmação da importância de o candidato estar em um ambiente tranquilo, não

realizar consultas, e responder às questões em, idealmente, 30 minutos, mas não mais que 45 minutos. Essas instruções poderiam ser adicionadas no cabeçalho do teste também por escrito, em língua portuguesa.

Por fim, a pergunta central da pesquisa foi: “é possível o oferecimento de um pré-teste com validade de critério em relação ao EPPL e disponível em dispositivos tecnológicos móveis?”. Acreditamos, considerando as respostas às subperguntas, que a exploração inicial realizada nesse estudo aponta para que seja possível o desenvolvimento de um teste com tais características, embora haja ainda algumas limitações no pré-teste a serem investigadas mais profundamente. Afinal, esse é apenas o segundo estudo realizado sobre o pré-teste ao EPPL, após o primeiro, de Agüena (2016).

Assim, como encaminhamentos, sugerimos estudos semelhantes com participantes de menor proficiência, para que haja uma variedade maior nos dados que nos permita relacionar melhor as faixas mais baixas. Além disso, como descrevemos na seção de avaliações diagnósticas, seria interessante que o pré-teste fornecesse um feedback mais completo aos participantes, indicando-os áreas em que poderiam investir mais estudos. Aplicar o pré-teste a candidatos de maior proficiência também enriqueceria as informações que poderiam ser dadas nesse tipo de feedback.

Outro encaminhamento possível estaria no estudo dos problemas de sub-representação de construto. O pré-teste pretende prever produção e compreensão oral, mas suas tarefas são de compreensão oral e produção escrita. Embora tenhamos tomado cuidados com relação a isso, seria ideal encontrar uma plataforma, como o *Whatsapp*, que foi sugerido no capítulo anterior, que fornecesse recursos para a gravação da produção oral do candidato. Seria de grande valia um estudo que colocasse essa ideia em prática para que pudéssemos analisar melhor as características de um teste dessa natureza, e como ele seria visto pelos candidatos. Mesmo com essa problemática, o pré-teste conseguiu ser capaz de prever o EPPL satisfatoriamente. Acreditamos, no entanto, que a inclusão da produção oral aumentaria a consistência e especificidade dos resultados.

Sobre o formato das respostas, que é de produção escrita, comentamos a contaminação (“*muddying*”) que pode acontecer nesse tipo de avaliação: as respostas, dadas por escrito e em inglês, acabam por ser um obstáculo ao entendimento de qual é de fato a habilidade do candidato. Assim, não seria impensável que um corretor se perguntasse: “essa resposta está incorreta porque o participante não compreendeu o excerto, ou porque não possui habilidade de escrita em LI suficiente para expressar suas respostas?”. Assim, seria necessário minimizar

esse problema da contaminação através da escrita de modo que o candidato pudesse apenas marcar uma resposta, por exemplo, ou, mais idealmente, produzir oralmente, o que resolveria os problemas de sub-representação e de contaminação simultaneamente. Mais pesquisas nesse direção seriam extremamente valiosas para o aprimoramento do pré-teste.

A aplicação do pré-teste a candidatos de níveis de proficiência mais variados é também outra investigação relevante. Em nosso estudo, uma das limitações foi o fato de nossos resultados não serem tão diversos que nos permitissem realizar inferências sobre as faixas mais baixas de proficiência do pré-teste e do EPPLE oral com certa segurança.

Ainda um último ponto digno de mais estudos é a escrita através de DTMs, que acontece com o amparo – ou o embargo – do programa de autocorreção disponível no aparelho. Em se tratando de realizar um teste em língua estrangeira, é importante estudar o impacto desse tipo de escrita, comparando modelos em que o autocorretor está habilitado para a língua portuguesa, para a língua inglesa, ou para ambas.

Por fim, nossa pesquisa demonstrou grandes problemas para se produzir um teste em meio eletrônico, especialmente em DTMs, no Brasil. A área é bastante nova em nosso país, e muitas vezes não há recursos básicos disponíveis nas universidades (como foi o caso dos fones e a má qualidade da conexão com a internet). Apesar dessas limitações, fomos capazes de produzir um teste que consegue prever com certa precisão os resultados do EPPLE oral. Ainda há muito a ser feito em direção à melhoria deste instrumento, mas os primeiros passos já foram dados.

## REFERÊNCIAS

AGUENA, D. M. **Avaliação da proficiência linguístico-comunicativa-pedagógica do professor de línguas: operacionalização de construto no Exame de Proficiência para Professores de Línguas Estrangeiras - ETAPA II.** 2016. 22 f. Relatório de Iniciação Científica (Graduanda em Licenciatura Em Letras) - Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), São José do Rio Preto, 2016.

ALDERSON, J. C. The challenge of (diagnostic) testing: do we know what we are measuring? In: FOX, J. et al. (ed.). **Language Testing Reconsidered.** Ottawa: University of Ottawa Press, 2007, p. 21-39.

ALDERSON, J. C.; CLAPHAM, C.; WALL, D. Validation. IN: \_\_\_\_\_. **Language Test Construction and Evaluation.** New York: Cambridge University Press, 1995, p. 170-196.

ALDERSON, J. C.; HUHTA, A. The development of a suite of computer-based diagnostic tests based on the Common European Framework. **Language Testing**, v. 22, n.3, p. 301-320, 2005.

ALMEIDA FILHO, J. C. P. O professor de Língua Estrangeira sabe a língua que ensina? A questão da instrumentalização lingüística. **Contexturas**, v. 1, n. 1, p. 77-85. São Paulo: APLIESP, 1992.

