



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

Camila Sthéfanie Colombo

Avaliação de Proficiência Oral em Língua Estrangeira:

Uma Proposta de Abordagem Avaliativa em um Exame para Professores de Línguas com base em Métodos de Estruturação de Problemas e em Métodos Multicritério de Análise da Decisão

São José do Rio Preto

2019

Camila Sthéfanie Colombo

Avaliação de Proficiência Oral em Língua Estrangeira:

Uma Proposta de Abordagem Avaliativa em um Exame para Professores de Línguas com base em Métodos de Estruturação de Problemas e em Métodos Multicritério de Análise da Decisão

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

Orientador: Prof. Dr. Douglas Altamiro Consolo.

São José do Rio Preto

2019

Colombo, Camila Sthéfanie.

C718a Avaliação de Proficiência Oral em Língua Estrangeira: Uma Proposta de Abordagem Avaliativa em um Exame para Professores de Línguas com Base em Métodos de Estruturação de Problemas e em Métodos Multicritério de Análise da Decisão / Camila Sthéfanie Colombo. -- São José do Rio Preto, 2019

262 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto

Orientador: Douglas Altamiro Consolo

1. Linguística Aplicada. 2. Avaliação de Proficiência. 3. Proficiência de Professores de Línguas. 4. Proficiência Oral. 5. Inglês como Língua Estrangeira. I. Título

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Camila Sthéfanie Colombo

Avaliação de Proficiência Oral em Língua Estrangeira:

Uma Proposta de Abordagem Avaliativa em um Exame para Professores de Línguas com base em Métodos de Estruturação de Problemas e em Métodos Multicritério de Análise da Decisão

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

Comissão Examinadora

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

Prof. Dr. Alexandre Bevilacqua Leoneti
USP – Ribeirão Preto

Profa. Dra. Vera Lúcia Teixeira da Silva
UFSCar – Câmpus São Carlos

Profa. Dra. Solange Aranha
UNESP – Câmpus de São José do Rio Preto

Profa. Dra. Suzi Marques Spatti Cavalari
UNESP – Câmpus de São José do Rio Preto

Suplentes

Profa. Dra. Teresa Helena Buscato Martins
UFBA – Câmpus Universitário Ondina

Profa. Dra. Paula Tavares Pinto
UNESP – Câmpus de São José do Rio Preto

Profa. Dra. Sandra Regina Butros Gattolin de Paula
UFSCar – Câmpus São Carlos

São José do Rio Preto
29 de agosto de 2019

Às mulheres da ciência brasileira.

AGRADECIMENTOS

A Deus, pela vida;

Ao meu orientador, Prof. Dr. Douglas Altamiro Consolo, pela oportunidade, pela confiança, pelo incentivo, pela compreensão e pelo exemplo de profissionalismo.

À Profa. Dra. Solange Aranha, à Profa. Dra. Suzi Marques Spatti Cavalari, à Profa. Dra. Vera Lúcia Teixeira da Silva e ao Prof. Dr. Alexandre Bevilacqua Leoneti pelas contribuições e pela disponibilidade para debater este trabalho.

À Profa. Dra. Maria Helena Vieira Abrahão, à Profa. Dra. Teresa Helena Buscato Martins, à Profa. Dra. Paula Tavares Pinto e à Profa. Dra. Sandra Regina Butros Gattolin de Paula, pela disponibilidade para compor o corpo suplente do Exame de Qualificação e da Defesa deste trabalho. Em especial, à Profa. Sandra, pelas reflexões tecidas na ocasião do IX SeLin.

Ao Prof. Dr. Edson Rosa Francisco de Souza pelos conhecimentos compartilhados na orientação de meu trabalho de qualificação especial.

À Profa. Dra. Cláudia Maria Ceneviva Nigro, por despertar meu interesse e oportunizar minha iniciação no universo da pesquisa.

A todos os docentes e funcionários do IBILCE e, em especial, aos associados ao Programa de Pós-Graduação em Estudos Linguísticos, pela qualidade dos serviços prestados e por garantirem um ambiente de formação acadêmica de qualidade reconhecida e mentalmente saudável.

Ao IFSP – Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, por incentivar e fomentar a qualificação de seu corpo docente. Em particular, a meus colegas do Câmpus Barretos, pelo suporte constante a meus compromissos de pesquisa.

Ao meu marido, Kaio, pelo amor, pelo companheirismo, pelo suporte, pela inspiração e pelos incentivos constantes.

Aos meus pais, Aparecida e João, e a meu irmão, João Victor, por todo incentivo e carinho.

A minha sogra, Cecília, pelo apoio e pelo carinho.

A meus avós, Aparecida e Antônio, Pedro e Elydia (*in memoriam*), pelo carinho.

Aos membros do grupo de pesquisa ENAPLE-CCC, Aline, Elen, Rodrigo, Jéssica, William, Diego e José Ricardo, pelos conhecimentos compartilhados.

A minhas colegas de trabalho Aline, Marcela, Natalia, Pricila, Rita, Rozângela, Laura e Isabel, pela compreensão, pelo apoio e pela descontração. Em especial, à Marcela, pela amizade.

Às amigas Jaqueline, Michele, Adelaide e Carollina, pela amizade, pela torcida e pelo incentivo.

À Daniela Bianchi Pandim, Camila Uenoyama Moura Miranda Pereira e Mariana dos Santos Murra, pelos cuidados propiciados durante esta caminhada.

Aos participantes desta pesquisa, que, voluntariamente, dedicaram parte de seu tempo ao compartilhamento de conhecimento.

A todos aqueles que se dedicam ao desenvolvimento da pesquisa brasileira e à manutenção das garantias de acesso ao ensino público gratuito.

A todos aqueles que, direta ou indiretamente, contribuíram para o desenvolvimento desta pesquisa.

Eu vi um menino correndo
Eu vi o tempo brincando ao redor
Do caminho daquele menino

Eu pus os meus pés no riacho
E acho que nunca os tirei
O sol ainda brilha na estrada e eu nunca passei

Eu vi a mulher preparando outra pessoa
O tempo parou pra eu olhar para aquela barriga
A vida é amiga da arte
É a parte que o sol me ensinou
O sol que atravessa essa estrada que nunca passou
[...]

Força Estranha, Caetano Veloso (1978)

RESUMO

Neste trabalho investiga-se a abordagem de avaliação da proficiência oral do professor de línguas estrangeiras empregado por um exame brasileiro voltado a essa finalidade específica, o EPPLÉ – Exame de Proficiência para Professores de Línguas Estrangeiras. O estudo tem por objetivo propor uma abordagem que atenda a necessidades levantadas pela literatura da área acerca da importância da democratização e transparência do processo avaliativo e do emprego de processos decisórios que garantam os princípios de validade sem causar interferências na praticidade das avaliações. Assim, a abordagem proposta subordina-se à elaboração de uma escala de mensuração e à reestruturação da escala de proficiência do exame, agregando-se a opinião de usuários com diferentes graus de especialidade sobre a avaliação de proficiência. Trata-se, portanto, de um estudo de caso de natureza híbrida e cunho exploratório-descritivo que se utiliza de dados secundários e primários para auxiliar a investigação. Os dados secundários consistem em estudos e documentos sobre a caracterização da proficiência linguística para finalidades gerais e para finalidades de ensino, bem como sobre a avaliação dessas proficiências; em dados de fala do teste oral do EPPLÉ referentes à participação de formandos em Letras de uma universidade pública paulista e de duas universidades públicas mineiras, na qualidade de candidatos ao exame, nos anos de 2015 e 2017; e nos respectivos conceitos atribuídos aos candidatos, seguindo-se os princípios determinados pela escala de proficiência oral original do exame. Os dados primários, por sua vez, consistem nas respostas obtidas por meio da aplicação sucessiva de dois questionários estruturados com base nos princípios do método de estruturação de problemas VFT – Pensamento Focado no Valor; o primeiro, com a finalidade de coletar informações a respeito do perfil dos participantes da pesquisa e de suas opiniões sobre as características da proficiência do professor de línguas e de sua avaliação e o segundo, para refinar e confirmar a opinião dos participantes, fornecendo embasamento para a elaboração da escala de mensuração. A determinação dos critérios que integram a escala de mensuração proposta, bem como de seus pesos para a avaliação, foi realizada com base nos valores informados pelos usuários do exame, nos valores determinados pela literatura e na criterialidade encontrada por meio da análise das características de fala dos candidatos. Os perfis linguísticos que representam os níveis de variação da proficiência na escala de mensuração, por sua vez, foram definidos com base nos resultados fornecidos pela análise multivariada de agrupamento, mais especificamente, pelo método de Ward, desses desempenhos. A escala de proficiência foi elaborada considerando-se os quadros de referência internacional CEFR – Quadro Europeu Comum de Referência para Línguas e CETF – *Cambridge English Teaching Framework*. A reavaliação dos desempenhos de fala foi realizada considerando-se a escala de mensuração proposta combinada ao método multicritério de análise da decisão *Fuzzy TOPSIS* – Técnica de Ordenamento de Preferências por Similaridade à Solução Ideal. A comparação do nível de proficiência sugerido pela aplicação do método em relação à escala de proficiência proposta com base nos quadros de referência mostrou-se confiável e, em relação à avaliação original, permitiu o tratamento de discrepâncias.

Palavras-chave: Avaliação. Proficiência oral. Proficiência do professor de línguas. Métodos de estruturação de problemas. Métodos multicritério de análise da decisão.

ABSTRACT

This research study investigates the assessment model used by a teaching-oriented language examination, the EPPLE – Proficiency Examination for Foreign Language Teachers, in the assessment of foreign language teachers' oral proficiency. We aim at evaluating the model used by the oral test and proposing a new one that meets current concerns raised by the literature regarding the importance of democratizing the assessment process, guaranteeing transparency and using decision models that assure validity principles without causing interferences in the test's practicality. Therefore, the model we propose relies on the reorganization of the oral proficiency scale used in the exam through the inclusion of its stakeholders' opinions. The investigation conducted consists on a quali-quantitative and exploratory-descriptive case study that uses primary and secondary data. Secondary data consist of research studies and documents on the concept of general and teaching purpose language proficiency; speaking data from candidates to the oral test of the EPPLE examination; as well as the respective rates assigned to such candidates according to the proficiency scale used in the exam. The candidates to the EPPLE examination that participate in this study are Letters course undergraduate students from Brazilian public universities, one located in the state of São Paulo and two in the state of Minas Gerais, in the years of 2015 and 2017. Primary data, on the other hand, consist of the answers obtained through the application of two questionnaires, which we structured following the principles of a problem structuring method, VFT – Value-Focused Thinking. We applied the first questionnaire aiming at collecting information on the research study participants' profiles and opinions concerning foreign language teachers' proficiency and its assessment. The second questionnaire supported the reorganization of the scale and was used to refine and confirm the participants' opinions. We selected the criteria included in the rubric scale, as well as its assessment weights, through the values of the stakeholders, the values pointed out in the literature and their criteriality, determined through the analysis of candidates' speech samples. Language profiles representing the variation of the levels of the proficiency in the rubric were defined through performance cluster analysis, more specifically through Ward's method. The proficiency scale was created through the principles presented in the CEFR – Common European Framework of Reference for Languages and in the CETF – Cambridge English Teaching Framework. Finally, the reassessment of the performances was conducted through the combination of the rubric and the multi-criteria decision analysis method, Fuzzy TOPSIS – Technique for Order of Preference by Similarity to Ideal Solution. The comparison of the proficiency level suggested by the application of the method based on the proficiency scale resulted in mainly reliable results, allowing the management of error measurements.

Keywords: Language testing. Oral proficiency. Problem Structuring Methods. Multicriteria Decision-Analysis methods. Teachers' oral proficiency

LISTA DE ILUSTRAÇÕES

Figuras

Figura 1. Ligações entre argumentos interpretativos.	47
Figura 2. Validade de construto da interpretação dos resultados.	48
Figura 3. Interação entre construto, critérios e o teste.	50
Figura 4. Etapas do processo de análise da decisão.	80
Figura 5. Modelo de matriz de decisão	85
Figura 6. Passos para aplicação do método TOPSIS clássico.	86
Figura 7. Passos para aplicação do método Fuzzy TOPSIS proposto por Chen (2000).	89
Figura 8. Definição dos pesos agregados e do número de critérios e de decisores.	108
Figura 9. Avaliação dos critérios em relação às alternativas.	108
Figura 10. Resultado da classificação sugerida.	108
Figura 11. Triangulação dos dados do estudo.	110
Figura 12. Competências e critérios relevantes para a atuação do professor de línguas segundo usuários do exame.	117
Figura 13. Competências e critérios relevantes para a atuação do professor de línguas segundo os envolvidos com a elaboração do exame.	120
Figura 14. Tarefa de apresentação pessoal da versão 3.0 do teste oral do exame EPPLE.	143
Figura 15. Tarefa de apresentação pessoal da versão 3.1 do teste oral do exame EPPLE.	143
Figura 16. Tarefa descrição de figuras da versão 3.1 do teste oral do exame EPPLE.	150
Figura 17. Tarefa de discussão de tópico da versão 3.0 do teste oral do exame EPPLE.	157
Figura 18. Tarefa de discussão de tópico da versão 3.1 do teste oral do exame EPPLE.	158
Figura 19. Rubrica da Tarefa 4 do teste oral do exame EPPLE (Versão 3.1).	163
Figura 20. Tarefa de metalinguagem da versão 3.0.1 do teste oral do exame EPPLE.	170
Figura 21. Tarefa de metalinguagem da versão 3.1 do teste oral do exame EPPLE.	171
Figura 22. Relação entre os componentes da competência linguístico-comunicativa avaliados.	180
Figura 23. Acento nuclear de DES.	183
Figura 24. Acento nuclear de ROB.	183

Quadros

Quadro 1. Atividades linguístico-comunicativas do CEFR.	56
Quadro 2. Estratégias linguístico-comunicativas do CEFR e criterialidade.	57
Quadro 3. Competências linguístico-comunicativas do CEFR e criterialidade.	58
Quadro 4. Criterialidade e tipos de erros gramaticais do CEFR.	60

Quadro 5. Quadro-resumo das principais características avaliativas dos exames internacionais.....	67
Quadro 6. Resultados dos exames em relação ao CEFR.....	68
Quadro 7. Etapas de construção da rede de objetivos meios e da hierarquia de objetivos fundamentais.....	83
Quadro 8. Quadro-resumo dos dados de fala.....	101
Quadro 9. Códigos empregado nas transcrições.....	104
Quadro 10. Relação entre as competências e os critérios do CEFR e do CETF e a opinião dos participantes.....	115
Quadro 11. Relação entre as competências e os critérios do CEFR e do CETF e a opinião dos participantes.....	118
Quadro 12. Resumo da importância dos critérios na opinião dos stakeholders.....	121
Quadro 13. Hierarquização dos objetivos do exame EPPLE.....	124
Quadro 14. Variáveis de uso e conhecimento sobre regras linguísticas da escala do exame EPPLE.....	135
Quadro 15. Hierarquização dos objetivos fundamentais segundo os princípios do VFT.....	136
Quadro 16. Definição dos atributos/critérios de mensuração.....	136
Quadro 17. Limite superior e inferior do atributo “clareza” para o objetivo “avaliar o fornecimento de informações sobre experiências”.....	137
Quadro 18. Limite superior e inferior dos atributos “velocidade” e “ritmo” para o objetivo “avaliar a fluência”.....	137
Quadro 19. Limite superior e inferior dos atributos “inteligibilidade”, “compreensibilidade” e “proximidade” para o objetivo “avaliar a pronúncia”.....	137
Quadro 20. Limite superior e inferior dos atributos “acurácia” e “complexidade” para o objetivo “avaliar o uso de estruturas gramaticais”.....	137
Quadro 21. Limite superior e inferior dos atributos “acurácia” e “variedade” para o objetivo “avaliar o uso de estruturas lexicais”.....	138
Quadro 22. Limite superior e inferior do atributo “adequação” para o objetivo “avaliar a expressão de conhecimento sobre regras linguísticas”.....	138
Quadro 23. Limite superior e inferior do atributo “clareza” para o objetivo “avaliar a seleção de informações para esclarecer dúvidas linguísticas”.....	138
Quadro 24. Limite superior e inferior dos atributos “adequação” e “compreensibilidade” para o objetivo “avaliar o fornecimento de explicação sobre regras linguísticas”.....	138
Quadro 25. Limite superior e inferior do atributo “adequação” para o objetivo “avaliar o uso de terminologia específica”.....	138
Quadro 26. Especificações das tarefas do teste oral do exame EPPLE (Versão 3.0 e 3.1).....	141
Quadro 27. Recursos terminológicos empregados no texto instrutivo.....	169

Quadro 28. Terminologia empregada na Situação 1 da Tarefa 3 do teste oral do EPPLE.....	178
Quadro 29. Terminologia empregada na Situação 2 da Tarefa 3 do teste oral do EPPLE.....	178
Quadro 30. Terminologia empregada na Situação A da Tarefa 5 do teste oral do EPPLE.....	178
Quadro 31. Terminologia empregada na Situação B da Tarefa 5 do teste oral do EPPLE.....	179
Quadro 32. Escala de mensuração para o teste oral do EPPLE.	197
Quadro 33. Escala de proficiência oral sugerida ao exame EPPLE com base em quadros de referências internacionais.....	200
Quadro 34. CEFR em relação a exames consagrados e ao EPPLE.....	204
Quadro 35. Definição dos pesos dos critérios de avaliação do teste oral do EPPLE.....	205
Quadro 36. Variáveis linguísticas associadas às faixas de proficiência.....	206
Quadro 37. Seleção de candidatos para reavaliação.	208
Quadro 38. Perfil de desempenho de LID.	209
Quadro 39. Perfil de desempenho de THY.	210
Quadro 40. Perfil de desempenho de DES.	212
Quadro 41. Perfil de desempenho de LIG.	212
Quadro 42. Perfil de desempenho de MAC.	212
Quadro 43. Perfil de desempenho de ROB.....	212
Quadro 44. Perfil de desempenho de LAC.....	212
Quadro 45. Quadro-resumo dos resultados da avaliação original e da reavaliação dos desempenhos de fala.	214

Gráficos

Gráfico 1. Competências da avaliação da proficiência do professor de línguas na opinião dos usuários do exame.....	123
Gráfico 2. Dendrograma da quantidade de funções comunicativas utilizadas na tarefa de apresentação pessoal.	146
Gráfico 3. Dendrograma da quantidade de emprego de tópicos expandidos na tarefa de apresentação pessoal.	147
Gráfico 4. Dendrograma do emprego de funções comunicativas na tarefa de descrição de figuras.....	153
Gráfico 5. Dendrograma da frequência do emprego de função comunicativa por nível de proficiência.....	154
Gráfico 6. Dendrograma do emprego de funções comunicativas na tarefa de discussão de tópico.....	160
Gráfico 7. Dendrograma da frequência de extensão de tópico na tarefa de discussão.....	161
Gráfico 8. Dendrograma do emprego de funções no fornecimento de instruções.....	166
Gráfico 9. Dendrograma do emprego de recurso terminológico na tarefa de instrução.....	168

Gráfico 10. Dendrograma do emprego de recursos metalinguísticos.....	174
Gráfico 11. Dendrograma da porcentagem de acurácia da explicação metalinguística.....	176
Gráfico 12. Dendrograma do desempenho de fala com relação ao acento nuclear.....	182
Gráfico 13. Dendrograma dos desempenhos de fala com relação à velocidade de fala.....	184
Gráfico 14. Dendrograma dos desempenhos de fala com relação ao alongamento de sons.....	185
Gráfico 15. Dendrograma dos desempenhos de fala com relação às pausas silenciosas.....	186
Gráfico 16. Dendrograma dos desempenhos de fala com relação às pausas preenchidas.....	187
Gráfico 17. Dendrograma dos desempenhos de fala com relação às repetições.....	188
Gráfico 18. Dendrograma dos desempenhos de fala com relação às reformulações.....	189
Gráfico 19. Dendrograma dos desempenhos de fala com relação à precisão linguística.....	190
Gráfico 20. Dendrograma dos desempenhos de fala com relação à medida global.....	191
Gráfico 21. Dendrograma dos desempenhos de fala com relação ao índice de coordenação.....	191
Gráfico 22. Dendrograma dos desempenhos de fala com relação ao Índice de Guiraud.....	192
Gráfico 23. Dendrograma dos desempenhos de fala com relação à frequência de palavras.....	193

LISTA DE TABELAS

Tabela 1. Tabela de contingência de conhecimento e participação no exame EPPLE.....	112
Tabela 2. Resumo descritivo do emprego mediano da quantidade de funções comunicativas abordadas.....	146
Tabela 3. Resumo descritivo do emprego mediano de tópicos expandidos.....	148
Tabela 4. Resumo descritivo do emprego mediano de funções comunicativas para descrição.....	153
Tabela 5. Resumo descritivo do emprego mediano de funções comunicativas para opinião.....	160
Tabela 6. Resumo descritivo do emprego mediano de organização do texto instrutivo.....	166
Tabela 7. Resumo descritivo do emprego mediano de variedade terminológica no texto instrutivo.....	168
Tabela 8. Resumo descritivo do emprego mediano de explicação metalinguística com acurácia.....	174
Tabela 9. Resumo descritivo do emprego mediano de explicação metalinguística com acurácia.....	177
Tabela 10. Resumo descritivo do emprego mediano de explicação metalinguística com acurácia.....	177
Tabela 11. Frequência do emprego de acento nuclear.....	182
Tabela 12. Resumo descritivo do emprego mediano de velocidade de fala.....	184
Tabela 13. Resumo descritivo do emprego mediano de alongamento de sons.....	185
Tabela 14. Resumo descritivo do emprego mediano de pausas silenciosas.....	186
Tabela 15. Resumo descritivo do emprego mediano de pausas preenchidas.....	187
Tabela 16. Resumo descritivo do emprego mediano de repetições.....	188
Tabela 17. Resumo descritivo do emprego mediano de reparos.....	189
Tabela 18. Resumo descritivo do emprego mediano de precisão linguística.....	190
Tabela 19. Resumo descritivo do emprego mediano de medida global.....	191
Tabela 20. Resumo descritivo do emprego mediano do índice de coordenação.....	192
Tabela 21. Resumo descritivo do emprego mediano de Índice de Guiraud.....	193
Tabela 22. Resumo descritivo do emprego mediano de frequência de palavras.....	194
Tabela 23. Mediana ou acima do limiar de A.....	207
Tabela 24. Limiar inferior de A.....	207
Tabela 25. Limiar superior de B.....	207
Tabela 26. Mediana de B.....	207
Tabela 27. Limiar inferior de B.....	207
Tabela 28. Limiar superior de C.....	207
Tabela 29. Mediana de C.....	207
Tabela 30. Limiar inferior de C.....	207
Tabela 31. Limiar superior de D.....	207
Tabela 32. Mediana ou inferior ao limiar de D.....	207
Tabela 33. Desempenho insuficiente.....	207

LISTA DE ABREVIACÕES

- AFT – *Alternative Focused Thinking*
- AHP – *Analytic Hierarchy Process*
- CAE – *Certificate of Advanced English*
- CEFR – *Common European Framework of Reference for Languages*
- CETF – *Cambridge English Teaching Framework*
- CLAN – *Computerized Language Analysis*
- CLC – Competência Linguístico-Comunicativa
- CPE – *Certificate of Proficiency in English*
- DT – *Drama Theory*
- EPPLÉ – Exame de Proficiência para Professores de Língua Estrangeira
- ETS – *English Testing Service*
- FCE – *First Certificate in English*
- IELTS – *International English Language Testing System*
- KET – *Key English Test*
- LE – Língua(s) Estrangeira(s)
- MCDA – *Multicriteria Decision Analysis* (Método Multicritério de Análise da Decisão)
- PLCPP – Proficiência Linguístico-Comunicativo-Pedagógica
- PSMs – *Problem Structuring Methods* (Métodos de Estruturação de Problemas)
- SCA – *Strategic Choice Approach*
- SODA – *Strategic Options Development and Analysis*
- SSM – *Soft Systems Method*
- TEPOLI – Teste de Proficiência Oral em Língua Inglesa
- TKT – *Teaching Knowledge Test*
- TOEFL iBT – *Test of English as a Foreign Language*
- TOPSIS – *Technique for Order of Preference by Similarity to Ideal Solution*
- VFT – *Value Focused Thinking*