ANCHIETA, P. P. **Relações entre o construto e as características do teste escrito do EPPLÉ eletrônico: um processo de validação.** 2015. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista “Júlio de Mesquita Filho”, São José do Rio Preto.

BACHMAN, L. F. **Fundamental Considerations in Language Testing.** Oxford: Oxford University Press, 1990.

BAFFI-BONVINO, M. A. **Avaliação da proficiência oral em inglês como língua estrangeira: foco na competência lexical e uma proposta para o processo de validação do descritor “vocabulário” de um teste de proficiência para professores de língua inglesa.** 2010. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista “Júlio de Mesquita Filho”, São José do Rio Preto.

CANALE, M.; SWAIN, M. Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. **Applied Linguistics**, v. 1, n.1, p. 1-47, 1980.

CELANI, M. A. A. Ensino de línguas estrangeiras: ocupação ou profissão. In: LEFFA, V. J. (Org.). **O professor de línguas estrangeiras: construindo a profissão.** Pelotas: EDUCAT, 2008.

CELANI, M. A. A. Perguntas ainda sem resposta na formação de professores de línguas. In: GIMENEZ, T.; MONTEIRO, M. C. G. (Orgs.). **Formação de professores de línguas na América Latina e transformação social.** Campinas: Pontes Editores, 2010.

CHAPELLE, C. A. **Technology in language testing**. 2010. 6 min, son., color. Disponível em: <http://languagetesting.info/video/main.html>. Acesso em: junho de 2018.

CHAPELLE, C. A.; VOSS, E. 20 years of technology and language assessment in Language Learning & Technology. **Language Learning & Technology**, v. 20, n. 2, p; 116–128, 2016. Disponível em: <http://llt.msu.edu/issues/june2016/chapellevoss.pdf>.

CHAPELLE, C; DOUGLAS, D. **Assessing language through computer technology**. Cambridge: Cambridge University Press, 2006.

CHOMSKY, N. **Aspects of the Theory of Syntax**. The MIT Press, Cambridge, Massachusetts. Trad. de José Antonio Meireles e Eduardo Paiva Raposo. Coimbra: Armênio Amado, 1975.

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

CONSOLO, D. A. Assessing EFL teachers' oral proficiency: On the development of teacher education programmes and testing policies in Brazil. IN: CONSOLO, D. A.; GATTOLIN, S. R. B.; TEIXEIRA DA SILVA, V. L. (orgs.), **Perspectivas em Avaliação no Ensino e na Aprendizagem de Línguas: pesquisas e encaminhamentos na formação e na prática docente**, Campinas: Pontes Editores, 2017, p. 59-71.

CONSOLO, D. A. Classroom oral interaction in foreign language lessons and implications for teacher development. **Linguagem & Ensino**, Pelotas, v. 9, n. 2, p. 33-55, 2006a.

CONSOLO, D. A. On a (re)definition of oral language proficiency for EFL teachers: perspectives and contributions from current research. **Melbourne Papers in Language Testing**, v. 1, p. 1-28, 2006b.

CONSOLO, D. A.; AGUENA, D. M. Pré-testagem ao Exame de Proficiência para Professores de Língua Estrangeira (EPPL): desenho e desafios tecnológicos para avaliação de proficiência oral em dispositivos tecnológicos móveis. **Revista Polifonia**, Cuiabá, v. 24, n. 35/2, p. 30-49, 2017.

CONSOLO, D. A.; ALVARENGA, M. B.; CONCÁRIO, M.; LANZONI, H. P.; MARTINS, T. H. B.; SILVA, V. L. T. An examination of foreign language proficiency for teachers (EPPL): The initial proposal and implications for the Brazilian context. In: **II Congresso Internacional da ABRAPUI**, 2009, São José do Rio Preto-SP. The Teaching of English: Towards an Interdisciplinary Approach Between Language and Literature. São José do Rio Preto-SP: ABRAPUI, 2009. p. 1-15.

CONSOLO, D. A.; BERTOLDO, E. S.; COX, M. I. P.; PETERSON, A. A. A.; VIANA, N.; FÉLIX, A. Desafios e perspectivas na formação de professores de línguas no Brasil. **Estudos Linguísticos** (São Paulo), São Paulo - SP, v. 28, n.1, p. 112-117, 1999.

CONSOLO, D. A.; SILVA, V. L. T. Em defesa de uma formação linguística de qualidade para professores de línguas estrangeiras: o exame EPPL. **Revista Horizontes de Linguística Aplicada**, v. 13, p. 63-87, 2014.

CORRÊA, M. L. G. Heterogeneidade da escrita: a novidade da adequação e a experiência do acontecimento. **Filologia e linguística portuguesa**, São Paulo, n. 2, p. 269-286, 2006.

CORRÊA, M. L. G. Letramento e heterogeneidade da escrita no ensino de português. In: SIGNORINI, I. (Org.). **Investigando a relação oral/escrito e as teorias do letramento**. Campinas: Mercado de Letras, 2008, p. 135-166.

COSTA, G. S. **Mobile Learning**: Explorando potencialidades com o uso do celular no ensino-aprendizagem de língua inglesa como língua estrangeira com alunos da escola pública. 2013. Tese (Doutorado em Letras) - Universidade Federal de Pernambuco, Recife.

COUNCIL OF EUROPE. **Common European Framework of Reference for Languages: Learning, Teaching, Assessment**. Cambridge: Cambridge University Press, 2001.

CRYSTAL, D. **Language and the internet**. Cambridge: Cambridge University Press, 2004.

DEANE, P. On the relation between automated essay scoring and modern views of the writing construct. **Assessing Writing**, v. 18, p. 7-24, 2013.

DOUGLAS, D. **Assessing languages for specific purposes**. New York: Cambridge University Press, 2000.

DOUGLAS, D. ESP and Assessment. IN: PALTRIDGE, B.; STARFIELD, S. (Ed.). **The Handbook of English for Specific Purposes**. Malden; Oxford: John Wiley and Sons, Inc., 2013, p. 367-383.

DOUGLAS, D.; HEGELHEIMER, V. Assessing language using computer technology. **Annual Review of Applied Linguistics**, v. 27, p. 115–132, 2007.

EBEL, R. L.; FRISBIE, D. A. Validity: Interpretation and Use. IN: \_\_\_\_\_. **Essentials of Educational Measurement**. 5. ed. Englewood Cliffs: Prentice-Hall International, 1991, p. 100-113.

ERICKSON, F. Qualitative methods in research on teaching. In: WITTROCK, M. C. (Ed.). **Handbook of research on teaching**. New York: Collier-Macmillan, 1986, p. 119-161.

FERNANDES, A. M. A (meta)linguagem para explicação gramatical em língua inglesa: subsídios para elaboração de tarefas do teste oral do EPPL (Exame de Proficiência para Professores de Língua Estrangeira). 2016. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista “Júlio de Mesquita Filho”, São José do Rio Preto.

FLICK, U. **An Introduction to Qualitative Research**. 4th ed. London: Sage Publications, 2009.

FREEMAN, D.; ORZULAK, M. M.; MORRISSEY, G. Assessment in Second Language Teacher Education. In: BURNS, A.; RICHARDS, J. C. **Second Language Teacher Education**. New York: Cambridge University Press, 2009, p. 77-90.

GILLHAM, B. **Case Study Research Methods**. London: Continuum, 2000.

GOMES, H. S. Smartphone passa PC e vira aparelho nº 1 para acessar internet no Brasil. **Portal G1**. Abr. 2016. Disponível em: <http://g1.globo.com/tecnologia/noticia/2016/04/smartphone-passa-pc-e-vira-aparelho-n-1-para-acessar-internet-no-brasil.html>. Acesso em: maio de 2017.

GREEN, A. Washback in language assessment. **International Journal of English Studies**, v. 13, n. 2, p. 39-51, 2013.

GUO, H. **Analysing and Evaluating Current Mobile Applications for Learning English Speaking**. 2014. Dissertação – University of Birkbeck, Londres. Disponível em: <https://englishagenda.britishcouncil.org/research-publications/elt-masters-dissertations/2014-2015-winners>. Acesso em: dezembro de 2018.

HUGHES, A. Validity. IN: \_\_\_\_\_. **Testing for language teachers**. Cambridge: Cambridge University Press, 1989, p. 22-28.

HYMES, D. On Communicative Competence. IN: PRIDE, J.; HOLMES, J. **Sociolinguistics**. Harmondsworth: Penguin, 1972.

JAMIESON, J. Trends in computer-based second language assessment. **Annual Review of Applied Linguistics**, v. 25, p. 228–242, 2005.

JANG, E. E. Diagnostic assessment in language classrooms. In: FULCHER, G.; DAVIDSON, F. (ed.). **The Routledge Handbook of Language Testing**. New York: Taylor & Francis Group, 2012, p. 120-133.

KENSKI, V. M. **Educação e tecnologias: o novo ritmo da informação**. Campinas: Editora Papirus, 2007.

KESSLER, G. Authoring Tools for Language Assessment. IN: CHAPPELLE, A. (Ed.). **The Encyclopedia of Applied Linguistics**. Oxford, UK: Wiley-Blackwell, 2013.

KUKULSKA-HULME, A.; VIBERG, O. Mobile collaborative language learning: State of the art. **British Journal of Educational Technology**, v. 49, n. 2, p. 207-218, 2018.

LABORDA, J. G.; ROYO, T. M.; BAKIEVA, M. Looking towards the Future of Language Assessment: Usability of Tablet PCs in Language Testing. **Journal of Universal Computer Science**, v. 22, n. 1, p. 114-123, 2016.

LAZARATON, A. Qualitative research methods in language test development and validation. IN: MILANOVIC, M.; WEIR, C. J. (Ed.) **Studies in Language Testing 18: European Language Testing in a global context**. Cambridge University Press, 2004.

LINK, Redação. WhatsApp chega a 120 milhões de usuários no Brasil. **Estadão**, São Paulo, 29 de maio de 2017. Disponível em: <https://link.estadao.com.br/noticias/empresas,whatsapp-chega-a-120-milhoes-de-usuarios-no-brasil,70001817647>. Acesso em: outubro de 2018.

LLURDA, E. On competence, proficiency, and communicative language ability. **International Journal of Applied Linguistics**, v. 10, n. 1, p. 85-96, 2000.

LONG, M. H.; PORTER, P. A. Group Work, Interlanguage Talk, and Second Language Acquisition. **TESOL Quarterly**, v. 19, n. 2, p. 207-228, 1985.

MANS, M. Uso de smartphones já alcança 80% dos brasileiros, aponta estudo. **Estado de São Paulo**. Out. 2016. Disponível em: <http://link.estadao.com.br/noticias/cultura-digital,uso-de-smartphones-ja-alcanca-80-dos-brasileiros-aponta-estudo,10000082680>. Acesso em: maio de 2017.

MARCUSCHI, L. A. Gêneros textuais emergentes no contexto da tecnologia digital. IN: MARCUSCHI, L. A.; XAVIER, A. C. (Org.). **Hipertexto e gêneros digitais: novas formas de construção de sentido**. 3ª ed. São Paulo: Cortez, 2010, p. 15-80.