SUMÁRIO

1 INTRODUÇÃO.....	25
1.1 A pesquisa	29
1.2 Organização da tese	33
2 CAPÍTULO I: FUNDAMENTAÇÃO TEÓRICA	35
2.1 Avaliação de proficiência de línguas: conceitos iniciais e desafios	35
2.1.1 Percurso histórico dos modos de avaliação da proficiência linguística: da era pré-científica à contemporaneidade	39
2.1.2 O desenvolvimento de exames de proficiência linguística: o modelo de utilidade do teste e suas qualidades	43
2.1.2.1 A avaliação da produção oral em língua estrangeira: o Quadro Comum de Referências para Línguas (CEFR) e os critérios de exames de proficiência consagrados	53
2.2 Avaliação da proficiência do professor de línguas estrangeiras: competências e perspectivas	68
2.2.1 O tratamento da linguagem pedagógica no Cambridge English Teaching Framework no exame de formação de professores de línguas TKT	73
2.3 A ciência da decisão: da neurociência à neuroeconomia	74
2.3.1 Estruturação de problemas e análise da decisão	77
2.3.1.1 Métodos de Estruturação de Problemas (PSMs)	81
2.3.1.2 Métodos Multicritério de Análise da Decisão (MCDA)	84
2.3.1.2.1 Técnica para Avaliar o Desempenho de Alternativas através de Similaridade com a Solução Ideal: TOPSIS e Fuzzy TOPSIS	85
2.4 A decisão na avaliação de proficiência em Linguística Aplicada: a possibilidade de tratamento interdisciplinar das discrepâncias e da incerteza	91
2.5 O capítulo em resumo.....	95
3 CAPÍTULO II: METODOLOGIA DA PESQUISA	97
3.1 Natureza da pesquisa	97
3.2 Contexto e participantes da pesquisa.....	98
3.2.1 O teste oral do exame EPPLÉ	98
3.2.2 Os candidatos ao exame EPPLÉ.....	100
3.2.3 Os usuários (<i>stakeholders</i>) especialistas do exame	102
3.3 Procedimentos e instrumentos de coleta e geração de dados	103
3.3.1 Transcrições das gravações em áudio do desempenho de fala de candidatos ao teste oral do exame EPPLÉ	103
3.3.2 Questionários para usuários (<i>stakeholders</i>) especialistas do exame EPPLÉ.....	105
3.4 Procedimentos e instrumentos de análise e triangulação de dados	105
3.5 O capítulo em resumo.....	110

4 CAPÍTULO III: DISCUSSÃO E ANÁLISE DE DADOS	111
4.1 Em busca de critérios para avaliação da proficiência oral do professor de línguas: os interesses dos usuários (<i>stakeholders</i>) do exame	111
4.1.1 As respostas fornecidas ao questionário de levantamento de critérios	111
4.1.2 As respostas fornecidas aos questionário sobre mensuração da proficiência	121
4.1.3 A estruturação do problema: considerações gerais sobre a opinião dos usuários.....	124
4.2 Avaliação da proficiência do professor de línguas pelo EPPLE.....	125
4.2.1 O construto informado e representado do exame: hierarquização dos objetivos da escala de proficiência oral.....	125
4.2.2 Definição dos critérios de avaliação do desempenho oral	135
4.2.3 Avaliação da proficiência oral no EPPLE: o desempenho dos candidatos nas tarefas de produção oral.....	139
4.2.3.1 Avaliação do uso da linguagem com finalidade geral no exame EPPLE	139
4.2.3.1.1 A tarefa de apresentação pessoal do teste oral do exame EPPLE.....	143
4.2.3.1.2 Avaliação do desempenho na apresentação pessoal no teste oral do exame EPPLE	145
4.2.3.1.3 A tarefa de descrição de figuras do exame EPPLE.....	150
4.2.3.1.4 Avaliação do desempenho na descrição de figuras no teste oral do exame EPPLE	152
4.2.3.1.5 A tarefa de discussão de tópico do exame EPPLE.....	157
4.2.3.1.6 Avaliação do desempenho na discussão de tópico no teste oral do exame EPPLE	159
4.2.3.2 Avaliação do uso da linguagem com finalidade pedagógica no exame EPPLE.....	162
4.2.3.2.1 A tarefa de fornecimento de instruções no teste oral do exame EPPLE.....	163
4.2.3.2.2 Avaliação do desempenho no fornecimento de instruções no teste oral do exame EPPLE.....	165
4.2.3.2.3 A tarefa de explicação gramatical do EPPLE	169
4.2.3.2.4 Avaliação do desempenho na explicação gramatical pela escala de proficiência oral do exame EPPLE	172
4.2.3.3 Avaliação das competências linguístico-comunicativas no exame EPPLE.....	180
4.3 Proposição de uma abordagem avaliativa da proficiência oral do professor de línguas para o EPPLE.....	194
4.3.1 A proposição de escalas para o teste oral do exame EPPLE	194
4.3.1.1 A escala de mensuração analítica para o teste oral do exame EPPLE	195
4.3.1.2 A escala de proficiência oral para o exame EPPLE.....	198
4.3.2 A avaliação da proficiência oral com auxílio do método multicritério de análise da decisão <i>Fuzzy TOPSIS</i>	204
4.3.3 Os resultados da reavaliação dos desempenhos de fala	208
4.4 O capítulo em resumo	215

5 CONSIDERAÇÕES FINAIS	217
REFERÊNCIAS	221
APÊNDICES	234
APÊNDICE A – Características distintivas da escala de proficiência oral do EPPLE	234
APÊNDICE B - Questionário de levantamento de interesses sobre o exame EPPLE I.....	236
APÊNDICE C - Questionário de levantamento de interesses sobre o exame EPPLE II.....	245
ANEXOS	259
ANEXO A - Escala Global do CEFR.....	259
ANEXO B - A escala do IELTS.....	260
ANEXO C - Escala de proficiência oral do EPPLE.....	261

1 INTRODUÇÃO

O ato de avaliar consiste em uma característica historicamente universal da vida em sociedade (McNAMARA, 2000) e, sendo uma atividade inerente ao ser humano, permeia o cotidiano em diversos níveis de complexidade (MESSICK, 1994; SHOHAMY, 2001). Na Linguística Aplicada, a avaliação é entendida como uma construção social (McNAMARA; ROEVER, 2006), apresentando, basicamente, três grandes propósitos: avaliação do rendimento escolar de indivíduos (*assessment*), testagem de capacidades (*testing*) e avaliação de qualidade educacional (*evaluation*) (FULCHER; DAVIDSON, 2007). A avaliação escolar tem finalidades formativas e seus resultados são provenientes da interação existente entre os elementos do tripé educacional ensino-aprendizagem-avaliação ao longo do processo de instrução. Basicamente, essa forma de avaliar consiste na coleta e análise de informações a respeito do rendimento do aprendizado de alunos por meio de instrumentos de avaliação diversos, como provas, portfólios, avaliação entre pares, autoavaliação, dentre outros (GENESEEE, 2001). A testagem, por sua vez, não tem por foco o processo instrutivo e, portanto, preocupa-se apenas com os resultados gerados pelo próprio instrumento de avaliação empregado. Desse modo, por meio da identificação e coleta de informações consideradas relevantes, visa a uma padronização das potenciais capacidades do examinado em relação a critérios previamente estabelecidos ou aos demais examinados (McNAMARA, 2000). Seus resultados, contudo, geram impactos que podem promover alterações relevantes nos cenários educacionais e sociais. A avaliação de proficiência linguística, objeto de estudo deste trabalho, consiste em um exemplo representante da categoria de testagem. Por fim, a avaliação de qualidade apoia-se em padrões estabelecidos para o processo educacional e pode ser empregada com a finalidade de prestação de contas sobre o processo instrucional; de nivelamento, promoção ou delimitação do status de um indivíduo em um curso ou programa; ou de orientação da instrução em sala de aula e consequente melhoria do processo de aprendizagem (GENESEEE, 2001).

Independentemente do propósito específico da avaliação, o ato de avaliar em si consiste em uma estratégia que visa à coleta, análise e interpretação de informações consideradas relevantes que possibilitem a realização de inferências para tomadas de decisão apropriadas (McNAMARA, 2000; GENESEEE, 2001). Na avaliação de proficiência linguística cabem aos avaliadores decidirem sobre os graus de adequação, representatividade e acurácia das amostras de língua fornecidas pelos candidatos e, por extensão, sobre a melhor faixa de proficiência que representa o potencial de uso da língua desse candidato com base no desempenho demonstrado

(McNAMARA, 2004); e, aos demais usuários (*stakeholders*) – como o próprio candidato, pesquisadores, líderes educacionais, formuladores de políticas educacionais, dentre outros –, as implicações sociais dos resultados apresentados (ELDER *et al.*, 2017). Assim, é importante ressaltar que o ato de avaliar, com seus propósitos e modelos, inevitavelmente, pressupõe tanto na etapa de avaliação *per se* quanto no período de pós-avaliação – aquele envolvendo o uso, pelos *stakeholders*, dos resultados fornecidos pelos examinadores – o ato de tomar decisões (KANE, 2002).

A decisão corresponde a uma atividade de posicionar-se em relação ao futuro (GOMES; GOMES; ALMEIDA, 2009), considerando-se o entendimento que se tem dessas projeções em um determinado momento político e sócio-histórico em relação a um ou mais critérios (BUEDE, 2013). No entanto, Munda (1993) destaca que sistemas sociais – aqueles nos quais os seres humanos têm um papel central – são marcados por imprecisão e incerteza devido a seu grande número de variáveis e de circuitos de retorno. Nesse sentido, quanto maiores os níveis de complexidade de um sistema, mais os conceitos de precisão e significância dos julgamentos realizados acerca de um comportamento tornam-se mutuamente excludentes (ZADEH, 1975), dificultando a tomada de decisões.

Para Clemen e Reilly (2013), a dificuldade inerente a algumas decisões está intimamente relacionada ao grau de singularidade do problema para o qual se busca uma solução. O problema consiste em uma oportunidade de decisão e estabelece-se quando se tem um objetivo em mente, mas não se sabe como atingi-lo (FISCHER; GREIFF; FUNKE, 2012). Desse modo, inicialmente, poder-se-ia pensar que a experiência com o tipo de problema seria suficiente para resolver a questão. No entanto, Clemen e Reilly (2013) ressaltam que, na verdade, o conhecimento proporcionado pela experiência acaba apenas por permitir o reconhecimento da dificuldade em problemas similares e a compreensão de como o problema consegue se encaixar em um cenário mais amplo para que a escolha das ferramentas apropriadas possa ser realizada. A dificuldade, desse modo, consiste não apenas na seleção dos elementos que compõem o processo, mas na tentativa de garantir que todos os elementos importantes sejam analisados. Assim, os autores destacam quatro fatores como potenciais geradores de dificuldade decisória: a complexidade do objeto, a incerteza sobre as informações, os múltiplos objetivos da decisão e os pontos de vista conflitantes por parte dos decisores.

A complexidade estabelece-se na relação de interdependência existente entre diversas características que compõem um objeto, implicando uma dificuldade de definição e consideração apropriada de todos os componentes críticos que integram problemas autóctones na análise da decisão diante da presença constante de elementos de ordem dinâmica e difusa

(MUNDA, 1993; CLEMEN; REILLY, 2013). A incerteza, por sua vez, consiste na ausência de informações desejadas para a decisão e, quando reflete probabilidades já conhecidas, por não permitir que o decisor visualize como os fatores que não estão sob seu controle irão se comportar, colocam a efetividade do processo decisório em risco (SMITH; HUETTEL, 2010; CLEMEN; REILLY, 2013). O trabalho com múltiplos objetivos, ante à limitação cognitiva dos seres humanos, pode fazer com que o foco em um objetivo específico acabe por impedir o alcance dos demais (CLEMEN; REILLY, 2013) e os pontos de vista conflitantes, por fim, podem ocorrer diante de alterações no insumo fornecido à compreensão do problema ou, simplesmente, por discordância entre os decisores a respeito dos valores subjetivos dos componentes, incluindo-se, aqui, fatores de interação social, como cooperativismo e honestidade (SMITH; HUETTEL, 2010; CLEMEN; REILLY, 2013). Por subjetividade, entenda-se a importância garantida aos critérios por cada um dos decisores envolvidos no processo, a qual é influenciada, especialmente, pelos graus de incerteza e de imprecisão de dados sobre as alternativas disponíveis (WĄTRÓBSKI *et al.*, 2018).

Na avaliação de línguas, a intenção de uma decisão deve ser a de maximizar o ideal de uma sociedade democrática (FULCHER; DAVIDSON, 2007). Assim, a avaliação de proficiência linguística é considerada uma atividade relevante em termos políticos, sociais e educacionais para a sociedade contemporânea, o que faz com que a Linguística Aplicada se preocupe constantemente com a clareza, confiabilidade e atualidade das informações nela e por meio dela veiculadas (SHOHAMY; OR, 2017). Tendo em vista que o objetivo da avaliação de proficiência linguística é realizar inferências acerca dos potenciais desempenhos que o candidato será capaz de realizar na vida real com base em amostras de desempenho que podem ser observadas em um exame ou teste (McNAMARA, 2000), entende-se que o verdadeiro objeto de interesse da avaliação de proficiência linguística consiste em um comportamento que não pode ser diretamente observado pelo avaliador, apenas inferido por meio dos construtos que embasam o referido instrumento (McNAMARA, 2004).

Diante disso, McNamara (2000) ressalta que o processo avaliativo é naturalmente limitado e, portanto, sujeito a erros, seja por questões de ordem contextual ou interpretativa, ou, ainda, por fatores afetivos gerados pela interferência do ato de observar no desempenho do candidato. Faz-se necessário, portanto, que a autenticidade dos testes e, por extensão, a modelagem realizada do construto sejam gerenciadas, de modo a minimizar interferências na possibilidade de generalização dos resultados (McNAMARA, 2004). Assim, é imprescindível ressaltar a inerência da subjetividade e da incerteza na avaliação de testes com objetos complexos, como é o caso da avaliação de proficiência linguística. A complexidade do objeto

acaba dificultando a capacidade dos elaboradores de delimitarem a dinamicidade interacional de seus componentes, fazendo com que os critérios sejam de difícil mensuração e, portanto, podendo gerar discrepâncias (BACHMAN; LYNCH; MASON, 1995). Ou seja, os resultados de uma avaliação de proficiência estão sujeitos a discrepâncias por configurarem uma decisão tomada subjetivamente com base em amostras de língua, as quais consistem em variáveis complexas e difusas, elicitadas por meio de respostas abertas sobre as quais são realizados inferências e julgamentos mais ou menos influenciados pelas opiniões, experiências e percepções do decisor a respeito do construto do exame e da amostra (HAMP-LYONS, 1992; ANG-AW; GOH, 2011).

Discrepâncias consistem em variabilidades no julgamento da qualidade do desempenho do candidato entre os avaliadores e, geralmente, são geradas pelo nível de severidade ou leniência desses avaliadores, pelo tempo de experiência com ensino de línguas e avaliação que eles apresentam, pelo grau de consistência no entendimento e emprego da escala solicitada, pela eficácia do treinamento realizado, dentre outros fatores (ANG-AW; GOH, 2011; TOFFOLI; SIMON, 2018). Ademais, Pollitt e Murray (1996) e Brown (2000) ressaltam que os avaliadores costumam diferir com relação às próprias abordagens avaliativas que empregam. De maneira geral, esses avaliadores podem formar um conceito holístico nos primeiros minutos acerca dos avaliados de modo que, na sequência, passem a buscar evidências de desempenho que comprovem ou refutem suas impressões iniciais.

Assim, apesar de serem inerentes ao processo decisório, especialmente na presença de múltiplos critérios e múltiplos decisores, as discrepâncias podem inviabilizar a obtenção dos resultados da avaliação ou, ainda, gerar um emprego desnecessário de recursos para sua solução. Buede (2013) destaca que o processo de tomar decisões envolve o dispêndio e a alocação de recursos visando a uma busca por mudança ou manutenção de um *status quo*. Independentemente da natureza desses recursos – financeira, temporal, cognitiva, dentre outros –, faz-se necessário apontar que um pensamento mais estruturado a respeito do problema decisório é imprescindível para sua melhor compreensão e, conseqüentemente, para a economia dos recursos empregados em sua resolução e para a obtenção de um resultado mais confiável (CLEMEN; REILLY, 2013).

No caso da proficiência¹ do professor de línguas, a complexidade é ainda maior. Isso porque inúmeras são as competências que devem ser desenvolvidas e mobilizadas por esses profissionais para que o processo de ensino seja bem-sucedido (CONSOLO; TEIXEIRA DA

¹ O termo “proficiência” empregado neste trabalho relaciona-se estritamente à proficiência linguística, mesmo quando empregado sem o acompanhamento de seu modificador.

SILVA, 2007a; CONSOLO; PORTO, 2011). Dentre os principais problemas relacionados à sua avaliação, encontram-se a delimitação do domínio de uso da língua para finalidades específicas, a autenticidade das tarefas e das amostras elicitadas e a extensão da influência de fatores extralinguísticos, como competência estratégica e habilidades de ensino, nessas amostras (DOUGLAS, 2000, 2001; ELDER, 2001). Do mesmo modo, a interpretação arbitrária e insuficiente de algumas abordagens de ensino de línguas, especialmente da abordagem comunicativa, em uma tentativa de negar a tradicional abordagem gramatical, faz com que professores defendam a prevalência de uma ou outra competência sobre a competência linguística. Assim, o elemento linguístico acaba sendo visto como “vilão” no processo de ensino-aprendizagem de línguas, como se o trabalho com a língua que se ensina fosse elemento de não-criticidade, contradizendo os próprios princípios, a própria concepção de língua da referida abordagem, para a qual a língua é um elemento essencial por meio do qual a comunicação é instaurada (SCHULZ; CUSTÓDIO; VIAPIANA, 2012), e fazendo com que exames de proficiência de línguas sejam evitados e mal interpretados (FULCHER; DAVIDSON, 2007).

Scaramucci (2000) ressalta que a falta de objetivos reais e específicos e de critérios claros de orientação do processo contribuem para reforçar a subjetividade da avaliação e, portanto, torná-la injusta. Scaramucci (2000) e O’Sullivan (2011) destacam, assim, a importância de garantir um processo transparente e participativo, ou seja, que integre, em sua base desenvolvimental, a perspectiva dos candidatos a respeito dos parâmetros contextuais, dos processos cognitivos e dos recursos empregados.

Diante desse cenário, no qual as limitações cognitivas humanas podem interferir de maneira significativa no processamento de informações e impactar a aceitação social de avaliações de proficiência, faz-se necessário buscar abordagens que permitam uma modelagem estruturada dos problemas, ou seja, que considerem as opiniões pessoais do avaliador no processo decisório, uma vez que a sistematicidade do pensamento pode auxiliar o desenvolvimento apropriado do julgamento subjetivo (CLEMEN; REILLY, 2013). Em outras palavras, deve-se buscar uma abordagem que não almeje a eliminação da subjetividade do processo decisório, mas, sim, sua representação de modo compreensível e passível de tratamento.

1.1 A pesquisa

Esta pesquisa insere-se no contexto de investigações sobre a avaliação da proficiência dos professores de línguas realizadas pelo grupo de pesquisa ENAPLE-CCC – Ensino

e Aprendizagem de Língua Estrangeira: Crenças, Construtos e Competências (CNPq/UNESP). Historicamente, as buscas pela definição da proficiência do professor de línguas e pelas estratégias de mensuração de seus componentes iniciaram-se por meio da apresentação, no ano 2000, do TEPOLI – Teste de Proficiência Oral em Língua Inglesa, desenvolvido, exclusivamente para avaliar a proficiência do professor de línguas em sua modalidade oral e suscitado pelo questionamento acerca da definição da proficiência do professor de línguas (CONSOLO, 2004; CONSOLO; TEIXEIRA DA SILVA, 2007b). A primeira aplicação do TEPOLI ocorreu a alunos formandos de um curso de Letras no ano de 2001 e o uso desse teste estendeu-se até o ano de 2007 para os referidos alunos de cursos de licenciatura em Letras e para professores em atuação.

Em 2008, foi proposto o exame EPPLE – Exame de Proficiência para Professores de Língua Estrangeira, que avalia a proficiência do professor de língua estrangeira na língua em que ensina em suas modalidades oral e escrita (CONSOLO, 2008; CONSOLO, 2011a, 2011b). O exame surgiu no formato de entrevista em pares face-a-face, da qual participavam dois examinadores – um responsável pela interação e outro examinador-observador. Em 2011, evoluiu para o formato eletrônico. Neste formato, a interação dá-se exclusivamente com o computador por meio de perguntas e respostas e avaliação passa a ser realizada *a posteriori*, analisando-se a interação gravada por dois avaliadores – na presença de discrepância, um terceiro avaliador é solicitado e, assim, sucessivamente até a obtenção de uma faixa em consenso. A partir de 2012, as aplicações do exame passaram a ser realizadas em ambos os formatos. No entanto, apesar de a abordagem avaliativa do exame ser amplamente utilizada na avaliação de proficiência oral (MAROCHI, 2008), tem-se que ela (i.) é pouco prática, uma vez que o resultado final é dependente da disponibilidade de ao menos duas pessoas para avaliação, e (ii.) apresenta validade de resultados questionável, já que a ausência de pesos e de atributos bem definidos para os critérios acaba por ampliar a subjetividade dos julgamentos, deixando-os sujeitos à interpretação dos avaliadores.

Esta pesquisa dedica-se ao estudo da abordagem avaliativa da proficiência oral de professores de inglês como língua estrangeira empregado por um exame de finalidades específicas, o EPPLE, visando a adequá-lo às exigências contextualmente democráticas impostas pelo momento sócio-histórico e, mais especificamente, por seus usuários (*stakeholders*), garantindo maior transparência e apoio às decisões avaliativas. Assim, tem por objetivo geral reestruturar a abordagem de avaliação e, conseqüentemente, de tomada de decisão para classificação do candidato em faixas de proficiência oral de um exame que tem

por objetivo a avaliação da proficiência em língua estrangeira de professores de línguas. Para tanto, faz uso de métodos de estruturação de problemas e de métodos de apoio à decisão.

Os objetivos específicos são (i) propor uma escala de proficiência oral que incorpore os interesses dos usuários (*stakeholders*) do exame, bem como as características de fala dos candidatos e das tarefas, e que esteja alinhada a indicadores de proficiência linguística internacionais e (ii) propor uma estrutura avaliativa que potencialize a precisão, a transparência, a praticidade e a confiabilidade dos resultados.

Assim, visa responder às seguintes perguntas de pesquisa:

1. Quais critérios devem integrar a avaliação da proficiência linguístico-comunicativo-pedagógica do professor de línguas?
2. Quais critérios da proficiência do professor de línguas são passíveis de mensuração – quantitativa ou qualitativamente?
3. Como mensurar a proficiência linguístico-comunicativo-pedagógica do professor de línguas, garantindo os princípios de validade, confiabilidade e praticidade das avaliações e agregando os valores do grupo de usuários (*stakeholders*) do exame sobre os critérios que a integram?

A escolha pelo trabalho com a avaliação da proficiência linguístico-comunicativo-pedagógica justifica-se, neste estudo, pelo fato de esse conceito de proficiência tratar-se de uma noção recente nos estudos em Linguística Aplicada (CONSOLO, 2011a, 2011b; CONSOLO; TEIXEIRA DA SILVA, 2014, 2016), carecendo, ainda, de maior compreensão. A justificativa estende-se para o fato de a avaliação dessa proficiência, conseqüentemente, impor, por sua natureza, inúmeras dificuldades decisórias. O exame EPPLE, por sua vez, foi escolhido por ser o único exame de bases comunicativas que trabalha com a proficiência em língua estrangeira do professor de línguas (CONSOLO, 2011a, 2011b). Já a modalidade oral da linguagem, por sua natureza dinâmica, aberta a níveis maiores de incerteza interpretativa (PLOUGH; BRIGGS; BONN, 2010), além de ser considerada essencial para a efetividade do processo de ensino de línguas (CONSOLO, 2004; CHAMBLESS, 2012).

Além disso, diante do fato de os resultados apresentados pelo EPPLE poderem ser utilizados para decidir se um professor está apto ou não a ministrar determinadas aulas, informando-lhe e a seu empregador sobre a qualidade e sobre os limites de sua proficiência, faz-se necessário que esses resultados apresentem altos índices de validade e, por extensão, de confiabilidade. Desse modo, na avaliação de proficiência e, mais especificamente, no EPPLE, há a necessidade de se minimizar os impactos da discrepância, ou seja, de adequar a interpretação dos resultados do exame ao seu construto e gerenciar a inerente influência da

subjetividade na tomada de decisão, tornando o processo avaliativo mais estruturado e compreensível. Apesar de ainda não existirem estudos sobre a presença e o gerenciamento das discrepâncias no exame EPPLÉ, especificamente, tal necessidade mostra-se fundamental em virtude do fato de ele consistir em uma avaliação de potencial alto impacto (KANE, 2002; KUNNAN, 2012), desempenhando um papel decisivo na determinação da vida dos candidatos (SCARAMUCCI, 2011). Ademais, apesar de haver técnicas estatísticas clássicas indicadas e reconhecidas para a avaliação dos níveis de concordância entre avaliadores em diferentes momentos e finalidades avaliativas, garantindo, assim, maior confiabilidade ao processo, é imprescindível ressaltar que tais técnicas trabalham com conceitos de diferença e relação, os quais não são necessariamente sinônimos de concordância e podem gerar falsas interpretações (RANGANATHAN; PRAMESH; AGGARWAL, 2017). Do mesmo modo, é fundamental destacar que, por ser um exame em fase de implementação, o exame EPPLÉ costuma contar com uma quantidade de amostras limitada, comprometendo o tamanho da amostra mínimo requerido – igual ou superior a 15 amostras – para a aplicação da maioria dos testes estatísticos paramétricos.

Ademais, diante do caráter atual pouco institucionalizado do EPPLÉ e da pretensão de se tornar um exame de abrangência nacional, o presente estudo traz contribuições ao promover uma oportunidade de remodelagem dos componentes decisórios do exame, propondo uma abordagem mais prática e mais consistente de decisão, e possibilitando sua reavaliação e sua adequação às exigências sociais por transparência e democratização. A modelagem estruturada proposta neste estudo justifica-se, então, como uma tentativa de promover uma maior compreensão do construto, na medida em que engloba elementos que são considerados relevantes por uma ampla gama de usuários (*stakeholders*) do exame. Igualmente, é necessário ressaltar que, por meio da metodologia empregada neste estudo, contribui-se, ainda, para a geração de uma abordagem decisória estruturada que pode servir de apoio inicial à futura elaboração de um sistema automatizado de avaliação. É importante mencionar, contudo, que a abordagem avaliativa proposta por meio deste trabalho não tem a pretensão de eliminar a subjetividade do processo avaliativo, mas, sim, orientar os procedimentos de avaliação dos dados dos examinados no EPPLÉ.

Assim, além da preservação, em termos éticos, da face do candidato, faz-se possível também preservar a face do próprio exame, ampliando sua validade e podendo, inclusive, vir a ampliar sua aceitação. Ressalta-se, por fim, que um melhor entendimento do conceito de proficiência linguístico-comunicativo-pedagógica e de sua avaliação faz-se crucial para cursos

de formação de professores de línguas, possibilitando a criação de estratégias de ensino para promover seu desenvolvimento nos docentes em formação.

1.2 Organização da tese

Este trabalho encontra-se organizado em cinco partes. Esta seção introdutória teve por objetivo apresentar o tema da pesquisa – a saber, a proficiência linguístico-comunicativo-pedagógica do professor de língua estrangeira, especificamente da língua inglesa –, juntamente com o objeto da investigação realizada – a avaliação da proficiência oral do professor de línguas por um exame de finalidades específicas. Diante disso, abordou-se o conceito de avaliar por meio do viés da realização de julgamentos e inferências em paralelo ao conceito de tomada de decisões. A dificuldade e a subjetividade inerentes aos processos decisórios e, por extensão, avaliativos também foram apresentadas, especialmente quando relacionadas a questões de complexidade, incertezas, múltiplos objetivos e decisões em grupo – situações essas que permeiam o ambiente de avaliação de proficiência linguística, especialmente em sua modalidade oral, e que possibilitam questionamentos sobre a confiabilidade desses instrumentos. Na sequência, foram introduzidos os objetivos e as perguntas de pesquisa que orientam o desenvolvimento da investigação, bem como as justificativas para a escolha do objeto de pesquisa.

O primeiro capítulo, referente à fundamentação teórica, visa, como o próprio título sugere, apresentar as teorias que embasam e orientam a investigação. Inicialmente, compõem o capítulo considerações acerca das abordagens de avaliação e de desenvolvimento e operacionalização de avaliações de proficiência linguística ao longo da história, considerando-se suas naturezas, suas finalidades, seus construtos e seus usos, bem como os critérios empregados em avaliações de proficiência consagradas e sugestões da literatura. Na sequência, são discutidos os conceitos e elementos integrantes da proficiência do professor de línguas para ensinar, bem como seus princípios avaliativos. Por fim, é introduzida a noção de análise da decisão, com a apresentação dos métodos de estruturação de problemas e de análise da decisão multicritério, além de considerações acerca das interfaces existentes entre a Linguística Aplicada e a ciência de Teoria da Decisão.

O segundo capítulo, referente à metodologia da investigação, tem por objetivo prover informações a respeito dos contextos e dos participantes da pesquisa, bem como dos instrumentos e procedimentos de coleta, análise e triangulação dos dados. Ressalta-se que, apesar de entender a fundamentação teórica como uma etapa da metodologia da investigação, este trabalho optou por respeitar a ordem canônica de estruturação de trabalhos acadêmicos e

apresentar a fundamentação teórica anteriormente à metodologia diante da necessidade de se explicarem os princípios que embasam os métodos de estruturação de problemas e de análise da decisão multicritério empregados.

O terceiro capítulo comporta a apresentação da análise e discussão dos dados à luz das teorias previamente apresentadas. Primeiramente, são exibidas as respostas obtidas por meio dos questionários aplicados aos participantes da pesquisa, especialistas diretos ou indiretos da avaliação de proficiência linguística de professores, com o objetivo de elicitare os valores e objetivos que foram, por eles, considerados relevantes para integrar a referida avaliação. Na sequência, são expostas as análises conduzidas a respeito da estrutura do exame EPPLÉ quanto a seus critérios, objetivos e alternativas – levantados, especialmente, por meio de informações extraídas de textos de divulgação do exame, bem como de seus insumos, de suas tarefas, de sua escala de proficiência e de sua abordagem avaliativa –, de modo a se obter um panorama dos valores por ele representados. Em seguida, apresentam-se os resultados da análise das características de fala presentes nos textos do corpus composto pelos candidatos ao exame nos anos de 2015 e 2017.

O capítulo desenvolve-se para a proposta de uma abordagem avaliativa ao exame EPPLÉ. Inicialmente, expõem-se os procedimentos sugeridos para a preparação dos dados de fala dos candidatos para avaliação. Segue-se, a isso, a proposição de uma escala de mensuração de bases analíticas, com a finalidade de orientar as análises de desempenho, e de uma escala híbrida – holística e analítica – de proficiência visando à divulgação dos resultados. Apresentam-se e discutem-se, por fim, os resultados da avaliação dos dados de fala pelo método multicritério de análise da decisão selecionado em relação ao resultado por ele fornecido e à escala de proficiência proposta.

O trabalho é finalizado com uma seção destinada às considerações finais, em que são sintetizados os resultados alcançados pela investigação, apresentadas as contribuições e as limitações do trabalho, e sugeridos encaminhamentos para pesquisas futuras.

2 CAPÍTULO I: FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são apresentadas as teorias que embasam e orientam a investigação proposta. O capítulo encontra-se subdividido em três seções. Na primeira, são abordados conceitos e procedimentos empregados na construção, implementação e nos usos de modelos de avaliações de proficiência ao longo da história, com enfoque especial nos exames e quadros de referência consagrados. Assim, são tecidas considerações acerca dos instrumentos avaliativos com relação à natureza; às abordagens de língua, de proficiência linguística e do próprio ato de avaliar; aos propósitos e às finalidades; à estruturação; aos impactos por eles gerados, bem como às implicações éticas e legais do processo.

A segunda seção aborda os conceitos e elementos da proficiência linguística do professor de línguas e perpassa, portanto, a introdução do conceito de proficiência linguístico-comunicativa na literatura até o surgimento do conceito de proficiência linguístico-comunicativo-pedagógica, que acompanhou o desenvolvimento do exame EPPLE. A seção comporta, ainda, os conceitos e as abordagens conhecidas de avaliação dessa proficiência no cenário internacional.

A última seção introduz a noção de análise da decisão, com a apresentação de métodos de estruturação de problemas e de análise da decisão multicritério em grupo, os quais compõem a metodologia empregada neste estudo. Além disso, apresenta uma discussão a respeito do tratamento da incerteza nos estudos sobre avaliação da proficiência linguística, destacando as contribuições que a área de Teoria da Decisão pode trazer aos estudos sobre avaliação de proficiência na Linguística Aplicada.

2.1 Avaliação de proficiência de línguas: conceitos iniciais e desafios

A definição do conceito de proficiência linguística é fundamental para os estudos sobre avaliação em Linguística Aplicada (IWASHITA *et al.*, 2008; CONSOLO, 2011a, 2011b), dado que ela se encontra diretamente relacionada à visão de língua daquele que a define (SCARAMUCCI, 2000). No Brasil, na última década, o conceito técnico de proficiência proposto por Taylor (1988) tem sido amplamente aceito pelos pesquisadores da área (BORGES-ALMEIDA, 2009; BORGES-ALMEIDA; CONSOLO, 2011; CONSOLO, 2011a, 2011b; BRAGA, 2013; CARVALHO E SILVA, 2016; CONSOLO, 2004; CONSOLO; TEIXEIRA DA SILVA, 2016; DIAS, 2011; MAROCHI, 2008; SCHOFFEN, 2009). Em seu sentido técnico, a proficiência deve ser entendida como a capacidade de usar a língua para uma

finalidade específica, podendo ser subdividida em níveis de competência. Assim, tem-se que, ao se questionar a qualidade da proficiência de um indivíduo, deve-se levar em consideração as finalidades comunicativas para as quais essa proficiência é requerida, existindo a possibilidade de um falante ser linguisticamente proficiente para desempenhar determinados tipos de tarefas e não outros. Essa definição surge diante da necessidade de se considerar aspectos de cunho comunicativo da língua e em oposição ao conceito binário de proficiência conhecido e difundido pelo senso comum, também denominado conceito não-técnico, que considera apenas que um usuário possa ou não ser proficiente de maneira absoluta.

Apesar de fundamental para a Linguística Aplicada e, mais especificamente, à avaliação, o conceito de proficiência linguística é bastante controverso (LLURDA, 2000) e comumente confundido com as noções de competência (*competence*), capacidade (*ability*), habilidade (*skill*) e fluência (*fluency*). Embora as noções sejam complementares, há uma hierarquia semântica entre elas que deve ser considerada para uma compreensão mais adequada de suas relações. De modo geral, na definição de proficiência de sentido técnico proposta por Taylor (1988), as competências integram dimensões de ordem linguística, comunicativa, dentre outras, a depender do conceito de língua que se considera e, independentemente da orientação, são definidas por habilidades, como “usar”, “informar”, “narrar”, “descrever”, “comparar”. Deste modo, tem-se que a habilidade linguística se constitui em uma aptidão relacionada às diferentes modalidades da língua, mais especificamente, à oralidade e à escrita, que materializa o uso do conhecimento da língua na comunicação real (CANALE; SWAIN, 1980). Assim, tradicionalmente, considera-se a existência de quatro macro-habilidades: a fala, a escrita, a escuta e a leitura, sendo as duas primeiras relacionadas à capacidade de produção e as duas últimas, à compreensão. A capacidade, por sua vez, consiste no “saber fazer algo”, podendo mobilizar habilidades específicas ou a própria competência como um todo (BARBOSA, 2007). De maneira simplificada, a proficiência consiste, então, na capacidade de uso de um conjunto de competências, que são definidas por um conjunto de habilidades.

Por fim, tem-se que os descritores da proficiência, ou seja, seus qualitativos devem ser definidos a partir das necessidades de comunicação para a situação na qual se pretende fazer uso da língua-alvo (SCARAMUCCI, 2000). A fluência corresponde, assim, a um dos elementos que compõem a proficiência e, por extensão, uma competência específica, ou seja, a um dos elementos que compõem os descritores dessa proficiência (TEIXEIRA DA SILVA, 2000), sendo também definida por meio de habilidades.

Os conceitos de competência e desempenho são, assim, essenciais para os estudos de avaliação de proficiência. Na década de 1960, após a apresentação por Chomsky (1965) dos

conceitos de competência e desempenho – segundo o qual a competência corresponde ao conhecimento linguístico inato do falante e o desempenho, à performance durante o ato de fala resultante da competência –, Hymes (1966) introduziu na Linguística a noção de competência comunicativa. No modelo proposto, o autor considerava-a composta por competências de ordem (i.) linguística, que condensa elementos gramaticais, lexicais e ortográficos; (ii.) discursiva, representativa da capacidade de ordenação textual de modo coeso e coerente; (iii.) sociolinguística, relativa ao conhecimento das regras socioculturais de uso da língua; e (iv.) estratégica, que representa a capacidade de reconhecer e realizar reparos comunicacionais sempre que necessário.

O trabalho de Hymes (1966) foi pioneiro ao considerar, em oposição ao conceito estático de competência chomskyano, a importância do contexto nos estudos linguísticos. Contudo, apesar do pioneirismo, são Canale e Swain (1980) os responsáveis por uma das definições de competência comunicativa mais discutidas nos estudos de Linguística Aplicada, apresentando-a como um conceito focado em funções comunicativas, no qual as formas gramaticais seriam empregadas como meios de expressão apropriados das funções.

Para Canale e Swain (1980), a competência comunicativa representa o conhecimento subjacente às funções comunicativas, bem como a capacidade de uso desse conhecimento. Desse modo, a noção é composta por competências de ordem (i.) gramatical, envolvendo conhecimento lexical e de regras de uso da morfologia, sintaxe e semântica; (ii.) sociolinguística, que compreende as regras de uso da língua em relação ao contexto e as regras discursivas de coesão e coerência – esta última que, posteriormente, se desdobraria em uma competência independente: a competência discursiva; e (iii.) comunicativa propriamente dita ou, mais especificamente, a competência estratégica, que inclui elementos tanto de cunho verbal – em correlação com as outras duas competências – e de cunho não-verbal, ou seja, sinestésicos e paralinguísticos. O modelo de Canale e Swain (1980), apesar de bastante utilizado, recebeu críticas por seu caráter descritivo, ou seja, por não ter sido elaborado com base em estudos de ordem empírica (WIDDOWSON, 2001), tendo originado várias tentativas de validação.

Uma década depois, Bachman (1990) foi o responsável pela proposição de um dos modelos de competência comunicativa mais aceitos dada sua complexidade (AMIRIAN; MOQADDAM; MOQADDAM, 2017). No modelo, intitulado “componentes da competência linguística”, a competência linguística condensa-se em competência organizacional, composta pelas competências gramatical e textual; e competência pragmática, composta pelas competências ilocucionária e sociolinguística. Apesar de inovador, o modelo de Bachman (1990) dialoga com as competências apresentadas em Canale e Swain (1980), especialmente no

que tange à equivalência das noções de competência gramatical e à condensação das competências estratégica e discursiva destes na definição do conceito de competência textual daquele. É válido ressaltar que o termo “linguística” associado à competência apresentada por Bachman (1990) é de ordem comunicativa, ou seja, para o autor, linguístico é considerado como relativo à língua, e não, especificamente, como elemento de ordem estrutural, gramatical.

É também o modelo de Bachman (1990) proposto para caracterização do uso da língua para propósitos gerais que embasa a elaboração do conceito de competência linguístico-comunicativa desenvolvido por Almeida Filho (1992) para delimitar as características da linguagem docente, o qual, por sua vez, ilumina o surgimento do conceito de proficiência linguístico-comunicativo-pedagógica, foco de estudo deste trabalho. Discussões mais aprofundadas sobre as competências que integram a proficiência específica do professor de línguas são apresentadas na seção 2.2.

Além do conceito de proficiência e, por extensão, das noções de competência e desempenho, o entendimento dos objetos que integram o ato de avaliar a capacidade de uso da língua nos estudos sobre testagem também é crucial para a definição das características do teste. Nesse contexto, também existem controvérsias a respeito das definições de teste, prova, tarefa e exame. De maneira geral, testes consistem em instrumentos avaliativos da proficiência linguística do qual pelo menos uma das quatro habilidades da língua não participa e opõe-se ao conceito de exame, que é composto por todas as habilidades, de modo que um exame pode ser composto por um ou mais testes de habilidades específicas. Assim, apesar de utilizar a terminologia referente a testes em sua nomenclatura, o consagrado *Test of English as a Foreign Language* (TOEFL), ao menos em sua versão eletrônica (iBT), consiste, na verdade, em um exame. A tarefa, por sua vez, apresenta ao menos dois domínios: o do teste em si e o domínio-alvo de uso da língua. Com relação ao primeiro, a tarefa consiste em representações de atividades da vida real por meio das quais o desempenho do candidato é elicitado. No segundo, tarefas consistem em atividades desempenhadas na vida real em um contexto específico. A prova, por fim, consiste no objeto de avaliação empregado e condensa as tarefas selecionadas para o teste, ou seja, é o formato por meio do qual testes e exames são apresentados ao candidato.

Além da natureza conceitual, Bachman e Palmer (1996) destacam, por fim, alguns desafios para desenvolvedores de testes e avaliadores de proficiência linguística que podem gerar usos indevidos dos resultados fornecidos. São eles: a crença na existência de um teste ideal, a presença de expectativas irreais sobre o que o teste consegue fazer e sobre como ele deveria ser, e a confiança inaudita na tecnologia da mensuração. Tais ideias subordinam-se à

concepção que se tem a respeito do construto a ser mensurado e dos propósitos do instrumento avaliativo e serão abordados de maneira mais aprofundada no decorrer deste capítulo.

A avaliação de proficiência de línguas consiste em um ramo dos estudos de testagem da Linguística Aplicada que tem por finalidade a avaliação empírica de desempenho de indivíduos em relação a parâmetros previamente definidos. Cumpre, portanto, uma finalidade de seleção e controle institucional e sua elaboração é política e sócio-historicamente construída (McNAMARA, 2000). Devido a seu impacto social e a sua concepção tradicionalmente controversa, a avaliação de proficiência linguística acaba funcionando como um elemento de poder, de autoridade por parte daqueles que a empregam, mobilizando questões de dependência e medo por parte dos avaliados (SCARAMUCCI, 2000; McNAMARA, 2000).

Nesse sentido, McNamara (2000, p. 03) observa que o conhecimento especializado dos profissionais que trabalham com avaliação acaba sendo visto, nas palavras do autor, como algo “remoto” e “obscuro”. Desse modo, o autor ressalta que o ato de avaliar acaba constantemente evocando reações negativas de ansiedade e impotência por parte dos prováveis avaliados. As mudanças pelas quais a área de avaliação de línguas passou nas últimas décadas acabam por tentar minimizar os efeitos negativos gerados pelas formas de avaliação tradicionais. No entanto, McNamara (2000) destaca a importância de se investigar como ocorreu esse desenvolvimento e quais os princípios de desenho do teste que devem ser seguidos. Para tanto, as subseções a seguir apresentam, respectivamente, um panorama da história da avaliação e dos conceitos que embasam o desenvolvimento de instrumentos avaliativos.

2.1.1 Percurso histórico dos modos de avaliação da proficiência linguística: da era pré-científica à contemporaneidade

As primeiras práticas avaliativas da história da humanidade datam da época do Império Chinês e representam o surgimento de tentativas, de caráter estatal, de se estabelecer um sistema de avaliação controlado. Assim, em seus primórdios, as avaliações eram conduzidas com a intenção de selecionar candidatos, por mérito, a cargos remunerados no governo. Apesar de, inicialmente, não se encontrarem diretamente vinculados ao sistema educacional, tais testes eram extensos, complexos, compostos por conteúdos acadêmicos e fortemente respeitados e, desse modo, exerciam pressão indireta sobre a educação (SPOLSKY, 2017).

No ocidente, outro momento crucial para a história da avaliação dá-se por meio do Teste de Treviso (*Treviso Test*). Aplicado ao final do ano escolar nas escolas ao norte da Itália na Idade Média, o teste focava no processo escolar e tinha por base conteúdos curriculares. Seu objetivo era o de hierarquizar os alunos com base no domínio dos conteúdos ensinados ao longo

do ano e os professores eram remunerados de acordo com o sucesso de seus alunos na avaliação. O teste sofreu influências, no século XVI, dos procedimentos avaliativos da China, procedimentos esses que foram adotados pelo Teste de Treviso adaptados ao propósito de controle curricular nos séculos XVII e XVIII (SPOLSKY, 2017).

Ainda na Europa, a revolução francesa instaura a necessidade de um governo forte e centralizado, trazendo a escola, inclusive o ensino religioso e a avaliação curricular, sob domínio do governo, sendo esta última de responsabilidade de departamentos acadêmicos. Na Marinha Britânica, no século XVII, por sua vez, institui-se a avaliação oral. Apesar de ser considerada como não sendo confiável dada a ausência de juízes treinados para sua aplicação, detinha grande autoridade à época e combinava características do modelo chinês de seleção por mérito, em oposição ao apadrinhamento, e do Teste de Treviso, em que se avaliava o domínio de um corpo de conhecimentos, em oposição ao ranqueamento simples (SPOLSKY, 2017).

Já no século XIX, com base nos exames desenvolvidos em Oxford e Cambridge, que se caracterizavam pelo ranqueamento de um número pequeno de candidatos, surge um instrumento avaliativo com a finalidade de selecionar servidores civis ingleses e indianos. Em Oxford, a avaliação era empregada com finalidade escolar honorária e, em Cambridge, para seleção dos melhores alunos. Similarmente, na Prússia, a avaliação foi adotada para seleção de magistrados e na Inglaterra de maneira geral e em suas colônias, para verificação de desempenho escolar em instituições estaduais. No século XVI, a Universidade de Cambridge já realizava a aplicação de testes orais para a avaliação de conhecimentos de Latim ou da capacidade de emprego da fala (SPOLSKY, 2017).

No século XX, a área de avaliação passa a fazer uso de técnicas psicométricas. Na Inglaterra, modelos com foco na objetividade foram empregados em avaliações com finalidade de seleção de escolas para matrícula de alunos. Nos Estados Unidos, contudo, especialmente a partir de 1920, ocorreu a proliferação desse tipo de teste e de empresas especializadas na oferta dos serviços para escolas. O interesse estadunidense deu-se em virtude do sucesso dos testes de inteligência aplicados à época da primeira guerra mundial. Assim, surgiram os testes objetivos – também conhecidos por testes americanos – que eram compostos, inicialmente, por questões de verdadeiro ou falso e, posteriormente, por questões de múltipla escolha. Tais testes passaram a ser vistos como ideais e, apesar de encontrarem resistência inicial, foram rapidamente disseminados pelo mundo com o advento da globalização após a segunda guerra mundial (SPOLSKY, 2017).

Atualmente, Spolsky (2017) destaca que testes objetivos ainda têm sido empregados com finalidade de prestação de contas no ambiente escolar, especialmente atrelados ao princípio

de Treviso, segundo o qual professores devem ser pagos de acordo com os resultados dos alunos. No Brasil, a prática é adotada, por exemplo, pela Secretaria da Educação do Estado de São Paulo desde 2010, com o Programa de Valorização pelo Mérito².

A área de testagem, por sua vez, surgiu em oposição ao contexto educacional, ocupando-se da avaliação de competências. Inicialmente, a ênfase recaía sobre habilidades de letramento, as quais eram avaliadas por meio de tarefas que envolviam a redação ou a tradução de textos ou, ainda, a compreensão, a qual era avaliada por meio de testes objetivos. A avaliação das habilidades orais, quando utilizada, ressaltava a dificuldade de ajuste das competências às escalas objetivas e a necessidade de se desenvolver métodos que empregassem menos recursos para sua aplicação, dando origem às avaliações via computador ou telefone (SPOLSKY, 2017). Após a segunda guerra mundial, os procedimentos de testagem, principalmente relacionados às habilidades orais, intensificaram-se (FULCHER, 2003).

Em 1913, com a introdução do CPE – *Certificate of Proficiency in English*, a avaliação da modalidade oral da língua foi apresentada, com foco na avaliação de pronúncia por meio de ditado e de leitura e conversação. A avaliação tinha por objetivo dar suporte ao conteúdo programático das escolas e encorajar boas práticas de ensino e de aprendizagem. A introdução do *Lower Certificate* – conhecido por *FCE - First Certificate in English* ou, mais recentemente, *B2 First* – em 1939 também incluía a avaliação de proficiência oral. No entanto, apesar do emprego recente desse tipo de avaliação nos testes britânicos, existem poucas documentações acerca do desenvolvimento e uso desses testes na Inglaterra (FULCHER, 2003).

Até 1939, a avaliação de línguas nos Estados Unidos, por sua vez, preocupava-se com a busca por resultados que atendessem aos princípios de confiabilidade. Acreditava-se, assim, que a avaliação da produção oral não fornecia resultados confiáveis diante de sua subordinação a julgamentos orientados por fatores não controláveis. Assim, até o ano de 1926, a expressão “teste oral” era majoritariamente empregada em referência à testagem de pronúncia, e não necessariamente de produção oral, dada a inconsistência e impraticabilidade atribuídas à última. O primeiro teste estadunidense de avaliação da produção oral consistiu no *The College Board's English Competence Examination*, aplicado a alunos estrangeiros que buscavam uma oportunidade nas universidades norte-americanas (FULCHER, 2003).

A partir de 1940, por sua vez, a avaliação da produção oral é intensificada por conta dos esforços realizados para suprir necessidades comunicativas no período de guerra. Desse modo, surgem os fundamentos relacionados à estruturação dessas avaliações, os quais são conhecidos

² Mais informações podem ser obtidas em <<http://www.educacao.sp.gov.br/valorizacao>>. Acesso em 10 Set. 2018.

e empregados ainda na contemporaneidade. Destaca-se, assim, o foco na avaliação do desempenho oral; a necessidade de treinamento de avaliadores; o modelo de entrevista face-a-face na presença de um examinador interlocutor e de um examinador observador, sendo este último responsável pela atribuição de notas; a preocupação com o contexto da situação de testagem, como o discurso do examinador interlocutor e as condições de aplicação do teste; e o formato das entrevistas, iniciado com uma sessão de conversas informais, cujo desempenho não integrava a avaliação, e seguido por tarefas que elicitavam a participação em situações comunicativas sobre serviços de segurança e de fornecimento de informação (FULCHER, 2003).

Em 1960, surge a nova geração de estudos em testagem, tendo as universidades de Princetown e Cambridge centralizado e industrializado as aplicações em larga escala. Assim, a área de testagem progrediu de exames de cunho mais tradicional, voltados para a tradução escrita, redação, compreensão e gramática, para testes de cunho psicométrico, em que eram avaliadas estruturas linguísticas – como o *Michigan Lado Test of English* –, para uma abordagem de cunho mais integrativo, considerando-se aspectos de ordem sociocultural (SPOLSKY, 2017).