MARQUES-SCHÄFER, G. ROZENFELD, C. C. F. (Org.). **Ensino de línguas e tecnologias móveis: políticas públicas, conceitos, pesquisas e práticas em foco**. São Paulo: Edições Hipótese, 2018.

MARTINS, T. H. B.; SILVA, V. L. T.; SILVA, W. E. Desafios para se aplicar um exame de proficiência em meio eletrônico. IN: CONSOLO, D. A.; GATTOLIN, S. R. B.; TEIXEIRA DA SILVA, V. L. (orgs.), **Perspectivas em Avaliação no Ensino e na Aprendizagem de Línguas: pesquisas e encaminhamentos na formação e na prática docente**, Campinas: Pontes Editores, 2017, p. 191-206.

MCNAMARA, T. **Language Testing**. Oxford: Oxford University Press, 2000.

MCNAMARA, T. Language assessment as social practice: challenges for research. **Language Testing**, v. 18, n. 4, p. 333-349, 2001.

NORTH, B. Linking language assessments: an example in a low stakes context. **System**, v. 28, p. 555-577, 2000.

PAIVA, V. L. M. A WWW e o ensino de inglês. **Revista Brasileira de Linguística Aplicada**, v. 1, n.1, p. 93-116, 2001.

PATHAN, M. M. Computer Assisted Language Testing (CALT): Advantages, Implications and Limitations. **Research Vistas**, v. 1, n. 4, p. 30-45, 2012.

**Preliminary Test I of for the ECPE** (Examination for the Certificate of Proficiency in English). Michigan University. Disponível em: [www.hau.gr/resources/toolip/doc/2012/02/17/0506\\_prel1\\_test.pdf](http://www.hau.gr/resources/toolip/doc/2012/02/17/0506_prel1_test.pdf). Acesso em: maio de 2017.

RAHMAN, M. S. The Advantages and Disadvantages of Using Qualitative and Quantitative Approaches and Methods in Language “Testing and Assessment” Research: A Literature Review. **Journal of Education and Learning**, Toronto, v. 6, n. 1, p. 102-112, 2017.

ROEVER, C. Web-based language testing. **Language Learning & Technology**, v. 5, n. 2, p. 84-94, 2001.

SCARAMUCCI, M. V. R. O exame de proficiência em Língua Inglesa para controladores de voo do SISCEAB. **Aviation in Focus**, v. 2, n. 1, p. 3-12, 2011. Entrevista concedida à revista.

SCARAMUCCI, M. V. R. Proficiência em LE: considerações terminológicas e conceituais. **Trabalhos Em Lingüística Aplicada** (UNICAMP), Campinas, v. 36, p. 11-22, 2000.

SCARAMUCCI, M. V. R. Validade e consequências sociais das avaliações em contextos de ensino de línguas. **Linguarum Arena**, Porto, v. 2, p. 103-120, 2011.

SIGNORINI, I. Construindo com a escrita “outras cenas de fala”. In: SIGNORINI, I. (Org.). **Investigando a relação oral/escrito e as teorias do letramento**. Campinas: Mercado de Letras, p. 135-166, 2008.

SHOHAMY, E. Critical Language Testing and Beyond. **Studies in Educational Evaluation**, v. 24, n. 4, p. 331-345, 1998.

STERN, H. H. Models of second language learning and the concept of proficiency. IN: \_\_\_\_\_. **Fundamental Concepts of Language Teaching**. Oxford: Oxford University Press, 1987.

STOCKWELL, G. Using mobile phones for vocabulary activities: examining the effect of the platform. **Language Learning and Technology**, v. 14, n. 2, p. 95-110, 2010. Disponível em: <http://lt.msu.edu/vol14num2/stockwell.pdf>. Acesso em: dezembro de 2018.

SUVOROV, R.; HEGELHEIMER, V. Computer-Assisted Language Testing. IN: KUNNAN, A. J. (ed.). **The Companion to Language Assessment**. v. 2. John Wiley & Sons, Inc., 2014. Disponível em: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118411360.wbcla083>. Acesso em: junho de 2018.

TARIGHAT, S.; KHODABAKHSH, S. Mobile-Assisted Language Assessment: Assessing speaking. **Computers in Human Behavior**, v. 64, p. 409-413, 2016.

TUDOR, I. Learning to live with complexity: towards an ecological perspective on language teaching. **System**, v. 31, p. 1-12, 2003.

UNESCO. *Diretrizes de políticas para a aprendizagem móvel*. Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO). Tradução para o Português. Unesco no Brasil, 2014. Disponível em: <http://unesdoc.unesco.org/images/0022/002277/227770por.pdf>. Acesso em: agosto de 2018.

VIBERG, O. **Design and Use of Mobile Technology in Distance Language Education: Matching Learning Practices with Technologies-in-Practice.** Örebro: Örebro University, 2015.

WEIR, C. J. **Language testing and validation: an evidence-based approach.** London: Palgrave Macmillan, 2005.