Diante dos avanços e da importância social da avaliação de línguas, a partir da metade do século XX, a área passa a ser considerada um campo de estudos da Linguística Aplicada, consistindo em uma das áreas de estudo mais recentes (FULCHER, 2003). O nascimento da testagem moderna no referido século foi impulsionado, em grande parte, pelos trabalhos dos estadunidenses Lyle Bachman, J. D. Brown e Adrian Palmer e dos britânicos Alan Davies, Don Porter, Arthur Hughes, Cyril Weir e Charles Alderson, que buscavam o desenvolvimento de avaliações linguísticas com valores teóricos explícitos, distanciando-se das técnicas de ordem psicométrica. Diante disso, assuntos que envolvem questões de validação, localização, profissionalismo, reflexividade e politização tornam-se centrais (O’SULLIVAN, 2011).

No século XXI, por sua vez, os testes alcançaram ainda maiores poderes nas áreas educacional, política e social, de modo que eles passaram a ser empregados em larga escala como instrumentos confiáveis de seleção, promoção, reformas e prestação de contas. Diante desses usos, a Linguística Aplicada passa a primar pela busca por abordagens avaliativas mais progressistas e abertas (SHOHAMY, 2007). Assim, Spolsky (2017) destaca a existência de duas grandes tendências de estudo contemporâneas: o esforço para superar a inevitável incerteza que permeia os julgamentos avaliativos e a avaliação de habilidades integradas. Os estudos sobre a incerteza focam-se em tentativas de obtenção de níveis maiores de confiabilidade e de validade nos resultados e usos das avaliações. Basicamente, o alcance da confiabilidade é almejado por

meio do emprego de métodos estatísticos – emprego esse que pode ser considerado controverso diante dos problemas de interpretação dos resultados comumente cometidos a respeito das medidas de diferença e relação fornecidos pelos testes (RAGANATHAN; PRAMESH; AGGARWAL, 2017) –, sendo a validade avaliada em relação aos impactos sociais gerados pelo instrumento. A avaliação de habilidades de forma integrada, por sua vez, impõe dificuldades para mensurações de caráter mais objetivo das amostras de língua e, portanto, dá origem a dois impasses: a busca por escalas confiáveis diante do desenvolvimento de testes industrializados em larga escala e a tentativa de atingir as demandas do mercado, que preza por mensurações de proficiência válidas e autênticas.

Mesmo diante dos avanços, Spolsky (2017) acredita que a área ainda carece de estudos sobre a natureza da proficiência linguística e sobre modos possíveis de se avaliar sua relevância para propósitos sociais específicos e questiona a reascensão de debates acerca das potencialidades de emprego de testes de múltipla-escolha, a insistência na manutenção do perfil linguístico padrão representado pela figura do falante nativo devidamente escolarizado e na redução de perfis multidimensionais a escalas unidimensionais, a produção de testes em escala industrial, e a busca pela interpretação de conceitos não significativos, atribuídos aleatoriamente. Assim, o autor aponta a necessidade de estudos que (i.) auxiliem na redução do poder e no aumento da humanidade em testes industriais e computadorizados, (ii.) empreguem apenas testes válidos para decisões importantes envolvendo carreiras, (iii.) garantam a elaboração de escalas com perfis complexos e (iv.) tratem da questão do uso inadequado dos resultados para o desenvolvimento da área.

2.1.2 O desenvolvimento de exames de proficiência linguística: o modelo de utilidade do teste e suas qualidades

O modelo de utilidade introduzido por Bachman e Palmer (1996) condensa as características principais a serem consideradas no desenvolvimento e uso de um teste, as quais justificam sua necessidade de elaboração e orientam o controle de qualidade do mesmo. O modelo surge como uma tentativa de reduzir a abstração teórica dos modelos de validade apresentados até então, oferecendo orientações de ordem mais prática para a incorporação dos conceitos na criação de instrumentos avaliativos (XI; SAWAKI, 2017) e garantindo atuações éticas (DAVIES, 2017).

O modelo congrega seis qualidades – a confiabilidade, a validade de construto, a autenticidade, a interatividade, o impacto e a praticidade –, as quais subdividem-se em explicações de ordem deontológica, congregando valores intrínsecos ao exame, e de ordem

teleológica, por envolverem o conceito de consequências (DAVIES, 2017). Durante o desenvolvimento e uso do teste, essas qualidades devem ser maximizadas. No entanto, a maximização pode gerar algumas tensões entre os elementos, especialmente entre os conceitos de validade e confiabilidade e entre autenticidade e confiabilidade, cuja relação é estabelecida de modo inversamente proporcional. Assim, a maximização deve almejar o conceito de utilidade do teste em geral, sem necessariamente ater-se às características em específico, ou seja, a avaliação de cada uma das qualidades deve ocorrer de acordo com seu efeito combinado para a obtenção máxima da utilidade geral, cujo equilíbrio ideal deve ser ajustado de acordo com cada situação de testagem (BACHMAN; PALMER, 1996).

Os conceitos de autenticidade e de interatividade consistem em conceitos relativos. No que tange à autenticidade, sua relativização dá-se em virtude da inexistência de possibilidade de uma tarefa ser totalmente autêntica na medida em que ela corresponde a apenas uma representação das características de uma tarefa da vida real. Assim, o conceito congrega os conceitos de validade de construto, ou seja, do grau com que as evidências de desempenho elicitadas pelo instrumento são efetivamente relevantes para mensuração dos traços comportamentais objetivados, e de validade de conteúdo, ou seja, do grau com que o instrumento é capaz de evidenciar o domínio do conteúdo que ele pretende medir. O conceito de interatividade, por sua vez, consiste no envolvimento das características individuais do candidato à tarefa que ele busca realizar, englobando, assim, noções de capacidade linguística, como conhecimento linguístico, competência estratégica e metacognitiva, conhecimento de tópico e esquemas afetivos (BACHMAN; PALMER, 1996).

A qualidade referente ao impacto, de acordo com Bachman e Palmer (1996), influencia tanto os indivíduos, em um nível micro, quanto a sociedade e os sistemas educacionais, em um nível macro. Assim, tem-se que a administração, o desenvolvimento, a aplicação, os usos e a participação em testes implicam certos valores e objetivos e geram consequências. De maneira direta, afetam-se aqueles que participam do procedimento de testagem, como o candidato, o usuário (*stakeholders*) e os decisores e de maneira indireta, os futuros alunos, os colegas de trabalho, empregadores, membros de sistemas educacionais de maneira geral.

O candidato, em específico, é afetado de três maneiras: pela experiência de participar da avaliação, e, se for o caso, por sua preparação para a avaliação; pelo *feedback* que recebe acerca de seu desempenho; bem como pelas decisões que são tomadas a seu respeito com base nos resultados. Assim, o afetamento atinge seu conhecimento de tópico, sua percepção do domínio de uso da língua, suas áreas de conhecimento linguístico e sua competência estratégica. Diante disso, Bachman e Palmer (1996) advogam em favor da necessidade de que as decisões

tomadas sejam justas, de modo que haja uniformidade entre os procedimentos avaliativos e os critérios e um uso justo do teste, por meio da relevância e da propriedade do resultado para informar decisões futuras e transparência ao candidato sobre como as decisões são tomadas.

Davies (2017), assim como Bachman e Palmer (1996), atenta para a diferença existente entre impacto e efeito retroativo. Para o autor, o impacto consiste nos efeitos que o teste tem sobre os usuários (*stakeholders*) quando colocado em uso, congregando seu modo de funcionamento em uma determinada situação e suas influências sistêmicas e permitindo que o instrumento passe por reavaliações e explicita os padrões que ele promove e as visões de língua. O efeito retroativo, por sua vez, consiste em um aspecto do impacto que se delimita a apenas contextos de ordem instrucional. Em um nível macro, é determinado pelos valores e objetivos de cada cultura (BACHMAN; PALMER, 1996). O impacto tem, portanto, uma abrangência maior do que o efeito retroativo, que, no caso da avaliação de línguas, se limita à influência gerada em professores e alunos a respeito das medidas necessárias para promover o ensino (MESSICK, 1996). Nesse sentido, McNamara (2006) destaca que os valores políticos e sociais de um teste devem ser cuidadosamente avaliados em relação a seus impactos políticos, criticando modelos de proficiência linguística individualistas, que assumem que os construtos de discurso e seus efeitos podem ser definidos *a priori* em favor de uma noção de co-construção social por meio de interação, através dos quais interlocutores dividem suas responsabilidades na criação de identidades e significados. Shohamy (2006) vai além e ressalta que avaliações não são neutras e que, portanto, têm poderes sociais capazes de promover exclusão, promoção ou seleção de indivíduos, os quais precisam ser constantemente revistos e gerenciados. Ademais, segundo a autora, tais poderes se estendem para o ambiente linguístico e cultural, possibilitando a marginalização de algumas línguas e a hegemonização e perpetuação de outras ou, ao menos, de variáveis dominantes por meio da definição das características de fala desejadas ilustradas nos critérios de correção, como a variedade nativa, por exemplo. A autora questiona, assim, a necessidade de emprego de testes e a contribuição desse emprego para a justiça social.

O conceito de praticidade, por sua vez, envolve a implementação, o desenvolvimento e o uso do teste quanto à disponibilidade dos recursos requeridos e a alocação e o gerenciamento dos recursos disponíveis. A praticidade afeta as decisões em todos os estágios, sendo cíclica. Os recursos são de natureza humana, como escritores, avaliadores, administradores, suporte técnico, dentre outros; de natureza material, como espaço, equipamento; e materiais, como papel, caneta, figuras; e temporal, como o tempo de desenvolvimento da prova, tempo de aplicação, tempo de correção.

A validade, por sua vez, consiste em uma qualidade do processo de avaliação voltada às interpretações dos dados em cada uso particular do teste. Dada sua especificidade, configura-se em um processo cíclico, que busca continuamente a integração da teoria e da prática da testagem. A natureza da validade é de caráter integrativo e complexo, o que acaba por dificultar sua compreensão e implementação (SHEPARD, 1993).

O conceito de validade que imperou entre 1950 e 1970 focava-se nas predições – em contraste às explicações com embasamento teórico e empírico – realizadas com base em resultados e apresentava uma variedade limitada e independente de tipos de validade, como as validades de conteúdo, preditiva e concorrente. No final da década de 1970 e início da década de 1980, com o surgimento das primeiras noções de validade de construto, estudos em Linguística Aplicada voltaram-se a fatores que interferem no desempenho dos candidatos de avaliações de línguas. Esses estudos impulsionaram investigações sobre a interpretação dos resultados com base em evidências empíricas (XI; SAWAKI, 2017).

No entanto, até a década de 1980, estudos sobre os tipos de validade costumavam considerá-los como características independentes, que poderiam ser avaliadas de forma igualmente independente. Foi apenas a partir da introdução do conceito de validade de construto proposto por Messick (1989) que a literatura passou a aceitá-lo como um conceito unitário, embora multifacetado (SHEPARD, 1993). Para Messick (1989), a validade consiste em um julgamento avaliativo acerca do grau com que evidências de ordem empírica e racionalidade teórica dão suporte à adequação e propriedade de inferências e ações baseadas nos escores do teste e em outras modalidades de avaliação.

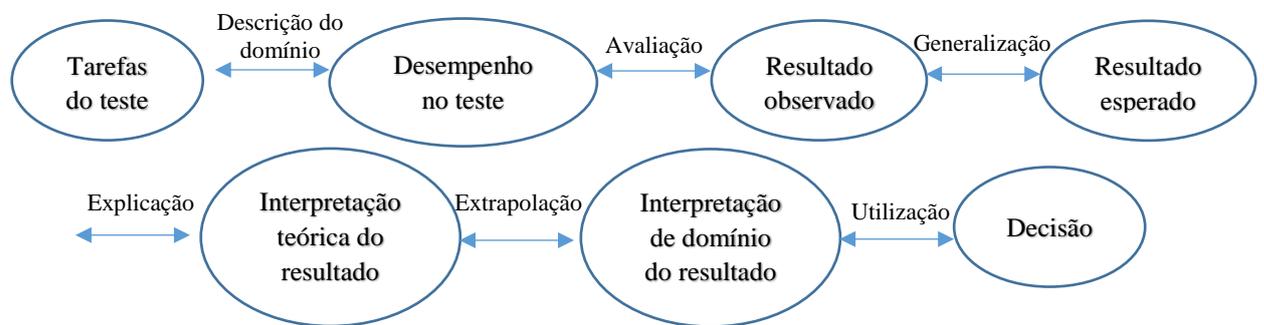
As facetas desse modelo iniciam-se pelo significado dos resultados e estendem-se para a relevância, os valores e as consequências sociais. Shepard (1993) tece críticas à ordenação proposta pelo autor argumentando que, ao iniciar a busca por evidências empíricas a partir dos resultados do teste, a conclusão do processo com a análise de consequências torna-se inviável. Diante disso, a autora advoga em favor de orientar a busca por evidências em torno do questionamento “o que uma determinada prática avaliativa se propõe a fazer?”, congregando características de uma outra abordagem de validade, a argumentativa.

A validade de argumento proposta por Kane (1992) e adaptada em Kane (2004) leva em consideração as afirmações que são tecidas a respeito das capacidades de um candidato, sobre as quais se avalia a credibilidade. A abordagem apresenta a crença na existência de interação entre os argumentos utilizados para interpretação dos resultados e baseia-se nas premissas de que os resultados e as informações providas aos usuários (*stakeholders*) são relevantes, úteis e

suficientes para a tomada de decisões; as tomadas de decisões são apropriadas; e a avaliação não incorre em consequências negativas.

O processo de validação de argumento encontra-se representado na Figura 1. Por validação, entenda-se o grau de defensabilidade das inferências que foram realizadas sobre os candidatos com base em seus desempenhos (McNAMARA, 2000). O conceito de validade de argumento ampara-se, assim, no conceito de validade para desenvolver e avaliar as evidências que foram empregadas para uma determinada interpretação e uso (XI; SAWAKI, 2017).

Figura 1. Ligações entre argumentos interpretativos.

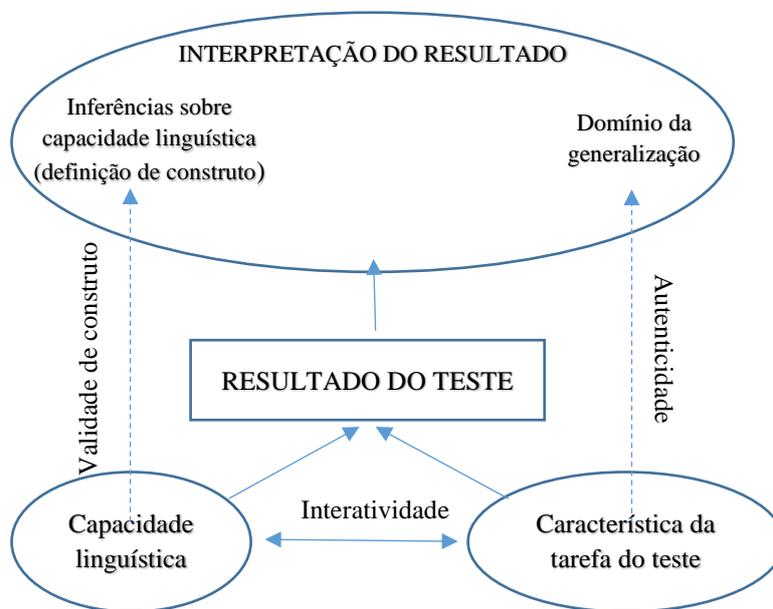


Fonte: Traduzido de Xi e Sawaki (2017, p. 197).

Diante da abstração do conceito proposto por Messick (1989), Bachman e Palmer (1996) propõem-se a oferecer uma noção mais prática e acessível para os avaliadores – introduzindo a teoria da utilidade do teste. Assim, a noção de validade de construto por eles apresentada consiste no nível de generalização possível para a interpretação de um resultado, o qual nunca é absolutamente válido.

O conceito congrega a definição de construto estabelecida pelo teste com as características, em termos de autenticidade e interatividade, das tarefas e baseia-se, majoritariamente, nos resultados fornecidos, conforme demonstrado pela Figura 2.

Figura 2. Validade de construto da interpretação dos resultados.



Fonte: Tradução de Bachman e Palmer (1996, p. 22).

Bachman e Palmer (1996) informam que as etapas de elaboração de um teste perpassam, ao menos, três fases: o desenho, a operacionalização e a administração. Na fase de desenho, são identificados, descritos e delimitados os construtos que compõem o instrumento avaliativo, bem como em que se planeja o modelo de utilidade do teste e procede-se à identificação, alocação e gerenciamento dos recursos necessários. De maneira geral, o desenho do teste reflete a maneira como as teorias sobre a natureza do uso da linguagem selecionadas para compô-lo no contexto de uso da língua para o qual o teste foi especificado foram compreendidas por seus desenvolvedores (McNAMARA, 2000) e consiste em uma fase essencial para minimizar inconsistências de confiabilidade (BACHMAN; PALMER, 1996). A fase de operacionalização, por sua vez, consiste naquela em que os planos são transformados em elementos materiais, por meio da escrita das tarefas, das instruções, dos pilotos e dos parâmetros avaliativos; e a fase de administração, na aplicação do teste em si.

O conceito de confiabilidade da avaliação de proficiência, por fim, é bastante controverso, dado que sua maximização pode gerar interferências em outras qualidades. Os tipos de confiabilidade mais investigados consistem na confiabilidade voltada aos avaliadores, dentre os quais se destaca a confiabilidade entre avaliadores e a confiabilidade dos resultados

atribuídos pelo mesmo avaliador em diferentes momentos do processo; e a confiabilidade voltada ao próprio instrumento, como a confiabilidade existente entre as tarefas do teste e entre conjuntos de testes. A confiabilidade refere-se ao nível de consistência da mensuração em diferentes características da situação de testagem e parte do princípio da existência de informações suficientemente relevantes acerca do candidato para a realização de decisões e de inferências (SMITH, 2003).

Desse modo, tendo em vista que o objetivo da validade é demonstrar significado aos resultados de um teste e justificar seus usos, verifica-se uma forte ligação entre os conceitos de validade e de confiabilidade das avaliações. Destaca-se, portanto, a centralidade dos resultados para a avaliação da validade de um teste, dado que são eles os responsáveis pela interpretação das capacidades do candidato e por extensão das decisões tomadas (CHAPELLE, 2012). Esses resultados, portanto, devem ser indicadores confiáveis da capacidade que o teste se propõe a medir, de modo que fontes potenciais de inconsistências de mensuração possam ser definidas, uma vez que elas não podem ser totalmente eliminadas (BACHMAN; PALMER, 1996).

Nesse sentido, Weir (2005a) entende a confiabilidade como um dos elementos que compõem as várias faces da validade de um teste, propondo que a noção seja abordada por meio do conceito de validade de resultados, a qual orienta esta investigação. Assim, Weir (2005a) determina a existência de cinco tipos de validade: validade de contexto, validade de teoria, validade de resultados, validade de consequências e validade de critério. A validade de resultados, diretamente relacionada à validade de contexto e de teoria – ou seja, preocupada com o que o teste efetivamente é capaz de medir – engloba todos os elementos da confiabilidade. A validade de resultados preocupa-se, desse modo, (i) com o grau com que os resultados fornecidos estão embasados em critérios apropriados, (ii) em exibir opiniões consensuais na classificação, (iii) com o grau em que os resultados estão livres de erros de mensuração, consistentes em termos das amostras de conteúdo e estáveis através do tempo, e (iv) em engendrar confiança, como indicador de tomada de decisão confiável. A validade de critério, por sua vez, indissociável da validade de resultados, é determinada por meio da equivalência comparativa entre os resultados do teste e um critério externo que possa medir a mesma capacidade, que pode ser aplicado no mesmo momento em que o candidato realiza o exame ou teste (validade simultânea) ou posteriormente (validade preditiva).

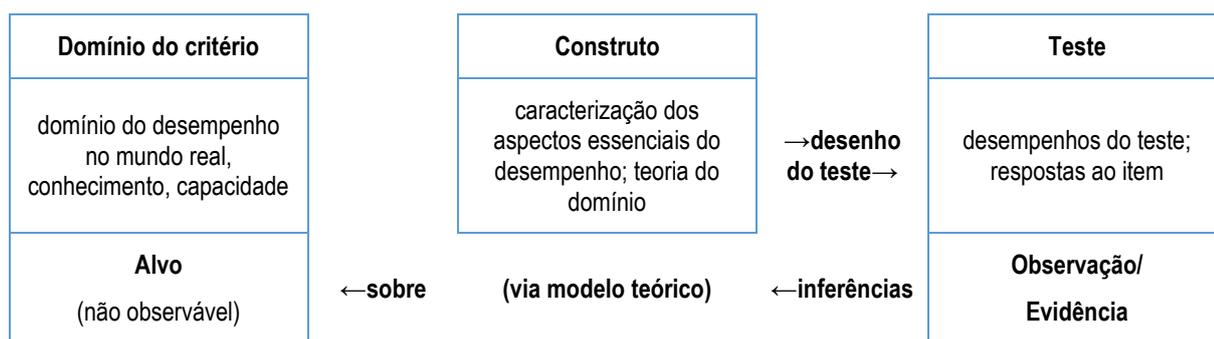
Por relacionar-se às validades de contexto e de teoria, para a mensuração da validade de resultados de um teste é necessário que se conheça qual a representação que esse teste apresenta do objeto a ser avaliado, ou seja, é necessário que se conheça o construto do referido instrumento de avaliação – presente em materiais de divulgação do exame, por exemplo – e

representado – presente em trabalhos científicos sobre o exame, em sua escala de proficiência e em suas tarefas. A partir da compreensão do construto, é possível determinar os critérios autóctones, ou seja, autenticamente representados (JACOBY; MCNAMARA, 1999) e as alternativas que serão empregadas no processo de avaliação e de tomada de decisão sobre o candidato. Assim, é imprescindível que haja validade de construto entre o que o teste se propõe a avaliar e os itens/as tarefas que o compõem, de modo a haver correlação adequada entre o objeto e os resultados da avaliação.

Diante disso, McNamara (2006) ressalta a finalidade de generalização de testes de línguas. Isso porque, segundo o autor, os resultados só fazem sentido quando inferências além da situação de teste podem ser realizadas com base nestes testes. Assim, ele destaca que o desempenho do candidato em si não é o interesse principal do processo avaliativo, pois não apresenta significado ou valor em si mesmo, mas, sim, no propósito de prever a atuação em contextos extra-teste, ou seja, no mundo real. McNamara (2000) define por critério os aspectos referentes à capacidade linguística do candidato no domínio do mundo real. O construto, portanto, consiste em uma representação abstrata desses critérios para o universo específico da testagem.

Para McNamara (2006), a base para a realização de inferências encontra-se na interação entre critério, construto e o teste em si, conforme representado na Figura 3.

Figura 3. Interação entre construto, critérios e o teste.



Fonte: McNamara (2006, p. 28).

A modelagem do domínio do critério origina, portanto, o construto, a partir do qual são definidos os procedimentos por meio dos quais os desempenhos serão observados, ou seja, o processo de avaliação, o que garante a possibilidade de inferências por parte do decisor. Desse modo, as inferências consistem em elementos essencialmente incertos e, portanto, sujeitos a questionamentos (MCNAMARA, 2006). É diante desses princípios que a Linguística Aplicada,

desde 1939, tem buscado elementos que garantam suporte à validade e à confiabilidade dos resultados das avaliações de línguas (FULCHER, 2003).

As escalas de proficiência são, portanto, essenciais para a avaliação por apresentarem uma relação direta com o construto (BACHMAN; PALMER, 1996; FULCHER, 2003; LUOMA, 2004), ou seja, deve haver uma relação direta entre o conteúdo do teste e os descritores da escala (ALDERSON, 1991), uma vez que são as escalas que oportunizam a definição operacional do construto linguístico. De acordo com Alderson (1991), considerando-se seus propósitos e funções, as escalas de proficiência podem ser divididas em três categorias e não deve haver desvios de uso de escalas desenvolvidas para propósitos específicos. Na primeira, a escala tem por propósito orientar os usuários (*stakeholders*) e função informativa, facilitando a interpretação dos resultados atribuídos ao candidato. O autor destaca que tais escalas garantem mais significado aos usuários (*stakeholders*) por fornecerem mais informações sobre os desempenhos e estão menos propensas a erros de interpretação por não impetrarem uma impressão irreal de precisão como as escalas numéricas. Na segunda categoria, o autor descreve escalas com propósito de orientar o examinador, cuja função é guiar o processo de avaliação. Nesse caso, as escalas garantem padronização das avaliações e, por consequência, maiores níveis de confiabilidade e validade ao processo. Na terceira categoria, por fim, o autor aborda escalas com propósito de orientar o elaborador, cuja função é orientar o processo de criação de um teste a níveis apropriados por meio do fornecimento de especificações.

Brown, Iwashita e McNamara (2005) destacam que o desenvolvimento de escalas de proficiência deve garantir que as escalas produzidas sejam relevantes para o contexto, significativas para os examinadores e observáveis no desempenho das tarefas. O desenvolvimento de escalas de proficiência segue, de maneira geral, duas orientações: uma abordagem intuitiva ou uma abordagem empírica. A abordagem intuitiva, geralmente, baseia-se em escalas pré-fabricadas, na literatura e/ou nas expectativas pessoais dos elaboradores com base em suas opiniões, ao passo que, na empírica, os elementos essenciais são definidos com base em observações de desempenho reais. North e Schneider (1998) acreditam que escalas de ordem intuitiva são válidas apenas para contextos de avaliação de baixo impacto, não devendo ser empregadas em testes de alto impacto. Dentre os procedimentos de elaboração de escalas empíricas, Brown, Iwashita e McNamara (2005) destacam a de julgamentos de especialistas a respeito do desempenho do candidato e a definição de características ou comportamentos-chave para decisão. Desse modo, além dos conceitos apresentados pela teoria de utilidade do teste proposta por Bachman e Palmer (1996), é importante destacar que as escalas de proficiência de exames de alto impacto devem preocupar-se, ainda, com uma propriedade adicional: a

criterialidade. Proposto por Hawkins e Buttery (2009, 2010) e incorporado à versão do CEFR – Quadro Comum de Referências para Línguas em 2018, o conceito diz respeito às características linguísticas típicas de níveis de proficiência específicos e que possibilitam a distinção entre níveis de proficiência adjacentes. Ressalta-se, contudo, que as características criteriosais não englobam todas as particularidades de uso da língua em um nível específico, apenas as características distintivas.

A avaliação de proficiência tradicionalmente segue, portanto, duas abordagens, de acordo com a visão de língua dos elaboradores do exame: a integrativa e a discreta, ambas realizadas com base na análise de desempenho (MCNAMARA, 1996). O conceito de avaliação integrativa foi introduzido por Carroll (1961) em contraposição ao de avaliação discreta, alvo de interesse da testagem até então por garantir altos índices de confiabilidade. O autor propôs que a avaliação tivesse por foco o efeito comunicativo global da fala dos candidatos, e não o desempenho em elementos estruturais isolados. Críticas diversas são, contudo, tecidas em reação à abordagem dicotomizada desses modelos, pois se acredita que, individualmente, eles não sejam capazes de agregar de maneira homogênea a relação entre formatos de tarefas e a natureza do teste, ou seja, como ressalta Farhady (1979), é possível a existência de testes de natureza gramatical, por exemplo, que empreguem tarefas integrativas ou vice-versa. Nesse sentido, Oller (1973) propõe que ambas as abordagens não sejam vistas de maneira dicotômica, mas, sim, como representantes de pontas extremas de um *continuum* de abordagens avaliativas.

A visão de língua determina, também, a natureza da escala de mensuração a ser empregada para avaliação da proficiência linguística. Tradicionalmente, a diferenciação é estabelecida entre escalas analíticas e holísticas. Escalas analíticas são aquelas que se ocupam de aspectos individuais das características da fala, como precisão, fluência e conteúdo; a holística, por sua vez, realiza uma avaliação global do desempenho dos candidatos, ancorando-se concomitantemente nas dimensões de impacto da mensagem produzida, de uso da língua e de desenvolvimento de tópico (OUNIS, 2017).

De acordo com Bachman (1990), o desenvolvimento de escalas de mensuração segue, geralmente, duas abordagens: a abordagem orientada por critérios e a abordagem orientada por normas. Enquanto a primeira preocupa-se com a mensuração do desempenho do candidato em relação a um nível de capacidade, ou seja, a domínios de conteúdo específicos, a segunda orienta-se por comparação de desempenho entre candidatos, os quais formam pontos de referência por meio da aplicação de testes padronizados a um grande número de candidatos. O autor destaca que escalas de mensuração devem, idealmente, apresentar quatro propriedades: potencialidade distintiva, ordenação de grandezas, intervalos e um ponto zero absoluto. No

entanto, o autor reconhece a impossibilidade de se atingir as quatro propriedades na avaliação de atributos de proficiência dadas a natureza das capacidades que são objetos da avaliação e às limitações de definição e observação de desempenho.

Assim, o autor define quatro tipos de escalas de mensuração, em ordem crescente de sofisticação, considerando-se suas capacidades informativas: a nominal, a ordinal, a intervalar e a de proporção. A escala nominal consiste na atribuição de códigos numéricos a categorias ou classes de um determinado atributo e apresenta propriedade distintiva, pois não trabalha com noções de hierarquização. Já a escala ordinal congrega o conceito de propriedade de ordenação, pois consiste na hierarquização comparativa do atributo em níveis. A escala intervalar, por sua vez, propicia a ordenação dos atributos por meio do estabelecimento de distâncias iguais entre os níveis, congregando as propriedades de distinguibilidade, ordenação e igualdade de intervalos. A escala de proporções, por fim, incorpora o conceito de ponto zero e permite o estabelecimento de comparações entre grandezas com unidades de mensuração distintas (BACHMAN, 1990).

Evidencia-se, deste modo, a importância do modo como os resultados de uma avaliação são atingidos e informados para o processo de validação. É válido ressaltar, assim, que testes não são passíveis de validação. Na verdade, são suas interpretações e seus usos que devem ser validados, não existindo, portanto, o conceito de “testes válidos” (MESSICK, 1989; KANE, 2002). Kane (2002) ressalta, ainda, que como interpretações têm caráter vago e implícito, elas sempre serão fragmentadas e imprecisas. A esse respeito, Chapelle (2012) destaca que a busca por “testes válidos” costuma ser um dos motivos que instigam a elaboração de novos instrumentos avaliativos. Assim, a autora destaca que a responsabilidade pela validade de um teste costuma ser atrelada apenas aos desenvolvedores especialistas, que baseiam suas definições de mensuração em fontes autoritárias. No entanto, ressalta que os usuários (*stakeholders*) do teste, especialmente os candidatos, desempenham um papel fundamental no processo de validação.

2.1.2.1 A avaliação da produção oral em língua estrangeira: o Quadro Comum de Referências para Línguas (CEFR) e os critérios de exames de proficiência consagrados

O CEFR – Quadro Comum de Referência para Línguas consiste, atualmente, em uma das especificações mais importantes para guiar a garantia de qualidade de cursos e avaliações de línguas estrangeiras. O quadro surgiu no início da década de 1970 como uma iniciativa do Conselho Britânico para determinar níveis funcionais mínimos de proficiência entre usuários da língua inglesa visando ao norteamo de cursos de ensino de línguas. No entanto, as

tentativas de definições dos níveis de proficiência na Europa datam de 1913, com a elaboração do CPE – *Certificate of Proficiency in English*, atualmente *C2 Proficiency*, pela *Cambridge Assessment English* e estende-se para além de 1939, com a criação do FCE – *First Certificate in English*, atualmente *B2 First* (TAYLOR; JONES, 2006; NORTH, 2006). Nessa época, conforme destacado por Fulcher (2003), a área de testagem recém surgida preocupava-se com a busca de padrões de confiabilidade avaliativa.

Em 1970, contudo, o Conselho Britânico lança o *Modern Language Project* em uma tentativa de definir um quadro de referências linguísticas. Em sua versão inicial, o quadro continha apenas dois níveis, o básico (*waystage*) e o intermediário (*threshold*). O nível básico descrevia um ponto de evolução da proficiência em direção ao nível intermediário. Este, por sua vez, indicava o nível de proficiência necessário para imigrantes e visitantes operarem de maneira efetiva na sociedade por meio da língua.

Na década de 1980, a *Cambridge Examinations*, o Conselho Britânico e a *BBC English* firmaram uma parceria visando à revisão dos níveis estabelecidos na década anterior diante da relevância que esses níveis haviam atingido para programas educacionais de aprendizado de línguas que permitiam a obtenção da cidadania europeia. A revisão proposta resultou nas especificações que guiaram a elaboração do PET – *Preliminary English Test*, atualmente *B1 Preliminary*, em meados da década de 1980 e o KET – *Key English Test*, atualmente *A2 Key*, no início da década de 90, ambos ofertados pela *Cambridge Assessment English* (TAYLOR; JONES, 2006).

Na década de 1990, uma parceria do Conselho Britânico com a ALTE – *Association of Language Testers in Europe*, que vinha desenvolvendo seu próprio quadro de referências, resultou na descrição linguística e funcional de um terceiro nível de proficiência, superior aos dois desenvolvidos anteriormente, o nível de usuário independente (*vantage*), que considerava as especificações apresentadas no FCE (atualmente, *B2 First*). Em 1991, mais especificamente, o desenvolvimento do CAE – *Certificate of Advanced English*, atualmente *C1 Advanced*, orientou a criação do nível da proficiência operativa eficaz (*effective operational proficiency*), que estabelecia um degrau de ligação entre o nível intermediário e o de domínio pleno (*mastery*), cujas especificações alinhavam-se ao CPE (atualmente, *C2 Proficiency*). Na mesma época, desenvolveu-se, também, o nível iniciante (*breakthrough*) (TAYLOR; JONES, 2006; NORTH, 2006).

O CEFR passou a ter, assim, a partir da década de 1990, uma escala de proficiência que conta com seis níveis distintos – A1, A2, B1, B2, C1, C2 – em uma estrutura crescente. Os níveis A1 – iniciante e A2 – básico, corresponderiam ao que tradicionalmente se define como

usuário básico, os níveis B1 – intermediário e B2 – usuário independente, ao usuário independente, e os níveis C1 – proficiência operativa eficaz e C2 – domínio pleno, ao nível de usuário proficiente. O desenvolvimento do quadro tem por objetivo promover transparência e coerência no ensino de línguas (UCLES; CUP, 2011) e orienta-se pela característica dinâmica e global da língua inglesa. Atualmente, encontra-se redefinido para abarcar descrições de níveis de proficiência para outras línguas estrangeiras que não apenas o inglês (TAYLOR; JONES, 2006; COE, 2018).

A proficiência comunicativa é entendida pelo CEFR como uma combinação de estratégias e competências em uma tentativa de compreender a complexidade do conceito. Desse modo, o projeto entende a proficiência como a

capacidade de desempenhar atividades linguístico-comunicativas (é capaz de...), ao mesmo tempo em que faz uso de **competências genéricas e linguístico-comunicativas** (linguística, sociolinguística e pragmática) e ativa **estratégias comunicativas** apropriadas³. (COE, 2018, p. 32, grifos originais)

As competências e estratégias são, assim, mobilizadas visando-se à realização de tarefas. As competências genéricas contempladas pelo quadro dizem respeito ao conhecimento declarativo (*savoir*), às habilidades e conhecimentos (*savoir-faire*), à competência existencial (*savoir-être*) e à capacidade de aprender (*savoir apprendre*) (COE, 2018).

As atividades e estratégias linguístico-comunicativas, por sua vez, representam-se por meio dos modos de comunicação receptivo, produtivo, interativo e mediador. No nível receptivo, as atividades englobam a compreensão oral, escrita e audiovisual; e as estratégias, a identificação de sinais e realização de inferências. Em termos de produção, tem-se que as atividades tratam da fala e da escrita; e as estratégias, do planejamento de gêneros, e da compensação e do monitoramento e reparo de lacunas de vocabulário e terminologia. No nível da interação, as atividades englobam a interação falada, a interação escrita, a interação online; e as estratégias, por fim, dizem respeito à tomada de turno, cooperação, e a pedidos de clarificação (COE, 2018). Os quadros 1 e 2, respectivamente, condensam as atividades e estratégias orais consideradas pelo CEFR e a criterialidade a elas atribuída.

³ No original: “[...] *ability to perform communicative language activities (can do...), whilst drawing upon both general and communicative language competences (linguistic, sociolinguistic, and pragmatic), and activating appropriate communicative strategies.*” (COE, 2018, p. 32, grifos originais)

Quadro 1. Atividades linguístico-comunicativas do CEFR.

	Atividades	Criterialidade
Recepção	<u>Compreensão oral</u> Interação entre outros falantes Membro de uma plateia ao vivo Comunicados e instruções Mídias de áudio e gravações <u>Compreensão escrita</u> Leitura de correspondência Leitura para orientação Leitura para informação e argumentação Leitura para instruções Leitura como atividade de lazer <u>Compreensão audiovisual</u> Assistir TV, filme ou vídeo	
Produção	<u>Fala</u> Monólogo sustentado: descrição de experiência Aspectos descritos Complexidade do discurso Monólogo sustentado: fornecimento de informação Tipo de informação Grau de precisão Monólogo sustentado: realização de defesa Tópicos Modo de arguição Modo de formulação Anúncios públicos Tipo de conteúdo Inteligibilidade Necessidade de preparação Endereçamento de público Tipo de endereçamento Consideração do público Tratamento de perguntas <u>Escrita</u> Escrita criativa Escrita de relatórios e redações	
Interação	<u>Interação falada</u> Compreensão do interlocutor Conversação Discussão informal Discussão formal Cooperação orientada a objetivos Obtenção de produtos e serviços Troca de informações Realização de e participação em entrevistas Uso de telecomunicação <u>Interação escrita</u> Correspondências Notas, mensagens e formulários <u>Interação online</u> Conversação e discussão Transação e colaboração orientadas a objetivos	
Mediação	<u>Mediação de texto</u> Paráfrase de informação específica Explicação de dados Processamento de texto Tradução de texto escrito Tomada de notas Expressão de resposta pessoal a textos criativos Análise e crítica de textos criativos <u>Mediação de conceitos</u> Colaboração em grupo: facilitação de interação colaborativa entre pares, colaboração para construção de significado	

	<u>Liderança de trabalho de grupo: gerenciamento de interação, encorajamento de fala conceitual</u> <u>Mediação de comunicação</u> <u>Facilitação do espaço pluricultural</u> <u>Atuação como intermediário em situações informais</u> <u>Facilitação da comunicação em situações delicadas e de discordância</u>	
--	---	--

Fonte: Elaborado pela autora com base em COE (2018).

Quadro 2. Estratégias linguístico-comunicativas do CEFR e criterialidade.

	Estratégias	Criterialidade
Recepção	<u>Inferências e identificação de sinais</u> Exploração de ilustrações, formatação, cabeçalho, etc. Deduzir significado do co-texto e do contexto linguístico Explorar pistas linguísticas	
Produção	<u>Planejamento de gêneros</u> Consideração sobre como expressar o assunto e ensaio da expressão Consideração sobre reação do interlocutor <u>Compensação de lacunas</u> Uso de gestual paralinguístico Uso de vocabulário inadequado e clarificação Definição de conceito ausente Uso de paráfrases e extensão da evidência do emprego do recurso <u>Monitoramento e reparo</u> Mudança de estratégia Autocorreção Extensão com que um problema é evidente antes do reparo	Óbvia nos níveis A e estável em C
Interação	<u>Tomada de turno</u> <u>Cooperação</u> <u>Pedidos de clarificação</u>	
Mediação	<u>Explicação de novo conceito</u> <u>Vínculo com conhecimento prévio</u> Colocação de perguntas para encorajar a ativação do conhecimento Estabelecimento de comparações ou de conexões entre o conhecido e o conhecimento novo Fornecimento de exemplos e definições <u>Discriminação de informações complicadas</u> Discriminação em uma série de etapas Apresentação de ideias ou instruções em tópicos Apresentação dos pontos principais separada da cadeia de argumentos <u>Adaptação de linguagem</u> Paráfrase Adaptação de fala Explicação de terminologia técnica <u>Simplificação de texto</u> <u>Ampliação de um texto denso</u> <u>Simplificação de um texto</u>	B1 B2 C B1: uma de cada vez C1: reforço e recapitulação / C2: metáforas B2 B1: + simplificação B2 C

Fonte: Elaborado pela autora com base em COE (2018).

A competência linguístico-comunicativa, por sua vez, subdivide-se em linguística, pragmática e sociolinguística. No âmbito da competência linguística, são consideradas questões de variedade geral, variedade lexical, precisão gramatical, controle lexical, controle fonológico e controle ortográfico. A competência pragmática trata do conceito estendido de fluência, que engloba flexibilidade, desenvolvimento temático, coerência, precisão proposicional e retoma a ideia de tomada de turno apresentada anteriormente na estratégia de interação, além do conceito reduzido de fluência de fala. A competência sociolinguística, por fim, representa-se pela propriedade sociolinguística (COE, 2018).

Quadro 3. Competências linguístico-comunicativas do CEFR e criterialidade.

	Subcompetências	Criterialidade
Linguística	<u>Variedade linguística geral</u>	A1-B2, depois irrestrito
	Variedade de cenários	
	Tipo de linguagem	
	Limitações	A1-B2, depois irrestrito
	<u>Variedade lexical</u>	
	Variedade de cenários	
	Tipo de linguagem	A1-B1 B1-B2 B2-C2
	<u>Precisão gramatical</u>	
	Controle de repertórios específicos	
	Proeminência de erros	A1-B1 B1-B2 B2-C2
	Graus de controle	
	<u>Controle lexical</u>	
	Familiaridade de tópicos	A1-B1 B2-C2
	Graus de controle	
	<u>Controle fonológico</u>	
Articulação (pronúncia de sons/fonemas)	Níveis inferiores	
Prosódia (entonação, ritmo, acento de palavras, acento de sentença, velocidade de fala/ <i>chunking</i>)		
Ausência de sotaque e desvio de padrões		
Inteligibilidade e compreensibilidade	Níveis inferiores	
<u>Controle ortográfico</u>		
Transcrição de palavras e sentenças		
Ortografia	Níveis inferiores	
Inteligibilidade, considerando-se ortografia, pontuação e apresentação (<i>layout</i>)		
Pragmática	<u>Flexibilidade</u>	
	<u>Desenvolvimento temático</u>	
	<u>Coerência</u>	
	<u>Precisão proposicional</u>	
Sociolinguística	<u>Fluência de fala</u>	
	<u>Propriedade sociolinguística</u>	
	Adaptação de registro	

Fonte: Elaborado pela autora com base em COE (2018).

A consideração das atividades, estratégias e competências para a reestruturação da escala global do CEFR – Anexo I – exibem um comportamento híbrido de definição da proficiência, ou seja, pautado tanto em orientações holísticas quanto analíticas. É importante ressaltar, contudo, que o CEFR, diante de sua característica neutra perante contextos e, por extensão, tarefas de desempenho linguístico, bem como perante a língua materna e estrangeira dos aprendizes, não tem utilidade operacional para a definição de construtos. O caráter genérico e, portanto, a falta de especificidade descritiva não é adequada para orientar processos de elaboração de avaliações, devendo-se ter cautela nesse processo (PAPAGEORGIOU, 2005; UCLES; CUP, 2011). Em uma tentativa de oferecer uma compreensão mais clara do CEFR, o Conselho Britânico em parceria com a Universidade de Cambridge, elaboraram um programa de pesquisa intitulado *English Profile Programme*. O programa tem por objetivo verticalizar

detalhadamente as características dos níveis horizontalizados no CEFR, apresentando descrições de níveis de referência para acompanhamento do quadro de referências para a língua inglesa em relação ao vocabulário, à gramática e às funções comunicativas (UCLES; CUP, 2011).

No que concerne à gramática, a criterialidade é definida em propriedades linguísticas corretas e propriedade incorretas/erros (UCLES; CUP, 2011). As informações são analisadas com base na frequência dos erros e no perfil dos alunos que cometem determinados tipos de erros, considerando-se, para tanto, o sistema de codificação de erros desenvolvido com base no *Cambridge Learner Corpus* apresentado em Nichols (2003). De maneira geral, do nível A2 ao B2, a criterialidade é definida com base na aquisição de novas estruturas por parte do falante, ao passo que, nos níveis C, a criterialidade é definida com base na precisão de uso e na complexidade dessas estruturas. O Quadro 4 resume as características distintivas principais com relação aos níveis do CEFR e os tipos de erros mais comuns presentes na adjacência de cada nível.

Quadro 4. Criterialidade e tipos de erros gramaticais do CEFR.

Níveis	Características distintivas	Tipos de erros comuns entre faixas adjacentes
C2	<u>Domínio do uso das estruturas, com uma ocorrência mínima de desvios</u> Verbos específicos (<i>declare, presume, remember, tough</i>) + objeto + infinitivo Enunciados mais longos e precisos Extensão média do enunciado: 19 palavras/sentença	-
C1	<u>Uso mais preciso de estruturas</u> Verbos específicos (<i>chance, believe, find, take, suppose, assume, feel, prove, discover</i>) + objeto + infinitivo Modal para permissão (<i>might</i>) Menor recorrência de erros de concordância, contabilidade ou formação de palavra Extensão média do enunciado: 17,3 palavras/sentença	Concordância de anáfora Concordância do determinante Concordância do verbo Estrutura do argumento Contabilidade do determinante Contabilidade do nome Contabilidade do quantificador Derivação de adjetivo Derivação de advérbio Derivação de anáfora Derivação de conectivo Derivação de determinante Derivação de proposição Derivação de quantificador Forma do adjetivo Forma da anáfora Forma do determinante Forma nominal Forma do quantificador Flexão de adjetivo Flexão de advérbio Flexão de anáfora Flexão de determinante Omissão de advérbio Omissão de determinante Omissão de nome Omissão de verbo Seleção de advérbio Seleção de quantificador

		Seleção de verbo Necessidade do conectivo Necessidade do determinante Necessidade do nome Necessidade do quantificador Erro complexo (agramatical ou incompreensível)
B2	<u>Ampliação da frequência do uso de estruturas mais complexas</u> Orações reduzidas de gerúndio (-ing) antes da oração principal It + verbo + oração reduzida de infinitivo Oração <i>wh</i> - como sujeito de uma oração principal Discurso indireto Verbos/Adjetivos específicos (<i>proved, turned out, likely, sure, prefer, expected, known, thought, hard</i>) + infinitive Extensão média do enunciado: 14,2 palavras/sentença	Concordância Concordância nominal Estrutura do argumento Contabilidade do determinante Derivação de anáfora Derivação de conectivo Derivação de determinante Flexão do determinante Forma do adjetivo Forma do advérbio Forma nominal Forma do quantificador Forma do verbo Flexão de adjetivo Flexão de anáfora Flexão de quantificador Omissão de adjetivo Omissão de conectivo Omissão de determinante Omissão de preposição Omissão de quantificador Omissão de verbo Necessidade do conectivo Necessidade da preposição Necessidade do quantificador Necessidade do verbo Formação de negação Erro complexo (agramatical ou incompreensível)
B1	<u>Evidência de domínio de estruturas básicas e tentativas de uso de estruturas mais complexas</u> Gerúndio (-ing) Orações relativas (<i>whose</i>) Perguntas indiretas Orações com <i>what</i> como sujeito/objeto	Derivação de conectivo Derivação de determinante Forma do determinante Flexão do determinante Flexão do verbo (regular/irregular)

	<p>Verbo (<i>order, seems, suppose, expect, like, want</i>) + objeto + infinitivo / <i>easy</i> + infinitive</p> <p>Auxiliares complexos (<i>would rather, had better</i>)</p> <p>Modais de permissão (<i>may</i>), necessidade (<i>must</i>) e probabilidade (<i>should</i>)</p> <p>Uso de quantificadores</p> <p>Extensão média do enunciado: 10,8 palavras/sentença</p>	
A2	<p><u>Emprego de estruturas básicas e relativamente simples</u></p> <p>Frases simples</p> <p>Orações coordenadas (<i>that</i>)</p> <p>Frases descritivas introduzidas por participio passado</p> <p>Perguntas simples e diretas (<i>wh-</i>)</p> <p>Frases simples com infinitivo / outros infinitivos</p> <p>Modais de possibilidade (<i>may, can, might</i>), obrigação (<i>must</i>), conselho (<i>should</i>).</p> <p>Orações relativas simples (<i>who, which</i>)</p> <p>Extensão média do enunciado: 7,9 palavras/sentença</p>	<p>Concordância anafórica</p> <p>Forma do determinante</p> <p>Omissão de adjetivo</p> <p>Omissão de advérbio</p> <p>Omissão de conectivo</p> <p>Omissão de quantificador</p> <p>Flexão do quantificador</p> <p>Seleção do quantificador</p> <p>Necessidade do quantificador</p>
A1	-em desenvolvimento-	-em desenvolvimento-

Fonte: Elaborado pela autora com base em UCLES e CUP (2011)

No que concerne às funções, tem-se que este aspecto da linguagem é abordado pelo aspecto funcional-nocional também empregado no CEFR. Emprega-se, portanto, uma escala generativa com base nas capacidades dos candidatos que consideram cinco níveis de análise: a função social que o candidato consegue realizar, os tópicos do desempenho, o texto de insumo, o texto produzido, a qualidade desse texto e as condições/restrições sociais em que a produção ocorre (UCLES; CUP, 2011).

No que se refere ao vocabulário, o projeto se preocupa com o significado de conhecimento de uma palavra, e o considera como um processo cumulativo de aquisição semântica. Assim, a definição dos perfis para os níveis C ainda se encontra em andamento. No entanto, atesta a existência de uma distinção entre o emprego de palavras em sentido literal para os níveis B e figurado para o C (UCLES; CUP, 2011).