# **ANEXOS**

## 1. Modelo do Termo de Consentimento

### Termo de Consentimento Livre e Esclarecido - TCLE

(Conselho Nacional de Saúde, Resolução 466/2012/Resolução 510/2016)

Você está sendo convidado a participar como voluntário do projeto de pesquisa “O pré-teste do EPPLE (exame de proficiência para professores de línguas estrangeiras): desafios tecnológicos, validade e histórico de implementação”, sob responsabilidade da pesquisadora Jéssica Nunes Caldeira Cunha. O estudo será realizado através da aplicação e coleta de resultados do Pré-Teste do EPPLE e exame EPPLE aos candidatos, além da utilização das respostas aos questionários entregues imediatamente após a realização das avaliações, para investigar em que medida o Pré-Teste é um instrumento de avaliação válido em relação ao EPPLE. Os riscos físicos, mentais ou emocionais para os envolvidos na pesquisa são mínimos. Você poderá consultar a pesquisadora responsável em qualquer época, pessoalmente ou pelo telefone da instituição, para esclarecimento de qualquer dúvida. Você está livre para, a qualquer momento, deixar de participar da pesquisa. Todas as informações por você fornecidas e os resultados obtidos serão mantidos em sigilo, e estes últimos só serão utilizados para divulgação em reuniões e revistas científicas, sem qualquer divulgação sobre as identidades dos participantes. Você será informado de todos os resultados obtidos, independentemente do fato de estes poderem mudar seu consentimento em participar da pesquisa. Você não terá quaisquer benefícios ou direitos financeiros sobre os eventuais resultados decorrentes da pesquisa. Este estudo é importante porque seus resultados fornecerão informações para estudos sobre avaliação da proficiência oral do professor de línguas estrangeiras, mais especificamente úteis à validação do Pré-Teste do EPPLE em relação ao exame EPPLE. Diante das explicações, se você concorda em participar deste projeto, forneça os dados solicitados e coloque sua assinatura a seguir.

Nome: \_\_\_\_\_ R.G. \_\_\_\_\_

Endereço: \_\_\_\_\_ Fone: \_\_\_\_\_

\_\_\_\_\_, de \_\_\_\_\_ de 20\_\_\_\_

---

Assinatura Do Participante /  
Usuário Ou Responsável Legal

---

Assinatura Da Pesquisadora  
Responsável / Coordenador Do Projeto Do  
Exame EPPLE

Nome da Pesquisadora: Jéssica Nunes Caldeira Cunha	Cargo/Função: Mestranda
Instituição: Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP/IBILCE	
Endereço: Rua Cristóvão Colombo, 2265 – São José do Rio Preto	
Projeto submetido ao Comitê de Ética em Pesquisa do IBILCE/UNESP São José do Rio Preto/SP – Fone 17-3221.2480 e 3221.2545	

## 2. Modelo do questionário

### QUESTIONÁRIO – PRÉ-TESTE

Pesquisadora responsável: Jessica Nunes Caldeira Cunha

Caro (a) participante,

O objetivo deste questionário é coletar informações sobre seu perfil e formação, além de suas impressões sobre o Pré-Teste ao EPPLE após realizá-lo. As informações coletadas são de extrema importância para a pesquisa *O pré-teste do EPPLE (Exame de Proficiência para Professores de Línguas Estrangeiras): desafios tecnológicos, validade e histórico de implementação*. Sua identificação nesse questionário é necessária para cruzamento de dados de pesquisa, mas não será divulgada.

Muito obrigada pela participação!

Nome: \_\_\_\_\_

Email de contato: \_\_\_\_\_

Estudante de:  Graduação  Pós-Graduação

Semestre/Ano de ingresso: \_\_\_\_\_

1. Idade: \_\_\_\_\_

2. Você possui algum certificado de proficiência internacional (TOEFL, FCE,...)?

Sim  Não

Se sim, qual? \_\_\_\_\_

Em que ano prestou o exame? \_\_\_\_\_

Que nota ou faixa de proficiência obteve? \_\_\_\_\_

3. Você atua como professor de língua inglesa atualmente?

Sim  Não

4. Em toda a sua vida, por quanto tempo você já foi professor de língua inglesa?

( ) Nunca

( ) De 0 a 1 ano

( ) De 1 a 2 anos

( ) De 2 a 3 anos

( ) Mais de 4 anos

5. Em que contexto atua/atuou como professor de língua inglesa? Marque todas as alternativas que forem válidas.

- ( ) Curso de extensão universitário
- ( ) Aulas particulares
- ( ) Aulas em cursinhos privado
- ( ) Aulas em escolas regulares públicas
- ( ) Aulas em escolas regulares particulares

6. Se concluiu, interrompeu ou ainda faz um curso em escola(s) de línguas, até que nível você chegou?

- ( ) básico
- ( ) intermediário superior
- ( ) pré-intermediário
- ( ) avançado
- ( ) intermediário
- ( ) outro: \_\_\_\_\_

7. Você autoriza a pesquisadora a contatá-lo(a) para realizar uma entrevista com o objetivo de esclarecer e explorar suas respostas aos questionários?

- Sim                       Não

**PRÉ-TESTE ao EPPLE**

8. O pré-teste avaliou satisfatoriamente a minha proficiência em língua inglesa.

Concordo totalmente	Concordo parcialmente	Não concordo nem discordo	Discordo parcialmente	Discordo totalmente

9. As orientações para a execução do pré-teste são claras.

Concordo totalmente	Concordo parcialmente	Não concordo nem discordo	Discordo parcialmente	Discordo totalmente

10. O tempo disponibilizado para responder as questões do pré-teste é suficiente.

Concordo totalmente	Concordo parcialmente	Não concordo nem discordo	Discordo parcialmente	Discordo totalmente

11. A plataforma do pré-teste (*Google Forms*) é de fácil utilização.

Concordo totalmente	Concordo parcialmente	Não concordo nem discordo	Discordo parcialmente	Discordo totalmente

12. O pré-teste do EPPLE avalia adequadamente a proficiência linguística de (futuros) professores de língua inglesa.

Concordo totalmente	Concordo parcialmente	Não concordo nem discordo	Discordo parcialmente	Discordo totalmente

13. A qualidade do áudio do teste permite boa compreensão de seu conteúdo.

Concordo totalmente	Concordo parcialmente	Não concordo nem discordo	Discordo parcialmente	Discordo totalmente

14. Você teria críticas ao pré-teste do EPPLE? (Você encontrou algum problema ao realizar o pré-teste?)

---

---

---

---

---

---

15. Você teria sugestões ou comentários sobre o pré-teste do EPPLE?

---

---

---

---

---

---

\*A alguns dos estudantes que realizaram o pré-teste no celular (grupo SP<sub>1</sub>), o questionário era o mesmo acrescido das três seguintes questões, posicionadas logo após a questão 13 e antes das duas últimas questões.