Exames de proficiência de reconhecimento internacional, como os produzidos pela empresa americana ETS – *English Testing Service* e pela britânica *Cambridge Assessment English* baseiam-se, assim, nos padrões linguísticos expressos pelo CEFR. O TOEFL iBT – *Test of English as a Foreign Language* consiste em um exame de proficiência desenvolvido pela ETS voltado à avaliação do uso das quatro habilidades da língua para finalidades acadêmicas. De acordo com informações contidas em seu site de divulgação⁴, o insumo fornecido aos candidatos engloba falantes nativos da América do Norte, União Europeia, Nova Zelândia e Austrália, os resultados do teste são válidos por dois anos e a pontuação varia de 0-120 – com o valor máximo de trinta pontos por habilidade. A parte de avaliação da produção oral tem duração total de vinte minutos e é composta por seis tarefas, cada uma com pontuação que varia de zero a quatro e são, posteriormente convertidas para uma escala de zero a trinta pontos. As tarefas, duas independentes e quatro integradas, consistem na expressão de opinião sobre um tópico familiar e na fala com base em tarefas de leitura e compreensão oral. A avaliação do desempenho dos candidatos nas tarefas orais é realizada por avaliadores voltados à análise do desenvolvimento do tópico e da entrega da mensagem. Os níveis de mensuração disponíveis são: fraco (*weak*), de zero a nove; limitado (*limited*), de dez a 17; razoável (*fair*), de 18 a 25; e bom (*good*), de 26 a trinta.

Ainda de acordo com informações do site⁵, a partir de primeiro de agosto de 2019, a seção de produção oral do teste foi reduzida para quatro questões, sendo uma independente e três integradas, com duração total de 17 minutos. Não foram realizadas alterações no valor total

⁴ Disponível em: <https://www.ets.org/toefl/ibt/about>. Acesso em 10 Mar. 2019.

⁵ Disponível em: https://www.ets.org/toefl/better_test_experience. Acesso em 10 Mar. 2019.

de scoring teste ou no valor de scoring individual de cada seção. Em relação ao CEFR, a escala de proficiência do TOEFL iBT varia dos níveis A2 ao C1 (PAPAGEORGIOU *et al.*, 2015; TANNENBAUM; WYLIE, 2008).

O IELTS – *International English Language Testing System*⁶, também produzido pela ETS, consiste em um exame que avalia as quatro habilidades da língua e se apresenta em duas versões: a versão Acadêmica, para finalidades acadêmicas e profissionais, e a de Treinamento Geral, que se estende para processos migratórios. Em ambas as versões, os testes de proficiência oral são os mesmos, apenas os de compreensão escrita diferem.

O teste de produção oral, por sua vez, tem duração de 11 a 14 minutos e ocorre em até uma semana antes ou depois da aplicação do teste de compreensão oral. Os testes são realizados no formato de entrevista individual, gravados e dividem-se em três partes. A primeira tem duração de quatro a cinco minutos e trata de assuntos familiares (opiniões, informações e experiências comuns). Na segunda parte, o candidato recebe um cartão sobre um tópico que indica o assunto sobre o qual ele deve discorrer. O candidato tem um minuto para preparar sua fala e pode, inclusive, fazer anotações durante esse tempo. A resposta deve ter duração de um a dois minutos e, na sequência, o candidato é solicitado a responder uma ou duas questões sobre esse mesmo tópico. A terceira parte retoma o tópico discutido na segunda parte de maneira mais genérica e abstrata. O candidato deve expressar e justificar suas opiniões e analisar, discutir e especular sobre o assunto no tempo de quatro a cinco minutos.

Na avaliação de desempenho, erros ortográficos e gramaticais são penalizados em todas as fases. Os resultados são convertidos e reportados em uma escala de nove faixas que mescla aspectos holísticos e analíticos por meio da atribuição de notas a quatro critérios, todos com o mesmo peso. A nota final da faixa é definida por meio do cálculo de média aritmética simples para os quatro componentes avaliados: compreensão oral, compreensão escrita, produção oral e produção escrita. Em relação ao CEFR, a escala de proficiência do IELTS varia dos níveis B1 ao C1.

Os exames produzidos pela *Cambridge Assessment English*, por sua vez, passaram por alterações de nomenclatura a partir de 2015. O teste de produção oral do FCE - *First Certificate in English*⁷, atualmente conhecido como *B2 First*, é realizado por meio de entrevistas em pares na presença de dois examinadores. O teste tem quatro partes e leva em média 14 minutos por par de candidatos. A primeira parte, com duração de dois minutos, consiste em uma entrevista na qual o candidato interage com o examinador-interlocutor e deve fornecer informações

⁶ Disponível em: <https://www.ielts.org/about-the-test/test-format>. Acesso em 10 Mar. 2019.

⁷ Disponível em: <https://www.cambridgeenglish.org/exams-and-tests/first/>. Acesso em 10 Mar. 2019.

peçoais e expressar sua opinião a respeito de tópicos diversos. Na segunda parte, o candidato deve, em formato de monólogo, estabelecer comparações, descrever, expressar opiniões e especular. O candidato recebe duas imagens e deve discorrer por um minuto sobre elas. Ao outro candidato são disponibilizados trinta segundos para comentar a respeito das imagens apresentadas a seu par. O outro candidato recebe outro par de imagens e a sequência é repetida. A terceira parte consiste em uma atividade de cunho colaborativo. Os candidatos recebem uma tarefa e devem conversar entre si e tomar uma decisão. A tarefa requer troca de ideias, expressão e justificativa de opinião, concordância e discordância, sugestão, especulação, avaliação, decisão por meio de negociação, dentre outros. Tem duração média de três minutos, sendo dois minutos dedicados à discussão e um minuto para tomada de decisão. A quarta e última parte retoma o tópico discutido na parte anterior e, por meio de perguntas feitas pelo examinador, os candidatos devem discutir o tópico expressando e justificando opiniões e expressando concordância ou discordância em quatro minutos.

O teste de produção oral do CAE - *Certificate of Advanced English*⁸, por sua vez, atualmente conhecido como *C1 Advanced*, tem duração total de 15 minutos e também é dividido em quatro partes. A primeira parte consiste em uma entrevista e tem duração total de dois minutos; na segunda parte, o candidato recebe três imagens e deve selecionar duas para comentar em três minutos; na terceira, o candidato recebe uma tarefa sobre a qual deve, juntamente com seu par, discorrer por dois minutos; e na quarta parte, os candidatos têm cinco minutos para conversar e tomar uma decisão em grupo sobre o tópico discutido anteriormente.

No teste de produção oral do CPE - *Certificate of Proficiency in English*⁹, atualmente *C2 Proficiency*, por fim, a duração é ampliada para 16 minutos, mas as partes do teste são reduzidas para três. A primeira parte continua sendo a entrevista, com duração total de dois minutos; a segunda, contudo, passa a ser uma tarefa colaborativa em que os candidatos devem tomar uma decisão em conjunto em quatro minutos; na terceira e última parte, os candidatos devem discorrer sobre um tópico em dez minutos.

A avaliação do desempenho dos candidatos é realizada por uma orientação analítica com base nos critérios de recursos lexicais, recursos gramaticais, gerenciamento discursivo, pronúncia e comunicação interativa – cada um com peso dois – e, de maneira holística, com base no critério de avaliação global, com peso quatro para o *B2 First* e cinco para os demais. Em relação ao CEFR, a escala de proficiência do *B2 First* varia de B1-C1, a do *C1 Advanced* de B2-C2 e a do *C2 Proficiency* de C1-C2 (TAYLOR; JONES, 2006). Os quadros 5 e 6 exibem,

⁸ Disponível em: <https://www.cambridgeenglish.org/exams-and-tests/advanced/>. Acesso em 10 Mar. 2019.

⁹ Disponível em: <https://www.cambridgeenglish.org/exams-and-tests/proficiency/>. Acesso em 10 Mar. 2019.

respectivamente, um resumo das principais características avaliativas dos exames consagrados e a relação dos resultados por eles fornecidos em relação ao CEFR.

Quadro 5. Quadro-resumo das principais características avaliativas dos exames internacionais.

	Formato	Partes	Valor	Total	Escala	Tempo	Tarefa	Tópicos	CrITÉrios	Valor	Pesos	Duração
TOEFL iBT	Individual	1	0-4	0-30 (25%)	0-9 = Fraco 10-17 = Limitado 18-25 = Razoável 26-30 = Bom	2 min.	Independente	Entrevista	Clareza e coerência	-	-	17 min.
		2	0-4			5 min.	Integrada	Exposição e discussão de tópico	Clareza e coerência	-	Iguais	
		3	0-4			5 min.	Integrada		Precisão informacional	-		
		4	0-4			5 min.	Integrada					
IELTS	Individual	1	0-9	0-9 (25%)	Anexo 2 ¹⁰	4-5 min.	Integrada	Apresentação e entrevista	Fluência e coerência	0-9	Iguais	11-14 min.
		2				3-4 min.	Integrada	Exposição de tópico	Recursos lexicais	0-9		
		3				4-5 min.	Integrada	Discussão de tópico	Variedade e precisão gramatical	0-9		
B2 First	Pares	1	0-60	0-60 (20%)	14 = 122 24 = 140 (B1) 36 = 160 (B2) 54 = 180 (C1)	2 min.	Independente	Entrevista	Gramática e vocabulário	0-5	2	14 min.
		2				3 min.	Integrada	Exposição de tópico	Gerenciamento discursivo	0-5	2	
		3				3 min.	Integrada	Tarefa colaborativa	Pronúncia	0-5	2	
		4				4 min.	Independente	Discussão de tópico	Comunicação interativa	0-5	2	
									Avaliação global	0-5	4	
C1 Advanced	Pares	1	0-75	0-75 (20%)	17 = 142 30 = 160 (B2) 45 = 180 (C1) 66 = 200 (C2)	2 min.	Independente	Entrevista	Recursos gramaticais	0-5	2	15 min.
		2				3 min.	Integrada	Exposição de tópico	Recursos lexicais	0-5	2	
		3				3 min.	Integrada	Tarefa colaborativa	Gerenciamento discursivo	0-5	2	
		4				5 min.	Independente	Discussão de tópico	Pronúncia	0-5	2	
									Comunicação interativa	0-5	2	
Avaliação global	0-5	5										
C2 Proficiency	Pares	1	0-75	0-75 (20%)	17 = 162 30 = 180 (C1) 45 = 200 (C2) 66 = 220 (C2)	2 min.	Independente	Entrevista	Recursos gramaticais	0-5	2	16 min.
		2				4 min.	Integrada	Tarefa colaborativa	Recursos lexicais	0-5	2	
		3				10 min.	Integrada	Exposição e discussão de tópico	Gerenciamento discursivo	0-5	2	
									Pronúncia	0-5	2	
									Comunicação interativa	0-5	2	
Avaliação global	0-5	5										

Fonte: Elaborado pela autora.

¹⁰ Disponível em: <https://www.ielts.org/about-the-test/how-ielts-is-scored>. Acesso em 10 Mar. 2019.

Quadro 6. Resultados dos exames em relação ao CEFR.

CEFR	TOEFL iBT	IELTS	B2 First	C1 Advanced	C2 Proficiency
C2	95	9			220-230 Grade A
					213-219 Grade B
C1		7-8		200-210 Grade A	200-212 Grade C
			180-190 Grade A	193-199 Grade B	180-199
B2	72	5,5-6,5	173-179 Grade B	160-179	
			160-172 Grade C		
B1	42	4,5-5	140-159		
A2		4			
A1					

Fonte: Elaborado pela autora.

O quadro-resumo com as mensurações de cada exame em relação aos níveis do CEFR permite visualizar que exames de caráter mais genérico, como o TOEFL iBT, apresentam uma limitação com relação à capacidade de mensuração de especificidades dos níveis C1 e C2, obrigando a escala a integrá-los. Exames com propósitos avaliativos mais específicos, por sua vez, ou seja, subdivididos em níveis, como os fornecidos pela *Cambridge Assessment English*, permitem o apontamento de particularidades dentro de cada um dos níveis-alvos.

2.2 Avaliação da proficiência do professor de línguas estrangeiras: competências e perspectivas

A proficiência do professor de línguas é um dos elementos essenciais das competências de ensinar desses profissional e sua definição é fundamental, portanto, para uma maior compreensão dessas competências (CONSOLO; TEIXEIRA DA SILVA, 2007b), além de ser importante para a ampliação do conhecimento sobre avaliação de línguas e para a formação de professores de línguas (CONSOLO, 2011a, 2011b).

No Brasil, a partir da década de 1990, as competências linguística e comunicativa relacionadas ao professor, mais especificamente, têm sido abordadas a partir de um viés integrador pela divulgação do conceito de competência linguístico-comunicativa (CLC) por Almeida Filho (1993). De acordo com o autor, a CLC “constitui[-se] na competência do professor de produzir linguagem em contextos de uso e comunicação, insumo de qualidade para

que os seus alunos tenham isso como insumo útil para produzir competência eles mesmos” (ALMEIDA FILHO, 1993, p. 38). Assim, tem-se que, com a inserção dos estudos sobre contextos de uso na Linguística Aplicada, despertou-se uma preocupação para a capacidade de o falante atingir os objetivos comunicacionais, utilizando-se, dentre outros elementos, das estruturas linguísticas para tal finalidade, em detrimento do foco que até então era unicamente empregado no conhecimento dos elementos estruturais. No caso do professor, mais especificamente, o conceito de CLC reforça o emprego de elementos de ordem linguística e comunicativa para um objetivo maior: o ensino de línguas.

Consolo (2004) ressalta a importância da consideração do aspecto linguístico, mais especificamente, metalinguístico na constituição da fala do professor. De acordo com Consolo (2007), a proficiência do professor de línguas é composta por uma proficiência de ordem mais genérica, que seria constituída pelo saber empregar a língua em diversos contextos comunicacionais, e por uma proficiência de ordem mais específica, derivada das necessidades da situação específica de uso da língua-alvo (DOUGLAS, 2000; ELDER, 2001) - no caso do ensino de línguas estrangeiras, típica da atribuição de ensinar línguas do professor, ou seja, a metalinguagem. Assim, para o autor, é fundamental que o professor seja capaz de utilizar a LE para explicar a língua que ensina.

Diante de estudos que ressaltam outras situações de uso pedagógico da LE por parte do professor de línguas, a competência específica definida por meio da metalinguagem e apontada em Consolo (2007) passou a ser denominada “competência pedagógica” (CONSOLO, 2011a, 2011b), ampliando, com isso, a possibilidade de inclusão de habilidades além da metalinguística, como o fornecimento de instruções para realização de atividades, o uso de dispositivos retóricos, o uso de estratégias de simplificação, dentre outras típicas da rotina de ensinar línguas (ANDRELINO, 2014; COLOMBO, 2014; ELDER, 2001).

Assim, recentemente, o termo proficiência linguístico-comunicativo-pedagógica do professor de línguas (PLCPP) (CONSOLO, 2011a, 2011b; CONSOLO; TEIXEIRA DA SILVA, 2014, 2016) tem ganhado destaque no cenário nacional, uma vez que ele condensa os principais elementos apontados por pesquisas brasileiras e internacionais sobre as atribuições docentes para o ensino de uma LE por meio da própria LE. O conceito foi apresentado pela primeira vez em Consolo (2011a, 2011b) e divulgado por Consolo e Teixeira da Silva (2014, 2016). Até então, a proficiência do professor de línguas era considerada exclusivamente a partir da perspectiva da proficiência linguístico-comunicativa, como sugerido em Almeida Filho (1993), sendo a instância pedagógica pensada como um fim e, muitas vezes, de forma dissociada – apresentando-se, inclusive, em uma escala à parte no TEPOLI (CONSOLO, 2004)

– e focada na competência metalinguística – excluindo-se, portanto, outros aspectos pedagógicos, como já mencionado. A proposição do novo conceito por Consolo em 2011 preocupava-se basicamente com o refinamento do mesmo e sua posterior incorporação em faixas de proficiência de caráter holístico e analítico do exame EPPL.

Nesse sentido, Consolo (2017) relata uma pesquisa conduzida com especialistas em ensino de línguas estrangeiras, mais especificamente, professores formadores de professores de línguas e selecionadores de professores, como coordenadores e diretores de escolas de línguas. Os resultados apresentados apontam a necessidade de o professor: ter clareza e correção estrutural, conhecimento de vocabulário, conhecimento de regras gramaticais e clareza ao ensiná-las, pronúncia compreensível – não precisa ser similar à de falantes nativos –, agilidade e naturalidade no uso da língua. O estudo também reforçou a crença de que professores de inglês brasileiros não sabem a língua que ensinam, a importância de se definir o propósito e a clientela do teste para que ele seja útil e a importância de se pensar as necessidades dos alunos e professores de línguas estrangeiras no Brasil.

Richards *et al.* (2013) destacam que, de acordo com o *American Council for the Teaching of Foreign Languages* (2002), nos Estados Unidos, altos níveis de proficiência linguística são pré-requisitos para o ingresso em cursos formais de formação de professores de línguas. Para esses autores, professores de línguas com alto nível de proficiência na LE que ensinam são mais capazes de (i.) oferecer explicações significativas aos alunos; (ii.) explorar recursos de ensino-aprendizagem de línguas de forma crítica, permitindo espaço para improvisos e adaptações; (iii.) oferecer insumo linguístico mais rico e apropriado, flexibilizando suas produções orais, de modo a adaptá-las às necessidades de compreensão dos alunos; (iv.) utilizar a língua-alvo para gerenciamento de sala de aula; (v.) responder de forma espontânea e especializada às questões formuladas pelos alunos sobre língua e sobre cultura, dentre outros, do que professores com menores níveis de proficiência. Desse modo, os autores acreditam que quanto maior o nível de proficiência linguística do professor, maior controle esse profissional deterá sobre o desenvolvimento e a operacionalização das demais competências relativas à atividade de ensinar línguas.

Na Linguística Aplicada, considera-se, comumente, a existência de dois tipos de avaliação: a avaliação para finalidades genéricas e a avaliação para finalidades específicas. De maneira geral, independentemente do tipo, todo processo avaliativo apresenta uma finalidade específica; no entanto, a proximidade do construto de um teste com abordagens de cunho comunicativo acaba por situá-lo em um *continuum* de especificidade que o aproxima ou distancia de noções de uso de língua para finalidades gerais ou específicas. Assim, a avaliação

de proficiência linguística do professor de línguas acaba por consistir em uma avaliação para finalidades específicas (ELDER, 2001) dadas as especificidades das competências linguísticas que a integram, a saber o uso da língua para finalidades gerais e para finalidades específicas de ensino (CONSOLO, 2004). A avaliação da proficiência do professor de línguas consiste em um tipo de testagem de língua para finalidades específicas, uma vez que seus conteúdos e métodos são oriundos da análise de situações específicas de uso dessa língua, representando-as com autenticidade. Desse modo, assim como os demais tipos de avaliação para finalidades específicas, sua construção impõe como uma dificuldade de alto nível a definição do grau de especificidade ou genericidade da língua-alvo (DOUGLAS, 2000).

Douglas (2000) ressalta, contudo, que, neste tipo de avaliação, existem dois tipos de construto: aquele que representa a capacidade linguística com finalidades específicas na situação de uso da língua e o construto que será efetivamente avaliado. O primeiro tipo de construto pode ser obtido por meio da observação e análise de situações reais de uso da língua, enquanto que o segundo consiste em um recorte dessa representação que seja passível de mensuração. A partir da situação de uso da língua-alvo ou, mais especificamente, das tarefas que são nela desempenhadas, definem-se os critérios que serão empregados no julgamento do desempenho dos candidatos. Um componente avaliativo consiste, desse modo, (i.) na definição do construto, (ii.) de um conjunto de critérios avaliativos e (iii.) de procedimentos de classificação/pontuação do desempenho. O autor ressalta a importância do emprego de critério autóctones, ou seja, autenticamente representativos da situação real, os quais devem ser elicitados por meio da análise discursiva e ser apresentados, de maneira transparente e explícita, no construto. Os critérios congregam tanto as características da capacidade de uso da língua no contexto real quanto as limitações de mensuração impostas pelo contexto avaliativo. Os critérios consistem, assim, na definição operacional do construto.

Um dos questionamentos apontados por Douglas (2000) acerca da necessidade de elaboração e aplicação de testes para finalidades específicas diz respeito à interferência do conhecimento de tópico na avaliação da capacidade de uso da língua. Assim, se uma determinada tarefa solicita que o candidato tenha conhecimento acerca de um assunto específico que não é de seu domínio, seu desempenho linguístico pode ser prejudicado e, conseqüentemente, sua avaliação final. Desse modo, o autor ressalta a importância de definição de objetivos avaliativos claros e de um recorte da situação real a ser representada no teste de maneira adequada. Na avaliação da proficiência dos professores de língua, por outro lado, faz-se importante ressaltar que, na avaliação da capacidade de uso da língua para finalidades pedagógicas é necessário que o candidato tenha conhecimento acerca dos assuntos abordados.

Isso porque, de acordo com Farrell e Richards (2007), o conhecimento de tópico é capaz de influenciar significativamente a habilidade de ensinar, uma vez que ela é responsável por determinar a capacidade de acessar recursos para ensino de línguas; de modelar o ensino de línguas; de fornecer *feedback* instrutivo; de gerenciar a sala de aula; de oferecer insumo apropriado; de oferecer explicações apropriadas; e de improvisar. Desse modo, a qualidade do conhecimento de tópico e a qualidade da linguagem pedagógica empregada pelo professor estariam intimamente relacionadas.

Rea-Dickins (2001) sugere que, em testes de avaliação de proficiência de futuros professores de inglês, por exemplo, seja essencial a avaliação de elementos como forma, estrutura, conhecimento de regras e adequação no uso da língua, de modo a se verificar se eles dispõem das ferramentas adequadas para a expressão da mensagem. No que concerne à metalinguagem, Clapham (2001) acredita que a avaliação dessa habilidade não pode ser realizada por meio de uma única tarefa. Para a autora é crucial que tarefas de avaliação de metalinguagem analisem a capacidade do candidato de reconhecer o termo gramatical requerido em mais de uma sentença, além de solicitar que ele exemplifique o uso do item. De acordo com a autora, o conhecimento metalinguístico do candidato pode variar a depender da complexidade da sentença e do item gramatical que é apresentado, bem como do termo gramatical e do modo como o conhecimento são solicitados e da língua e do contexto em que o elemento gramatical ocorre.

Por fim, é importante ressaltar que um exame de proficiência como o EPPLÉ não tem a pretensão – e nem deveria ter – de avaliar se um candidato é um bom professor de línguas, ou seja, não é capaz de fornecer inferências sobre o efetivo desempenho profissional de um candidato. Tal impossibilidade justifica-se, de acordo com Douglas (2010), pelo fato de avaliações de proficiência serem capazes de medir unicamente a competência linguística do candidato e nenhuma outra competência mais. Assim, o EPPLÉ corresponde a um instrumento de avaliação útil para fornecer inferências sobre apenas um aspecto das competências que compõem um bom professor de línguas: a competência linguístico-comunicativo-pedagógica, entendendo-se por “pedagógico” não as habilidades didáticas, mas, sim, a capacidade de utilizar a língua(gem) para ensinar por meio, por exemplo, de metalinguagem, do fornecimento de instruções, de *feedback*, de explicação de atividades, dentre outros usos linguísticos no ensino de línguas estrangeiras.

2.2.1 O tratamento da linguagem pedagógica no Cambridge English Teaching Framework no exame de formação de professores de línguas TKT

O CETF – *Cambridge English Teaching Framework*¹¹ consiste em um quadro de referência para línguas, alinhado aos parâmetros do CEFR, específico para as competências do professor de línguas. O quadro divide-se em cinco áreas de conhecimento e apresenta quatro níveis de desenvolvimento, que são, em ordem crescente, o *Foundation*, o *Developing*, o *Proficient* e o *Expert*. Na primeira área, define-se o conhecimento do professor sobre teorias e conceitos sobre ensino-aprendizagem de línguas; a segunda, trata dos princípios de ensino, aprendizagem e avaliação; a terceira aborda a competência linguística; a quarta, do conhecimento e da consciência de linguagem; e a quinta, do desenvolvimento dos valores profissionais (CAE, 2018).

A competência linguística com finalidade pedagógica é definida por meio (i.) da precisão, da fluência e da efetividade do uso de linguagem de sala de aula, definida por meio da introdução de tópicos, do fornecimento de instruções, do fornecimento de explicações linguísticas, do gerenciamento de sala de aula, do fornecimento de respostas aos alunos, dentre outros; (ii.) da capacidade de fornecer modelos de linguagem precisa para diferentes níveis do CEFR, destacando a forma, significado, o uso e a pronúncia da língua-alvo; (iii.) do reconhecimento de erros de aprendizes, em termos de forma, significado, uso e pronúncia; e (iv.) de se comunicar com outros profissionais; (v.) o nível mínimo do CEFR para atuar em cada nível de desenvolvimento, variando de A2 a C2. Na seção destinada à descrição do conhecimento e da consciência de linguagem, aborda-se (i.) a consciência de linguagem em termos de capacidade analítica, (ii.) a terminologia para descrição de línguas, e (iii.) o emprego de recursos de ensino (CAE, 2018).

O TKT – *Teaching Knowledge Test*¹², por sua vez, consiste em um teste de avaliação de conhecimentos do professor de inglês para atuar no ensino de crianças e adultos. A certificação é dividida em três módulos. No primeiro, é avaliado o conhecimento do candidato a respeito de terminologias e conceitos de ensino de línguas; de fatores do processo de aprendizagem de línguas; e das variedades de materiais, tarefas e atividades para o ensino de línguas. No segundo módulo, avalia-se o conhecimento sobre a capacidade e as habilidades necessárias para preparar planos de aula e sobre os tipos de avaliações e recursos que são capazes de guiar o planejamento

¹¹ Disponível em: <https://www.cambridgeenglish.org/teaching-english/professional-development/cambridge-english-teaching-framework/>. Acesso em 10 Mar. 2019.

¹² Disponível em: <https://www.cambridgeenglish.org/teaching-english/teaching-qualifications/tkt/>. Acesso em 10 Mar. 2019.

das aulas. No terceiro módulo, por fim, avalia-se o conhecimento a respeito da linguagem de professor e aluno e de gerenciamento de sala de aula. O teste sugere que o candidato tenha proficiência mínima compatível com o nível B1 do CEFR para realizar a prova. Com relação ao CETF, ao teste se insere nos estágios iniciais de *Foundation* e *Developing* do quadro.

Percebe-se, assim, que o TKT avalia tanto a competência profissional quanto a linguístico-comunicativa do professor de línguas – nos moldes propostos por Almeida Filho (1993). A proficiência linguística de ordem pedagógica é, portanto, especificamente avaliada no terceiro módulo do teste, para o qual são dedicados 1 hora e 20 minutos. O teste conta com 80 questões de caráter objetivo, como associação e múltipla-escolha, sendo 6 tarefas com 40 questões dedicadas à avaliação do conhecimento de linguagem de sala de aula e mais 6 tarefas com 40 questões dedicadas à avaliação da capacidade de gerenciamento. Para cada resposta correta, o candidato recebe 1 ponto.

Na avaliação da linguagem de sala de aula, verifica-se a capacidade do candidato de empregar a linguagem em uma variedade de funções pedagógicas de maneira efetiva, de empregá-la para o fornecimento sequencial de instruções, para identificar as funções da linguagem dos aprendizes e para classificar os erros dos alunos. Na avaliação do gerenciamento de sala de aula, por sua vez, analisa-se a capacidade de atuar enquanto professor em diferentes momentos, como gerenciamento de grupos e fornecimento de *feedback*, de modo a promover o aprendizado.

A avaliação ocorre por meio da identificação das funções utilizadas e dos propósitos comunicativos, da capacidade de identificar e classificar os erros, e de características linguístico-organizacionais, como a adequação da linguagem empregada em termos de sofisticação, estilo, adaptabilidade, dentre outros. Kol (2009) ressalta, contudo, o cuidado que se deve ter ao se empregar padrões avaliativos de exames internacionais em avaliações no contexto brasileiro. Isso porque, segundo o autor, em alinhamento às recomendações recentes do COE (2018), tais parâmetros não são capazes de esgotar as necessidades de diferentes contextos de ensino.

2.3 A ciência da decisão: da neurociência à neuroeconomia

Na história da humanidade, o interesse pelo estudo do funcionamento do cérebro tem suas origens datadas de 3000 a.C., no Antigo Egito, e consiste na primeira descrição do córtex cerebral e da geração de sintomas por lesão craniana realizada pelo “deus” da medicina egípcia, Imhotep. O relato foi escrito em meados de 1700 a.C. no que ficou conhecido como “Papiro Cirúrgico de Edwin Smith” (GROSS, 1987, 2009). No entanto, as civilizações grega e romana

também deixaram contribuições importantes. No primeiro milênio antes de Cristo, o médico grego Hipócrates, o filósofo e matemático grego Platão e o filósofo Aristóteles dedicaram-se ao estudo da relação entre órgãos sensoriais e a percepção, ou seja, entre corpo e alma. Para Hipócrates, considerado “o pai da medicina”, os seres humanos eram compostos por processos materiais, representados pelo corpo, e imateriais, representados pela alma humana, a qual governava todo o corpo. Com base em suas observações, o médico grego atribuiu à alma a responsabilidade pela produção do comportamento humano, essencialmente imprevisível. Platão, por sua vez, rejeitando métodos observacionais ou experimentais, advogou em favor da existência do plano metafísico, correspondente à realidade material e considerado inatingível, e dos sentidos físicos, que seriam gerados pela reflexão do mundo causal. Assim, a alma foi dividida hierarquicamente em três partes: na alma imortal, residiria a cabeça e, por extensão, a razão; na parte superior da alma mortal, o coração, essencialmente o executor da razão; e na parte inferior da alma mortal, o fígado, responsável pelas emoções e desejos animais. Aristóteles, por fim, também atribuiu ao coração um papel fundamental na geração de sensações, do intelecto e dos movimentos (GROSS, 1987, 2009; GLIMCHER, 2003, 2009).

Os estudos sobre o processo decisório que influenciaram a ciência moderna, contudo, têm suas origens atribuídas ao século XVII, mais especificamente à figura do físico, filósofo e matemático francês René Descartes, diante de sua inquietação a respeito da relação existente entre comportamento, cérebro e mente e dos questionamentos apontados por Francis Bacon na mesma época sobre as limitações do conhecimento científico dos gregos e romanos. No entanto, até o século XIX, foram os estudos datados do século II e conduzidos pelo médico e filósofo romano Cláudio Galeno, com base nos princípios sobre o corpo e alma, sobre material e imaterial de Platão e Hipócrates, que formaram a base da neurociência. Para Galeno, o comportamento consistia em fruto da alma imaterial (pneuma). Os estudos de Galeno foram questionados por Andreas Vesalius, médico belga considerado pai da anatomia moderna, no século XVI, por meio da introdução da anatomia experimental (GLIMCHER, 2003, 2009).

Apesar das teorias de Descartes não terem sido provadas empiricamente, o físico, contrariamente ao escolasticismo, defendia que o mundo fosse explicado por meio das leis da física. Assim, apoiando-se na relação entre alma e corpo, Descartes introduz uma visão dualística sobre o comportamento humano segundo a qual a parte simples do comportamento seria determinada e resultado das ações cerebrais reflexivas e a parte complexa seria resultado da alma imaterial e abriria espaço para volição. Para o físico, apenas os comportamentos determinísticos seriam passíveis de investigação científica. Na mesma época, com base no modelo elíptico do sistema solar do astrônomo e matemático polonês Nicolau Copérnico, o

astrônomo e matemático alemão Johannes Kepler introduz a noção de prova geométrica, reforçando a filosofia de que o mundo deve ser determinístico e pode ser descrito por meio da geometria analítica. Na mesma época, o filósofo e matemático alemão Gottfried Wilhelm Leibniz por meio da proposição do sistema de linguagem lógica, e o astrônomo e filósofo inglês Isaac Newton por meio da criação do cálculo, reforçaram o viés da matemática determinista (GLIMCHER, 2003).

No século XVIII, a matemática analítica é o paradigma que passa a dominar os estudos sobre o cérebro determinístico. O matemático, filósofo e engenheiro inglês Charles Babbage é o responsável por desenhar o que seria o primeiro computador para finalidades gerais utilizando-se das leis da física. No entanto, nessa época, as provas matemáticas ainda eram fruto da intuição e da criatividade dos cientistas. Apenas no século XIX, com o matemático e filósofo britânico George Boole e o matemático e filósofo alemão Friedrich Ludwig Gottlob Frege que as ferramentas matemáticas, com o desenvolvimento dos sistemas de lógica, passaram a ser empregadas no exame e na solução de problemas operacionais (GLIMCHER, 2003).

No século XX, contudo, os matemáticos e filósofos britânicos Bertrand Russell e Alfred North Whitehead apontaram a existência de contradições no sistema proposto por Frege e limitações no de Boole. Assim, ao sistema de Frege, incorporaram conceitos avançados da teoria dos conjuntos e defenderam a ideia de que o cálculo da lógica dos predicados poderia ser empregado na reconstrução de todo comportamento determinístico. Na mesma época, o matemático alemão David Hilbert apresentou alguns pressupostos que corroboravam as ideias de Russell e Whitehead. De acordo com Hilbert, a teoria dos números era completa, a matemática era consistente e os problemas matemáticos eram passíveis de decisão (verdadeiros ou falsos) (GLIMCHER, 2003).

A neurociência, enquanto disciplina, surge apenas no século XX, na década de 1960. Seus fundamentos encontram-se na teoria de base biológica do comportamento determinístico proposta por Charles Scott Sherrington, que revela a presença de incertezas como resultantes da interação entre comportamento, cérebro e mente. Dada sua passividade, em 1976, os neurocientistas David Courtnay Marr e Tomaso Poggio propõem uma abordagem em que o sistema é analisado como ponto de partida, sendo, a partir dele, determinado o modo como os objetivos comportamentais foram atingidos e o caminho de produção desses comportamentos para, só a partir de então, elaborar-se uma teoria (GLIMCHER, 2003). Com o surgimento da neurociência e, mais especificamente, com a intensificação dos estudos sobre neurociência cognitiva, psicologia e economia, nasce, em meados da década de 1990, a Neuroeconomia, fruto da interdisciplinaridade (GLIMCHER, 2003, 2009; GLIMCHER; FEHR, 2014;

DAMASIO, 2014). Desde seu surgimento, um dos seus grandes tópicos de interesse consiste nas preferências dos decisores (GLIMCHER *et al.*, 2009) e na busca por maneiras de revelar e orientar o processo de escolha (RUTISCHINI, 2009).

2.3.1 Estruturação de problemas e análise da decisão

A Teoria da Decisão consiste em um ramo da Pesquisa Operacional que trabalha com a escolha da melhor alternativa diante de um problema (CERTO, 2005). Por “problema”, entenda-se qualquer assunto ou questão diante da qual um decisor se encontra cercado por incerteza, dúvida ou dificuldade (FRANCO; MONTIBELLER, 2009; ALMEIDA; MORAIS; ALMEIDA, 2014). De acordo com Almeida, Morais e Almeida (2014), o ato de tomar decisão sempre envolve estar diante de um problema que possui mais de uma alternativa para sua solução. Assim, para os autores, o ato de decidir é equivalente ao de tecer recomendações, levando-se em consideração várias alternativas que induzem a determinado resultado.

Ensslin, Montibeller e Noronha (2001) destacam que, geralmente, decisões são frutos de interações entre preferências de indivíduos ou grupos, ou seja, de atores intervenientes do processo decisório, os quais têm interesses diversos e relevantes e agem diretamente por meio de seus sistemas de valores. Além dos atores intervenientes, os autores informam, ainda, a existência de atores passivos, ou seja, aqueles que não participam ativamente da decisão, mas que, contudo, também precisam ser considerados no processo decisório por serem afetados pelas decisões.

Assim, por envolver o sistema de valores dos atores do processo, a análise da decisão envolve, inevitavelmente, aspectos subjetivos – como objetivos, preconceitos, cultura e intuição –, uma vez que cada decisor percebe e interpreta de forma distinta o contexto decisório. Por isso, cada ator do processo tem uma representação mental individual do problema. Assim, no caso de decisões em grupo, existem vários problemas sobre um mesmo contexto decisório, sendo necessário que se empreguem estratégias de modelagem para se chegar a um único problema do grupo, ou seja, a um problema que seja aceito pelos atores, levando-se em consideração suas representações individuais (ENSSLIN; MONTIBELLER; NORONHA, 2001). Simon (1965), contudo, ressalta a racionalidade limitada do ser humano como sendo responsável por gerar modelos simplificados da realidade. Assim, para o autor, em um dado instante, muitos aspectos são irrelevantes ou não podem ser considerados graças à racionalidade limitada, gerando problemas mais ou menos estruturados, já que a racionalidade atua na seleção das alternativas ao problema.

Certo (2005) e Robbins, Judge e Sobral (2010) destacam que o processo de tomada de decisão é uma atividade passível de erros, pois ela é afetada pelas características pessoais e pela percepção do tomador de decisões. Desse modo, um processo estruturado não é garantia de sucesso, mas permite uma decisão lógica, coerente e, por extensão, minimiza as chances de ocorrência de erros.

Estudos do campo da teoria da decisão comumente dividem-se em duas abordagens de racionalidade relacionadas, mas distintas com relação às finalidades e à interpretação dos processos que integram o sistema decisório: a tomada da decisão, de base racional – Escola Americana – e a análise da decisão, de base não racional – Escola Francesa. A tomada da decisão trabalha com um paradigma normativista, de modo que as soluções propostas apresentam finalidades otimizatórias. O decisor é visto como um ser de capacidade cognitiva ilimitada e coerente, não influenciado por limitações de qualquer ordem e, portanto, suas soluções propostas são entendidas como absolutamente objetivas (WANG; TRIANTAPHYLLOU, 2008).

Por outro lado, a análise da decisão, embasamento teórico-metodológico deste trabalho, consiste em um ramo que, diferentemente da tomada de decisão, acaba por não trabalhar com conceitos de otimização e objetividade absoluta por considerar a subjetividade um elemento essencial para o processo decisório. Assim, trabalha com preceitos que envolvem limitações da mente humana em termos de conhecimento, tempo e memória. Tem por objetivo propor uma maneira estruturada de se pensar sobre os problemas e o contexto decisório no qual eles se inserem e, a partir dessa estrutura, desenvolver julgamentos subjetivos, na qualidade de opiniões e sensações, críticos a uma decisão de qualidade (CLEMEN; REILLY, 2013).

Clemen e Reilly (2013) destacam que a análise da decisão é composta por três etapas. Assim, inicialmente deve-se identificar a situação decisória e delimitar seus objetivos. Seguindo-se os princípios de Keeney (1992), os autores reforçam a importância de primeiramente se pensar nos valores e objetivos centrais e só então nas oportunidades de decisão. Os autores consideram como valores aquilo que é importante para o problema decisório, aquilo com o qual o decisor se importa, ou seja, o motivo pelo qual a decisão precisa ser tomada. Os objetivos, por sua vez, seriam as especificidades – relacionadas ou não – que definem esses valores e as quais o decisor busca alcançar. Assim, os autores destacam a importância do conhecimento dos objetivos na determinação de alternativas, no modo de mensuração dos resultados e na determinação das incertezas a serem consideradas.

A etapa seguinte consiste na modelagem do problema decisório, por meio da decomposição do problema em partes menores e mais gerenciáveis para que sejam

compreendidas sua estrutura e para mensuração de seus valores e incertezas. A modelagem consiste em uma etapa crítica para a análise da decisão e utiliza-se de diagramas de influência ou de árvores de decisão para representar o problema. De maneira geral, esta etapa preocupa-se em garantir uma maior compreensão do problema e de seus aspectos.

Diagramas de influência consistem em uma maneira de representar o problema decisório e de exibir a relação que existe entre os elementos que o integram, funcionando como uma fotografia de uma situação decisória particular, num dado momento da análise. São representados por meio de quatro nós (*nodes*), apresentados em um gráfico e conectados por setas e arcos, os quais representam sequências ou relevância. Decisões são representadas por retângulos (nós de decisão – *decision nodes*), eventos possíveis são representados por elipses (nós de possibilidades – *chance nodes*) e a consequência final ou o nó de troca são representados por diamantes (nós de troca – *tradeoff nodes*). Valores constantes ou cálculos matemáticos – consequências intermediárias – são apresentados por meio de retângulos de cantos arredondados (nós de consequências ou nós de cálculos – *consequence nodes* ou *calculation nodes*).

As árvores de decisão, por sua vez, permitem a apresentação de maiores detalhes acerca do problema decisório e representam todos os caminhos possíveis que podem ser tomados pelos decisores. Por esse motivo, é interessante que sejam apresentadas de acordo com uma sequência temporal - da esquerda para a direita. Similarmente aos diagramas de influência, quadrados são empregados na representação de decisões a serem tomadas e círculos representam os eventos. Os ramos que emanam dos quadrados representam as escolhas disponíveis ao decisor e devem ser apresentados de modo que o decisor possa escolher apenas uma opção. Os que emanam dos círculos representam os resultados possíveis de um evento possível e devem conceber resultados que sejam mutuamente excludentes – apenas um dos resultados pode ocorrer – ou coletivamente completos – não há outra possibilidade, um dos resultados deve ocorrer. Ao final dos ramos, são apresentadas as consequências.

O modelo gerado é essencialmente de natureza matemática e gráfica, possibilitando a análise. Assim, emprega-se probabilidade para o cálculo de incertezas/das chances de um evento ocorrer, modelos de rede ou hierárquicos para compreender a relação entre os objetivos, e funções de utilidade para entender a valoração dos resultados.

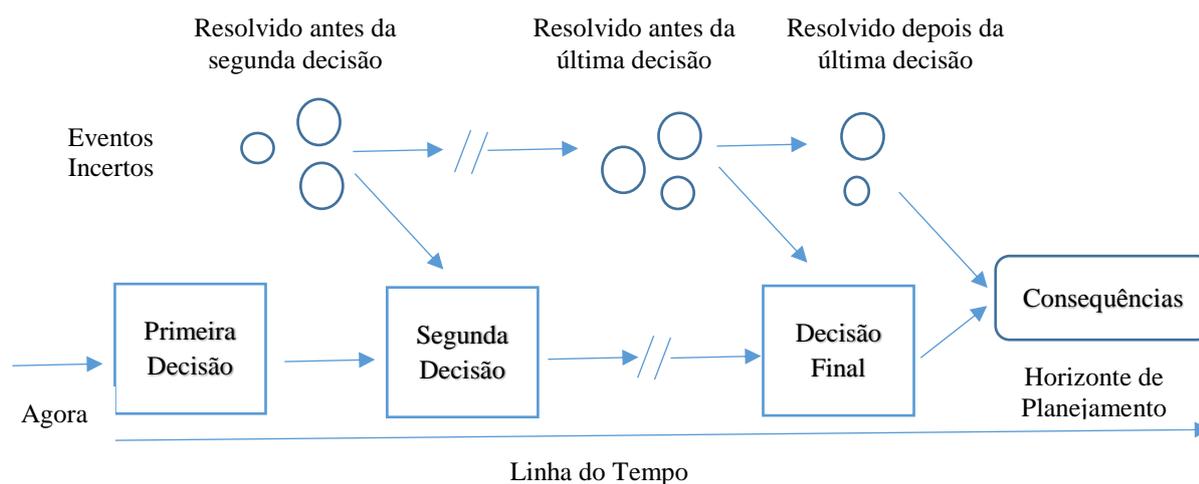
A incerteza é tratada de modo que, inicialmente, seus resultados devem ser estabelecidos e relacionados e, na sequência, deve-se verificar de que modo esses resultados impactam ou não nos objetivos delimitados na etapa anterior. Considera-se relevante e inclui-se na avaliação do processo apenas os resultados que apresentam algum impacto em pelo menos um dos

objetivos delimitados. Se possível, a solução para as incertezas deve ser apresentada de acordo com a sequência temporal das decisões que importam ao processo, em um horizonte de planejamento (*planning horizon*). Incertezas que não são importantes para o problema ou que não podem ser solucionadas devem permanecer explícitas, apesar de não integrarem a solução.

A última etapa consiste na obtenção da solução para o problema em si, que não tem caráter normativo, mas, sim, sugestivo. A solução consiste na decisão – imediata ou sequencial, para processos dinâmicos – tomada a respeito do problema por meio da avaliação das consequências que as alternativas impõem aos objetivos. Em problemas que requerem múltiplos objetivos, deve-se valorar as consequências de escolhas das alternativas possíveis para cada uma e considerar as alternativas que envolvem trocas (*trade-offs*) relevantes.

A Figura 4 condensa as etapas necessárias ao processo de análise da decisão.

Figura 4. Etapas do processo de análise da decisão.



Fonte: Traduzido de Clemen e Reilly (2013, p. 32)

A análise da decisão consiste em um processo iterativo, ou seja, após a obtenção da solução, deve-se realizar a análise de sensibilidade, que consiste na verificação de mudanças na solução gerada diante de pequenas alterações nas informações de entrada, percebendo-se, assim, o impacto que potencial de cada variável no modelo. A análise de sensibilidade, de acordo com Clemen e Reilly (2013), não precisa ser expressa quantitativamente e pode ser aplicada, inclusive, durante a fase de estruturação de problemas. No entanto, para que o modelo seja considerado robusto, ele deve resultar em resposta afirmativa à pergunta “estamos resolvendo o problema correto, real?”. Uma resposta negativa a essa questão faz com que o pesquisador incorra em um erro tipo III – em paralelo aos erros tipo I e II estatísticos. Caso

alterações sejam percebidas de maneira significativa, deve-se retornar às etapas do ciclo de análise da decisão e reavaliá-lo.

Durante a iteração, é natural que haja mudanças na percepção, por parte dos decisores, a respeito da natureza e da estrutura do problema, de suas opiniões e preferências. No momento da decisão, também é importante que o decisor pondere sobre quão extensa deve ser sua consideração temporal, levando-se em consideração o impacto que esses eventos e decisões futuras geram na decisão imediata, garantindo consistência ao contexto decisório e aos objetivos relevantes.

2.3.1.1 Métodos de Estruturação de Problemas (PSMs)

Como o próprio nome sugere, métodos de estruturação de problemas consistem em uma abordagem empregada desde meados da década de 1960 que têm o propósito de auxiliar a estruturação de problemas – e, por extensão, garantir uma melhor compreensão dos mesmos – de modo a propiciar a geração de consenso ou de facilitar a negociação em grupos (ROSENHEAD, 1996). Esses métodos surgiram em resposta às técnicas tradicionais de solução de problemas da Pesquisa Operacional, as quais empregavam procedimentos que visavam à otimização do problema, tais como representações com níveis hierárquicos rígidos.

Franco e Montibeller (2009) destacam a importância de se estruturar o problema, mais especificamente, de definir o problema e delimitar o escopo de participação de modo a se obter uma melhor compreensão da oportunidade decisória. Assim, retomando Eden e Sims (1979), os autores destacam a fundamentalidade dos métodos de estruturação de problemas especialmente para casos em que o problema não é facilmente estruturável, ou seja, em casos nos quais a estruturação é constantemente negociada entre os decisores.

De acordo com Rosenhead (1996), métodos de estruturação de problemas são vantajosos por garantirem uma melhor compreensão do problema por meio da geração de sistemas de representação a partir de interação e iteração, geralmente por meio de mapas cognitivos ou correlatos, baseados nas relações que existem entre as múltiplas perspectivas trabalhadas, evitando, assim, a resolução do “problema errado” (FRANCO; MONTIBELLER, 2009). Segundo Rosenhead e Mingers (2001) e Mingers e Rosenhead (2004), situações-problema não estruturadas são marcadas por múltiplos atores, múltiplas perspectivas, interesses conflitantes ou incomensuráveis, intangíveis importantes e incertezas-chave. Nesse sentido, Mingers e Rosenhead (2004) argumentam que essas situações não constituem originalmente problemas, mas, situações problemáticas em virtude unicamente da ausência de estruturação.

Diversos são os métodos de estruturação de problemas existentes, como SODA (*Strategic Options Development and Analysis*), SSM (*Soft Systems Methodology*), SCA (*Strategic Choice Approach*), DT (*Drama Theory*), dentre outros. No entanto, dentro da abordagem dos métodos de estruturação de problemas, Keeney (1992) destaca a existência de duas vertentes: uma, mais tradicional, com foco nas alternativas geradas pela situação-problema, denominada, pelo autor de *Alternative-Focused Thinking* (AFT) e outra, mais contemporânea, com foco no valor, o *Value-Focused Thinking* (VFT). Métodos como SODA, SSM, SCA e DT pertencem à vertente tradicional e estruturam seus problemas a partir da consideração das alternativas possíveis para compreensão do mesmo. O método que trabalha com a estruturação do problema a partir do valor e, por extensão, permite o delineamento de oportunidades de decisão e a geração de alternativas, é o VFT (KEENEY, 1992, 1996).

Por valor, entenda-se aquilo com o que o decisor se importa em um contexto decisório e é representado por meio de objetivos, que se apresentam pela estrutura “verbo + objeto”, como, por exemplo, “minimizar custos” (MORAIS et al, 2013). As alternativas, por sua vez, consistem nas possibilidades existentes para resolver a situação-problema e, de acordo com Keeney (1996), são geradas a partir da definição dos valores. O autor ressalta que, em abordagens de decisão estratégica, ou seja, aquelas que têm consequências importantes para os envolvidos, a definição de objetivos é um processo fundamental. No entanto, destaca que a listagem de objetivos não consiste em uma atividade simples, pois os envolvidos geralmente confundem objetivos meios, fins e fundamentais. Do mesmo modo, o autor destaca a fundamentalidade de se pensar em valores anteriormente à análise das alternativas, pois são estes valores muitas vezes os responsáveis por recriarem alternativas não-óbvias à situação-problema.

No caso da abordagem de pensamento focado no valor, da qual o método *Value-Focused Thinking* (VFT) proposto por Keeney (1992) é o representante, a estruturação de valores não ocorre apenas por meio da listagem dos objetivos, mas, sim, da categorização em objetivos estratégicos, fundamentais, meios e fins e, ainda, por meio da explicitação dos modos de mensuração dos alcances desses objetivos. A listagem dos objetivos é o primeiro passo da estruturação, seguida da organização desses objetivos em categorias maiores que os representem e que representem os valores dos decisores. Na sequência, os objetivos são separados nas categorias propostas por Keeney (1992) e retomada em Keeney (1996).