A. Quais foram as vantagens de realizar o teste através de um dispositivo móvel (celular)?

---

---

---

---

---

---

B. Quais foram as desvantagens de realizar o teste através de um dispositivo móvel (celular)?

---

---

---

---

---

C. Se pudesse escolher, você faria esse teste pelo computador ou pelo celular? Por quê?

---

---

---

---

---

### 3. Script do pré-teste 2

#### PRÉ-TESTE – 2017 – SCRIPT

*Welcome to the pretest for the EPPLE examination. You are going to hear 8 questions, and then write the answers in English in the following box. From questions one (1) to four (4), you will hear general questions. After question four (4), you will hear an audio recording. Questions five (5) to eight (8) are based on this recording. Your answer to each question must contain at least ten (10) words. Have a good experience in the EPPLE pretest!*

1 – What are the best things about the city where you live?

2 - What expectations do you have about you career in the near future?

3 – When did you start learning English?

4 – A student comes to you and asks the following question: “I wrote in my composition the sentence: ‘I want to speak English really good’. Why is it wrong?” Pretend you are his teacher and explain the mistake in the sentence that confused the student.

*Intro: “Listen to the linguist David Crystal talk about how English became a global language. You can play the audio twice. Then, answer questions 5 to 8 about the recording”.*

(AUDIO) Global English

*Audioscript: [A language becomes a global language because of the power of the people who speak it. It’s nothing to do with the structure of language, there’s nothing about the English grammar or vocabulary or pronunciation or spelling that makes English an appealing language at a global level. In fact, if you think about it, English spelling would put most people off, you’d think. So it’s nothing to do with the structure of language, no, it’s all to do with power. But power means different things at different times. English first became international because of political power, military power, the power of the British empire, really. But it isn’t just political that takes a language around the world, there gotta be other factors too. A century later, we’re talking about the 16<sup>th</sup>, 17<sup>th</sup> century there, we had the power of science, of technology, the Industrial Revolution, English is the language of science. That started then, as something like two thirds of the people who invented all the things that make modern society what it is did so through the medium of the English language. And then, in the 19<sup>th</sup> century, economic power, money talks always, and the language it was talking in the 19<sup>th</sup> century was English, because America and Britain, between them, had the money markets of the world so not, for the most part. And then, in the 20<sup>th</sup> century, we have cultural power, with English being the language of the vast majority of the inventions that make modern society what it is. So it’s English turning up in the, as it were, always at the right place at the right time during these last 400 years or so that has produced the enormous global status that it currently has.]*

5 – According to the speaker, why does a language become a global language?

6 – What does the speaker mean when he says “power means different things at different times”?

7 – The speaker says “So it’s English turning up in the right place and the right time during these last 400 years or so that has produced the enormous global status that it currently has”. Why does the speaker use the form “has produced” in this sentence?

In the following sentence, the speaker uses the form “has produced”: “*So it’s English turning up in the, as it were, always at the right place at the right time during these last 400 years or so that has produced the enormous global status that it currently has.*” Why was the verb used in this form in this context?

8 – The speaker says that some aspects of English would probably put most people off. What does the expression “put off” mean in this context?

#### **4. Construto do EPPL (Anchieta, 2015)**

O Exame de Proficiência para Professores de Línguas Estrangeiras (EPPL) é um exame comunicativo que objetiva avaliar a proficiência linguística oral e escrita do candidato em termos de uso contextualizado da língua, em situações reais de comunicação de ensino de língua estrangeira, tomando por base o desempenho global do candidato nas tarefas do exame, sendo seu resultado expresso em faixas de proficiência, com base em descritores de competência e desempenho linguístico comunicativo do candidato.

Adotamos no construto do EPPL a visão de Stern (1983) ao entender que o termo proficiência significa o desempenho efetivo de um falante de uma determinada língua, e envolve o domínio (a) das formas, (b) dos significados linguísticos, cognitivos, afetivos e socioculturais dessas formas, (c) a capacidade de utilizar a língua com máximo enfoque na comunicação, e (d) a criatividade no uso da língua. O uso de uma língua inclui ações realizadas pelos falantes/usuários, indivíduos que desenvolvem um conjunto de competências gerais e, particularmente, de competências comunicativas nessa língua, as quais implicam processos linguísticos para produzirem e reproduzirem textos relacionados com temas pertencentes a domínios específicos, de modo a ativar estratégias mais apropriadas para o desempenho de tarefas comunicativas.

Dessa forma, entende-se que a construção da competência de ensino do professor de língua estrangeira é um processo que se solidifica no interior das relações sociais, através da interação do professor com seus alunos, com colegas de disciplina, bem como com professores de outras disciplinas, coordenadores, diretores e supervisores, mediado pelo diálogo, construindo a partir dessas interações processos de significação. Por meio da interação há a construção de conhecimentos, e, para tanto os professores de LE desenvolvem um conjunto de competências de linguagem gerais e específicas.

Na competência geral, o uso da língua para tratar de assuntos da vida cotidiana, e para interagir socialmente na sala de aula, de modo a poder participar efetivamente de situações formais e informais tanto na fala quanto na escrita, compreendendo mensagens, desenvolvendo tópicos e opiniões, fazendo leitura de textos informativos, e na produção escrita de gêneros textuais variados. Na competência específica, o uso da linguagem metalinguística na terminologia dos assuntos tratados pelo professor e a discursiva para se adequar ao discurso da sala de aula.

## 5. Faixas de proficiência – Pré-teste 2

1	<p><b>Variação:</b> Tem uma gama suficiente de linguagem para produzir descrições claras, expressar pontos de vista sobre temas gerais, usando algumas formas de sentenças complexas para fazê-lo.</p> <p><b>Precisão:</b> Mostra um grau relativamente elevado de controle gramatical e não comete erros que causam mal-entendidos. Pode cometer erros de ortografia, causados ou não pelo autocorretor, desde que não alterem a inteligibilidade da mensagem. Marcas de oralidade também não devem ser penalizadas.</p> <p><b>Coerência/Coesão:</b> Pode usar um número limitado de ferramentas de coesão para conectar suas declarações contribuindo para um discurso claro e coerente.</p> <p><b>Compreensão oral:</b> Compreende bem um excerto de áudio e é capaz de selecionar informações para responder adequadamente a questões sobre o excerto.</p> <p><b>Metalinguagem:</b> Conhece bem as regras linguísticas, sendo capaz de explicá-las de modo claro.</p>
2	<p><b>Variação:</b> Tem conhecimento suficiente da língua para sobrevivência, com vocabulário para expressar-se sobre temas como família, hobbies e interesses, trabalho, viagens e eventos atuais.</p> <p><b>Precisão:</b> Utiliza padrões estruturais associados a situações de uso com precisão linguística razoável. Pode cometer erros de ortografia, causados ou não pelo autocorretor, desde que não alterem a inteligibilidade da mensagem. Marcas de oralidade também não devem ser penalizadas.</p> <p><b>Coerência/Coesão:</b> Pode relacionar uma série de elementos de conteúdo, curtos e simples, em uma sequência coerente e linear.</p> <p><b>Compreensão oral:</b> Compreende um excerto de áudio e é capaz de selecionar informações para responder a questões sobre o excerto, com algumas limitações.</p> <p><b>Metalinguagem:</b> Conhece as regras da língua, mas apresenta dificuldade em explicá-las de modo claro ou em selecionar informações relevantes para esclarecer dúvidas de alunos, fazendo uso limitado de terminologia específica de metalinguagem.</p>
3	<p><b>Variação:</b> Utiliza padrões básicos de sentenças e grupos de poucas palavras e fórmulas, a fim de comunicar informações limitadas em situações cotidianas simples.</p> <p><b>Precisão:</b> Utiliza algumas estruturas simples corretamente, mas ainda comete erros básicos esporadicamente. Pode cometer erros de ortografia, causados ou não pelo autocorretor, desde que não alterem a inteligibilidade da mensagem. Marcas de oralidade também não devem ser penalizadas.</p> <p><b>Coerência/Coesão:</b> Pode relacionar grupos de palavras com conectores simples tais como “and”, “but” e “because”.</p> <p><b>Compreensão oral:</b> Compreende um excerto de áudio com certa dificuldade, e não consegue responder completamente às questões sobre o excerto.</p> <p><b>Metalinguagem:</b> Conhece as regras de uso da língua parcialmente, mas não</p>

	consegue explicá-las de modo claro.
4	<p><b>Varição:</b> Tem um repertório básico de palavras e sintagmas simples relacionados a dados pessoais e situações concretas específicas.</p> <p><b>Precisão:</b> Mostra apenas um domínio limitado de algumas estruturas gramaticais e padrões de sentenças simples. Pode cometer erros de ortografia, causados ou não pelo autocorretor, desde que não alterem a inteligibilidade da mensagem. Marcas de oralidade também não devem ser penalizadas.</p> <p><b>Coerência/Coesão:</b> Pode relacionar palavras ou grupos de palavras com conectores lineares muito básicos como “and” ou “so”.</p> <p><b>Compreensão oral:</b> Apresenta dificuldade na compreensão de um excerto de áudio e não consegue responder a questões de compreensão sobre este excerto.</p> <p><b>Metalinguagem:</b> Pouco ou nenhum conhecimento sobre regras da língua, não conseguindo explicá-las.</p>
X	<b>Não respondeu à questão.</b>

## 6. Faixas de proficiência – EPPLE oral

**Proposta preliminar de**  
**FAIXAS DE PROFICIÊNCIA PARA O TESTE ORAL DO EPPLE**  
(com base em Baffi-Bonvino, 2010; Consolo, 2004; Consolo e Teixeira da Silva,  
2007)

---

### **FAIXA A**

A1) Atinge plenamente os objetivos de comunicação verbal, apresentando fluência em termos de velocidade e ritmo de fala.

A2) Exibe padrões de pronúncia bastante semelhante a falantes competentes da língua estrangeira, sem influências marcantes dos padrões de sua língua materna, nem causar qualquer desconforto ou incompreensão por parte do interlocutor.

A3) Fornece informações sobre experiências presentes e passadas de modo claro, e utiliza estruturas linguísticas (por exemplo, formas verbais de passado) e vocabulário, incluindo expressões lexicais ( *“the year before”* ) adequados e específicos para as necessidades de produção oral.

A4) Narra e descreve, de modo detalhado, uma cena de vídeo. Levanta hipóteses a respeito da cena e estabelece associações com outras experiências, sugerindo contribuições para o ensino e a aprendizagem de língua estrangeira.

A5) Expressa conhecimento sobre regras de uso da língua e seleciona, desse conhecimento, as informações necessárias para esclarecer dúvidas linguísticas de alunos. É capaz de explicar regras linguísticas de modo claro, fazendo uso de terminologia específica, por exemplo, sobre classes de palavras e estruturas gramaticais. As explicações seriam plenamente compreendidas por alunos de língua estrangeira.

A6) Não apresenta dificuldades de compreensão da fala na língua estrangeira em ritmo normal.

---

### **FAIXA B**

B1) Atinge plenamente os objetivos de comunicação verbal, apresentando fluência em termos de ritmo de fala, e se comete erros gramaticais, é capaz de se auto-corrigir.

B2) Exibe pronúncia bastante próxima aos padrões de falantes competentes da língua estrangeira, sem influências marcantes dos padrões de sua língua materna, e sem causar incompreensão do interlocutor.

B3) Fornece informações sobre experiências presentes e passadas, e utiliza estruturas linguísticas (por exemplo, formas verbais de passado) e vocabulário, incluindo expressões lexicais ( *“the year before”* ) adequados e específicos para as necessidades de produção oral.

B4) Descreve uma cena de vídeo. Levanta hipóteses a respeito da cena mas apresenta certa dificuldade em estabelecer associações com outras experiências e para sugerir contribuições para o ensino e a aprendizagem de língua estrangeira.

B5) Expressa conhecimento sobre regras de uso da língua mas apresenta dificuldade em explicá-las de modo claro, ou em selecionar as informações relevantes para esclarecer dúvidas

linguísticas de alunos, fazendo uso limitado de terminologia específica de metalinguagem. Suas explicações não seriam plenamente compreendidas por alunos de língua estrangeira.

B6) Não apresenta dificuldades de compreensão da fala na língua estrangeira em ritmo normal.

---

### **FAIXA C**

C1) Atinge os objetivos de comunicação verbal, podendo apresentar limitações na fluência, em termos de ritmo e de velocidade de fala, e no uso de estruturas linguísticas.

C2) Exibe pronúncia compreensível, porém com alguns desvios com relação aos padrões de falantes competentes da língua estrangeira.

C3) Fornece informações sobre experiências presentes e passadas utilizando estruturas linguísticas simples e pouca variedade lexical.

C4) Descreve uma cena de vídeo mas apresenta dificuldade em levantar hipóteses a respeito da cena ou em estabelecer associações com outras experiências, e para sugerir contribuições para o ensino e a aprendizagem de língua estrangeira.

C5) Expressa conhecimento parcial sobre regras de uso da língua mas não consegue explicá-las de modo claro. Tem dificuldade em selecionar informações para esclarecer dúvidas linguísticas de alunos, e sua fala seria compreendida apenas parcialmente por alunos de língua estrangeira.

C6) Pode apresentar dificuldades de compreensão da fala em língua estrangeira mas não a ponto de prejudicarem seu desempenho verbal.

---

### **FAIXA D**

D1) Atinge os objetivos de comunicação verbal com limitações. Exibe falta de fluência e de competência no uso de estruturas linguísticas.

D2) Exibe pronúncia compreensível, mas distinta, em alguns aspectos de sons e padrões de entoação, de falantes da língua estrangeira, podendo haver alguma interferência na compreensão e causar certo desconforto ao interlocutor.

D3) Fornece informações sobre experiências presentes e passadas utilizando estruturas linguísticas simples, e pouca variedade lexical.

D4) Descreve uma cena de vídeo mas apresenta dificuldade em levantar hipóteses a respeito da cena e em estabelecer associações com outras experiências, e também para sugerir contribuições ao ensino e à aprendizagem de língua estrangeira.

D5) Expressa conhecimento limitado sobre regras de uso da língua e não consegue selecionar quais informações são relevantes para esclarecer dúvidas linguísticas de alunos, ou não consegue explicar as regras da língua de modo claro. Sua fala seria de difícil compreensão por alunos de língua estrangeira.

D6) Apresenta dificuldades de compreensão da fala na língua estrangeira em ritmo normal, as quais podem, ocasionalmente, prejudicar o desenvolvimento de seu desempenho verbal.

---

### **FAIXA E**

E1) Não atinge satisfatoriamente os objetivos de comunicação verbal, apresentando falta de fluência e de competência na produção oral.

E2) Exibe pronúncia nitidamente distinta, em aspectos de sons e padrões de entoação, de falantes da língua estrangeira, com interferências marcantes de sua língua materna, que podem causar incompreensão e desconforto ao interlocutor.

E3) Fornece informações sobre experiências presentes e passadas utilizando estruturas linguísticas simples e pouca variedade lexical, e comete erros estruturais, com prejuízo para sua expressão oral.

E4) Apresenta dificuldade em descrever uma cena de vídeo e em levantar hipóteses a respeito da cena, bem como em estabelecer associações com outras experiências e para sugerir contribuições para o ensino e a aprendizagem de língua estrangeira.

E5) Expressa pouco ou nenhum conhecimento sobre regras de uso da língua, e não consegue explicar regras da língua de modo a esclarecer dúvidas linguísticas de alunos de língua estrangeira.

E6) Apresenta dificuldades de compreensão da fala em língua estrangeira em ritmo normal, as quais prejudicam seu desempenho verbal.

---