Assim, de acordo com Keeney (1992, 1994, 1996), o VFT estabelece uma distinção entre quatro tipos de objetivos: os objetivos estratégicos, os objetivos fundamentais, os objetivos meios e os objetivos fins. Os objetivos estratégicos consistem naqueles que

direcionam todo o processo de tomada de decisão. Os fundamentais, de acordo com Bana e Costa (1992), consistem em critérios, ou seja, representam os fins que os tomadores de decisão almejam em um contexto decisório e apresentam relação divisional, ou seja, subdividem-se em outros objetivos, os meios. Desse modo, os objetivos fundamentais englobam aqueles que refletem exatamente aquilo que se quer alcançar e são especificados por meio de objetivos meios, que auxiliam no alcance dos fundamentais. Os objetivos fundamentais são organizados de maneira hierárquica, do mais geral para os mais específicos, e os objetivos meios são organizados em redes (*networks*). Os objetivos fundamentais das camadas inferiores configuram-se na base sobre a qual as consequências são valoradas. Já os objetivos meios consistem em informações importantes à análise, especialmente quando são de fácil mensuração se comparados aos fundamentais que eles representam. Esses últimos objetivos, por sua vez, são formas, maneiras de se alcançar outros e apresentam uma relação causal, ou seja, eles são o motivo pelo qual um objetivo fundamental existe.

De acordo com Keeney (1992, 1996), para se encontrar um objetivo meio, deve-se realizar a pergunta “por que esse objetivo é importante?”. Se a resposta for “para atingir um objetivo maior”, ele é um objetivo meio; mas, se a resposta for, “porque sim”, ele é um objetivo fundamental. Os objetivos fins representam, então, a agregação das finalidades do objetivo meio. É a partir da rede de objetivos meios que as alternativas são obtidas.

O Quadro 7 resume as etapas de construção da rede de objetivos meios e da hierarquia de fundamentais:

Quadro 7. Etapas de construção da rede de objetivos meios e da hierarquia de objetivos fundamentais.

	OBJETIVOS FUNDAMENTAIS	OBJETIVOS MEIOS
Para mover-se: Pergunte:	<i>Para baixo na hierarquia “O que você quer dizer com isso?”</i>	<i>De encontro aos objetivos fundamentais “Como isso pode ser atingido?”</i>
Para mover-se: Pergunte:	<i>Para cima na hierarquia “Este é um aspecto de qual objetivo mais geral?”</i>	<i>Ao encontro dos objetivos fundamentais “Por que isso é importante?”</i>

Fonte: Traduzido de Clemen e Reilly (2013, p. 52).

A escala de mensuração dos objetivos é conhecida como escala de atributos (*attributes*) e subdivide-se em escala de atributos naturais – aqueles que usual e consentidamente é empregado para mensuração -, escala de atributos proxy ou indiretos – que são criados para medir o objetivo – e escala de atributos de alcance dos objetivos, por meio de níveis. De acordo com Bana e Costa (1992), atributos são, na verdade, descritores dos critérios.

Assim, o VFT consiste em um método híbrido, ou seja, de etapa qualitativa e quantitativa, as quais podem ser aplicadas separadamente. De acordo com Almeida, Morais e

Almeida (2014), este método essencialmente constitui-se de duas atividades: decidir o que você deseja e então descobrir como alcançá-lo. Assim, ele fornece uma forma estruturada de pensar sobre as decisões, de desenvolver e de apoiar julgamentos subjetivos que são fundamentais para decisões eficientes.

O VFT consiste em um método bastante empregado nas avaliações de caráter ambiental, como por exemplo, na construção de hidrelétricas (KEENEY; MCDANIELS, 1992; CUOGHI, 2015), nas investigações sobre os objetivos de grupos terroristas (KEENEY; VON WINTERFELD, 2009; SIEBERT; VON WINTERFELD; JOHN, 2015), dentre outros. Embora ainda não se tenha relatos de emprego do VFT na Linguística Aplicada, nota-se que o método consiste em uma estratégia de grande utilidade para a elaboração de rubricas e escalas de proficiência, uma vez que trabalha com diversos níveis de análise ao desmembrar os objetivos de acordo com sua função, permitindo uma maior compreensão do construto de língua empregado pelo exame e menos subjetividade no processo de medição por meio da geração organizada de critérios e atributos.

Franco e Montibeller (2009), bem como outros autores, como Cuoghi (2015) e Cuoghi e Leoneti (2018), destacam, contudo, a necessidade de se realizar a etapa de estruturação para se articular e definir o problema multicritério. Assim, esses autores sugerem o emprego de técnicas de estruturação de problemas em articulação com métodos MCDA para uma melhor abordagem das oportunidades decisórias.

2.3.1.2 Métodos Multicritério de Análise da Decisão (MCDA)

Os métodos MCDA são amplamente empregados em diversos estudos de avaliação, como os do setor de sustentabilidade, bélico, sistemas de Pós-Graduação, etc. (JOERIN *et al.*, 2001; DIAS; CLIMACO, 2000; MIRANDA; ALMEIDA, 2003; ALMEIDA; COSTA, 2003). De acordo com Zopounidis, Pentaraki e Doumpos (1998), métodos MCDA são instrumentos de análise da decisão flexíveis, pois permitem a incorporação de preferências à análise, lidam com fatores de ordem qualitativa e são capazes de se atualizarem, ou seja, de se adaptarem às mudanças do contexto dinâmico de decisão. Munda (1993) ressalta, ainda, que métodos MCDA consistem em técnicas bastante úteis no trabalho com problemas que envolvem incertezas por permitirem o tratamento da subjetividade.

Para problemas de decisão multicritério, a matriz de decisão consiste em uma representação, em linhas e colunas, da relação existente entre critérios e alternativas, como exibido na Figura 5. Os critérios empregados na mensuração do desempenho das alternativas são representados em colunas (C_1, C_2, \dots, C_n) e as alternativas de decisão, opções de escolha

pelos decisores, são apresentadas nas linhas (A_1, A_2, \dots, A_m). Os pesos de cada critério são apresentados fora da matriz inicial (W) ou abaixo da última alternativa.

Figura 5. Modelo de matriz de decisão

$$D = \begin{matrix} & C_1 & C_2 & \cdots & C_n \\ A_1 & x_{11} & x_{12} & \cdots & x_{1n} \\ A_2 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A_m & x_{m1} & x_{m2} & \cdots & x_{mn} \end{matrix} ,$$

$$W = [w_1 \ w_2 \ \cdots \ w_n],$$

Fonte: Extraído de Chen (2000, p. 01).

Desse modo, na Figura 5, tem-se que x_{ij} representa o valor da alternativa A_i com relação ao critério C_j . O peso do critério C_j é representado por w_j .

A escolha dos métodos MCDA mais adequados para a resolução de um determinado problema depende do tipo de contexto decisório envolvido. Uma vez que não existe um método que seja capaz de lidar com todos os problemas de maneira adequada, é importante que a escolha do método não se limite a justificativas que envolvam familiaridade e afinidade com um método específico (GUITOUNI; MARTEL, 1998).

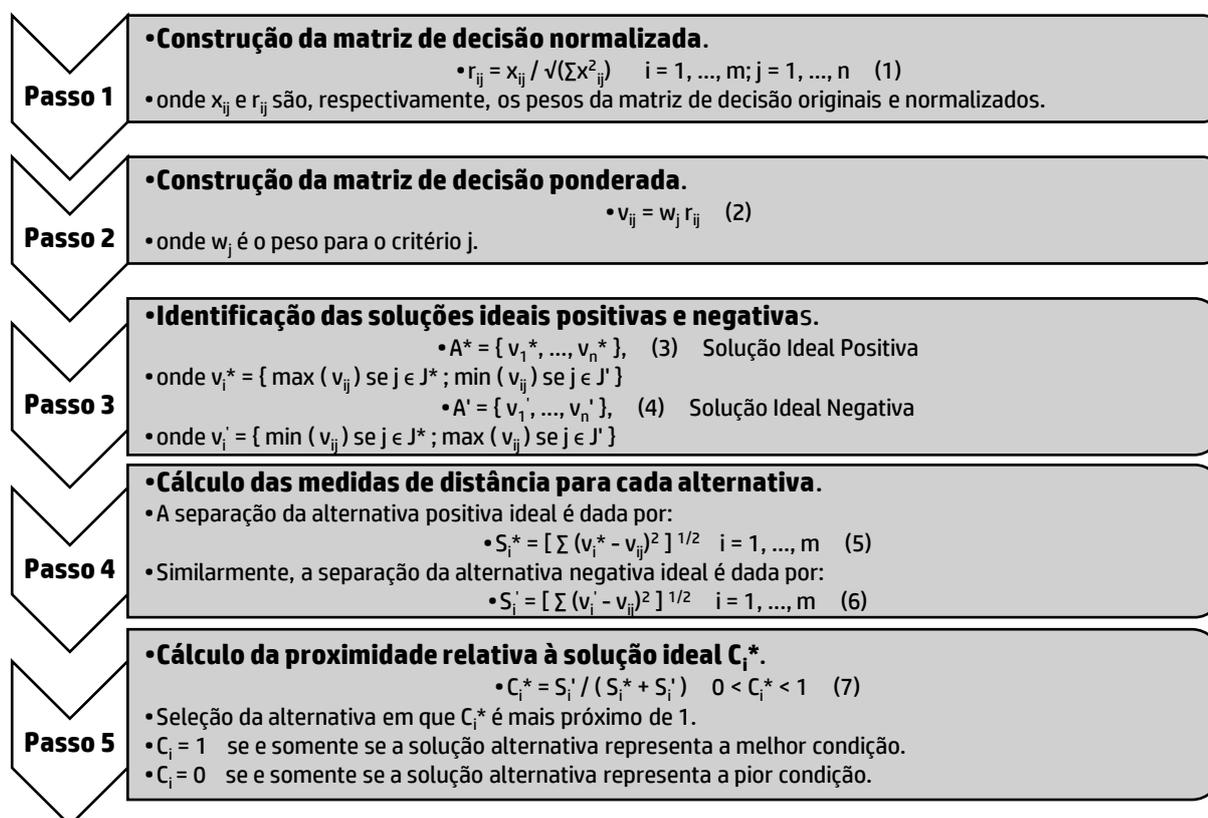
2.3.1.2.1 Técnica para Avaliar o Desempenho de Alternativas através de Similaridade com a Solução Ideal: TOPSIS e Fuzzy TOPSIS

O método TOPSIS (*Technique for Order of Preference by Similarity to Ideal Solution*, em inglês) consiste em um método de ranqueamento de preferências que, baseado na geometria de distância euclidiana, busca fornecer a melhor alternativa para um determinado problema de modo que essa solução seja o mais próxima possível da solução positiva ideal e o mais distante possível da solução negativa ideal. A solução é atingida com base na comparação de um conjunto de alternativas por meio da identificação de pesos atribuídos por um decisor a cada critério. Assim, a solução positiva ideal é composta pela maximização de critérios de benefício e a minimização de critérios de custo; ao passo que a solução negativa ideal consiste na maximização de critérios de custo e na minimização de critérios de benefício (ARESE *et al.*, 2017).

O método funciona com variáveis objetivas e quantitativas, desde que seus valores sejam representados numericamente, que os atributos aumentem ou diminuam uniformemente e que sejam passíveis de mensuração (BEHZADIAN *et al.*, 2012) e consiste em um algoritmo compensatório. Desse modo, garante uma modelagem mais realista ao permitir a ocorrência de trocas entre os critérios, de modo que um resultado insuficiente em um atributo pode ser negado graças a um resultado satisfatório em outro (ZAVADSKAS *et al.*, 2016).

Desenvolvido originalmente por Hwang e Yoon (1981) – e adaptado em Yoon (1987) e Hwang, Lai e Liu (1993) –, a aplicação do TOPSIS divide-se em sete etapas, as quais encontram-se representadas na Figura 6.

Figura 6. Passos para aplicação do método TOPSIS clássico.



Nota: O símbolo * corresponde ao valor máximo. O símbolo ' corresponde ao valor mínimo.

Fonte: Tradução livre de Behzadian *et al.* (2012, p. 13052) e de Zavadskas *et al.* (2016, p. 09).

Conforme descrito em Behzadian *et al.* (2012) e em Zavadskas *et al.* (2016), a primeira etapa consiste na normalização linear ou vetorial da matriz de decisão. É válido ressaltar que o conceito de normalização empregado pelos métodos de análise de decisão não coincide com o conceito estatístico usual. Na matemática, de maneira geral, o ato de normalizar uma variável consiste na transformação do valor dessa variável em distribuição normal. No caso do TOPSIS

e de outros métodos MCDA, o processo de normalização é equivalente ao de padronização, ou seja, por normalização entenda-se o processo de padronizar os valores da matriz de decisão, de modo que todos estejam na mesma escala de medida – no caso do TOPSIS, os valores devem estar entre 0 (zero) e 1 (um). A normalização é realizada dividindo-se um a um os valores originais da matriz (x_{ij}) pela soma, por critério, desses mesmos valores ao quadrado – como exibido pela Equação 1.

Na segunda etapa, os valores normalizados de cada coluna são multiplicados pelo vetor de peso correspondente subjetivamente atribuído pelos decisores – conforme exibido pela Equação 2. Forma-se, então, uma nova matriz decisória por meio da qual será possível realizar a terceira etapa, que consiste em verificar os melhores e os piores desempenhos de cada critério e separá-los em alternativa ideal positiva e negativa. O conjunto de solução ideal positivo ocorre em duas condições: (i.) se o valor é máximo e o critério pertence aos critérios de benefício e (ii.) se o valor é mínimo e o critério pertence aos critérios de custos – conforme exibido na Equação 3. Similarmente, o conjunto de solução ideal negativa ocorre (i.) se o valor é mínimo e o critério pertence aos critérios de benefício e (ii.) se o valor é máximo e o critério pertence aos critérios de custos – conforme exibido na Equação 4.

A quarta etapa consiste no cálculo da função de distância euclidiana para cada alternativa. As Equações 5 e 6, condensam o processo de cálculo da distância da alternativa-alvo em relação à melhor e à pior condição, respectivamente, em uma escala que varia de 0 (zero) a 1 (um). A quinta e última etapa – representada pela Equação 7 – consiste na seleção do valor da solução ideal, que deve ser o mais próximo de 1 (um), e no ranqueamento das alternativas com relação a este valor.

Graças à simplicidade de seu processo computacional, o TOPSIS consiste em um algoritmo de fácil implementação e entendimento, o que permite sua programação em planilhas eletrônicas. O método também apresenta característica intuitiva, sendo, portanto, solidamente embasado na racionalidade da escolha humana. Contudo, faz-se necessário ressaltar que o método é propenso à reversão de ordem (*rank reversal*), ou seja, a inserção ou remoção de alternativas não previstas no início do processo decisório pode desencadear uma modificação na ordem das alternativas sugeridas (GARCÍA-CASCALES; LAMATA, 2012). No entanto, mesmo diante desta inconveniência, o TOPSIS consiste, atualmente, no segundo método multicritério de análise da decisão mais utilizado, estando atrás apenas do AHP (ZAVADSKAS *et al.*, 2016) – que também sofre com problemas de reversão.

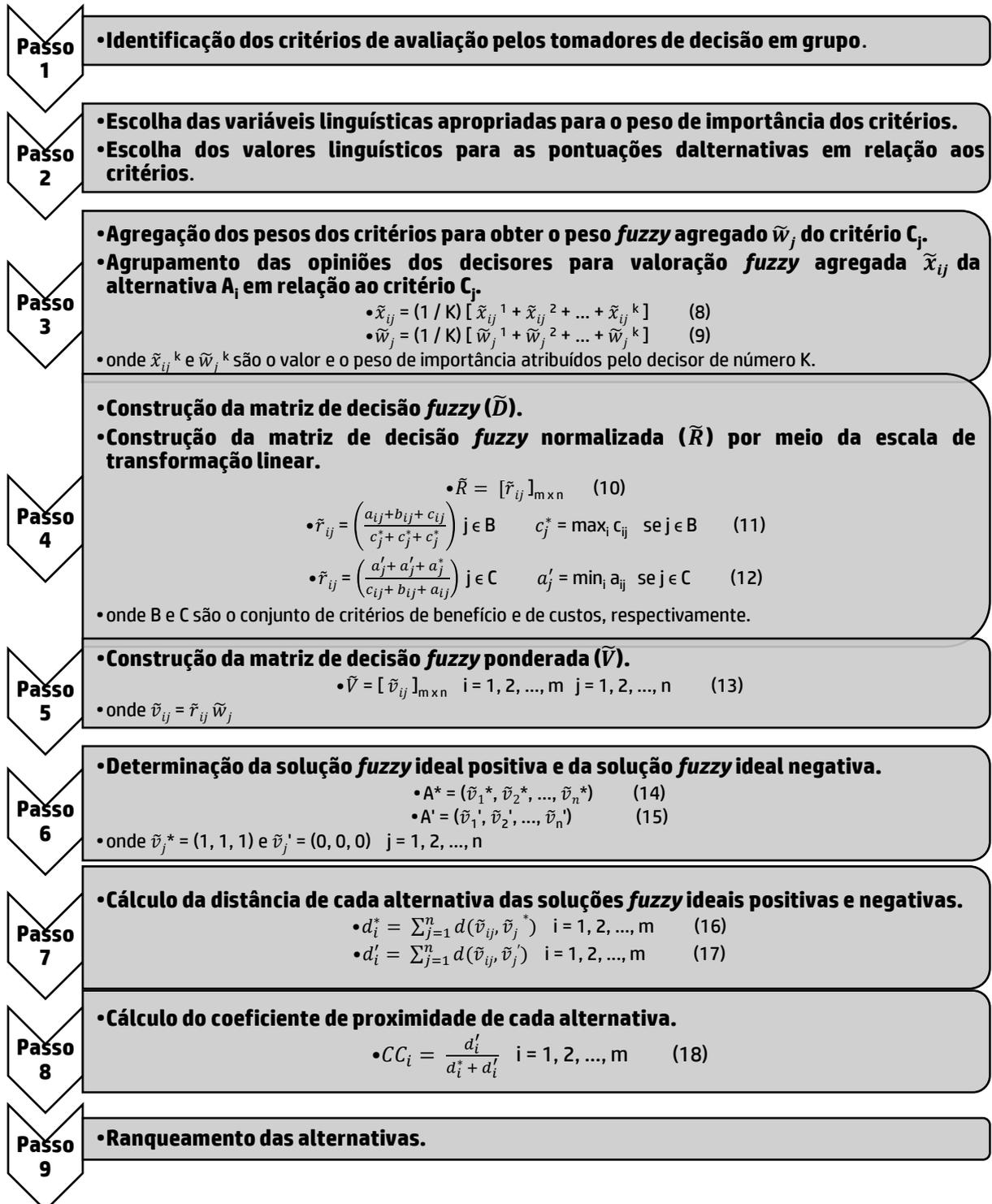
Além das reformulações realizadas por seus criadores – Yoon (1987) e Hwang, Lai e Liu (1993) – Behzadian *et al.* (2012) e Zavadskas *et al.* (2016) ressaltam uma série de extensões

realizadas por diferentes pesquisadores de modo a acomodar melhor suas necessidades de pesquisa, revelando a flexibilidade do algoritmo. Os autores informam que o método tem sido empregado de maneira eficiente em diversas áreas de conhecimento, como gerenciamento de recursos humanos; gerenciamento de transportes; construção; gerenciamento e logística de cadeias de fornecimento; design, engenharia e manufatura de sistemas; gestão hídrica; gerenciamento de negócios e de marketing; análises de localização; computação soft; controle de qualidade; pesquisa operacional; engenharia química; gerenciamento de meio ambiente, saúde e segurança; design de produtos; gerenciamento energético; e pesquisas sociais, em esporte e em educação.

Apesar da popularidade do método, o TOPSIS não consegue tratar adequadamente a incerteza e a imprecisão inerentes ao processo de tomada da decisão em uma grande variedade de situações de decisão na vida real (CHEN, 2000; KRHOLING; CAMPANHARO, 2011). Assim, dentre as modificações realizadas, encontra-se a técnica conhecida como Fuzzy TOPSIS que, em vez de estimar julgamentos pessoais utilizando-se de valores numéricos exatos (*crisp*), possibilita o emprego de variáveis linguísticas que podem ser convertidas em números difusos (*fuzzy*). A primeira proposta foi elaborada por Chen (2000), que combinou o TOPSIS com a teoria dos conjuntos difusos proposta por Zadeh (1965) para problemas que envolvem a decisão em grupo em ambientes *fuzzy*. Desse modo, o método proposto por Chen (2000) adaptou o algoritmo original do método para lidar com números difusos e empregou variáveis linguísticas para avaliar os pesos de cada critério e para valorar as alternativas em relação aos critérios. Outras extensões foram realizadas por diversos autores, embora o método de Chen (2000) ainda tenha maior visibilidade graças à sua simplicidade e facilidade de entendimento (LIMA JUNIOR, 2013).

O método combinado proposto por Chen (2000) encontra-se sintetizado na Figura 7 e consiste em 9 passos.

Figura 7. Passos para aplicação do método Fuzzy TOPSIS proposto por Chen (2000).



De acordo com Chen (2000), nesse modelo, inicialmente, a avaliação das alternativas em relação aos critérios e a atribuição de pesos por parte dos decisores é realizada utilizando-se de variáveis linguísticas previamente definidas. O peso de importância dos critérios pode ser atribuído diretamente ou por comparação de critérios.

Na sequência, as variáveis linguísticas são convertidas em números *fuzzy* triangulares positivos, cuja função de pertencimento é definida por $\mu_{\tilde{n}}(x)$, como exibido na Equação (19), e os quais se apresentam em tríade $\tilde{n} = (n_1, n_2, n_3)$.

(19)

$$\mu_{\tilde{n}}(x) = \begin{cases} 0 & x < n_1 \\ \frac{x-n_1}{n_2-n_1} & n_1 \leq x \leq n_2 \\ \frac{x-n_3}{n_2-n_3} & n_2 \leq x \leq n_3 \\ 0 & x > n_3 \end{cases}$$

A Equação 19 permite calcular as fronteiras dos limites da área passível de avaliação. Para tanto, consideram-se n_1 , n_2 e n_3 números reais, de modo que $n_1 < n_2 < n_3$. Assim, o valor de x em n_3 representa a nota máxima da função de pertencimento do conjunto *fuzzy* \tilde{n} em universo discursivo X , ou seja, representa o valor mais provável ($\mu_{\tilde{n}}(x) = 1$). O valor de x em n_1 , por sua vez, representa a nota mínima dessa função, ou seja, o valor menos provável ($\mu_{\tilde{n}}(x) = 0$). Desse modo, tem-se que quanto menor o intervalo $[n_1, n_3]$, menor a difusidade dos dados.

Antes de dar início à elaboração da matriz, é necessário que seja realizado uma agregação das preferências dos decisores, conforme exposto pelas Equações 8 e 9 na Figura 7.

A matriz de decisão e os pesos *fuzzy* são definidos e normalizados em uma escala $[0,1]$ por meio da escala de transformação linear, conforme exibido pelas Equações 10, 11 e 12 da Figura 7, para possibilitar, na sequência, a geração da matriz de decisão *fuzzy* ponderada, como exibido na Equação 13 da mesma figura.

As soluções *fuzzy* ideais positiva e negativa são calculadas por meio das equações 14 e 15, respectivamente. A distância de cada alternativa com relação a essas soluções é calculada por meio das Equações 16 e 17, respectivamente. O cálculo da proximidade de cada alternativa é dado pela Equação 18 e permite o ranqueamento.

2.4 A decisão na avaliação de proficiência em Linguística Aplicada: a possibilidade de tratamento interdisciplinar das discrepâncias e da incerteza

A consistência e a confiabilidade da tomada de decisão no processo de avaliação da proficiência linguística são assuntos que interessam à Linguística Aplicada desde a década de 1990 (BACHMAN; LYNCH; MASON, 1995; MCNAMARA, 1996). Ambos os processos são influenciados por questões que vão do desenho do teste e da definição dos parâmetros avaliativos, como o tipo de tarefa e as características de linguagem que devem ser avaliadas, até às características e atitudes do avaliador no momento decisório, como sua experiência com o ensino de língua estrangeira e com avaliações, treinamento prévio, familiaridade com as escalas avaliativas, capacidade de distinguir níveis de proficiência e conhecimento do processo (KUIKEN; VEDDER, 2014).

Eckes (2008, 2012), no que ficou conhecida como a hipótese do tipo de avaliador, argumenta que o modo de utilização dos critérios, especialmente com relação à importância desses critérios, acaba sendo um comportamento bastante fixo por parte dos avaliadores. Cai (2015) acaba definindo três perfis principais de avaliadores de desempenho oral: o orientado pela forma (*form-oriented*), o equilibrado (*balanced*) e o orientado pelo conteúdo (*content-oriented*).

Diversos estudos apontam diferenças no comportamento dos avaliadores, as quais podem, eventualmente, gerar discrepâncias nas avaliações realizadas. Tais divergências dão-se, especialmente, com relação à severidade avaliativa (MYFORD; WOLFE, 2000; STASSENKO; SKOPINSKAJA; LIIV, 2010; BORGER, 2014; HUHTA *et al.*, 2014; YAN, 2014; YOUN, 2018), à variabilidade de concordância entre avaliadores com relação a níveis específicos de proficiência (STASSENKO; SKOPINSKAJA; LIIV, 2010; GRANFELDT; AGREN, 2014; YAN, 2014), à importância atribuída aos critérios avaliativos (STASSENKO; SKOPINSKAJA; LIIV, 2010; TAGUCHI, 2011; KUIKEN; VEDDER, 2014) e à consistência quanto ao uso da escala avaliativa designada (MYFORD; WOLFE, 2000) durante o processo decisório.

Existem diferentes estratégias que visam a minimizar os impactos da discrepância nas avaliações de proficiência de línguas, como a delimitação de um número menor de faixas de proficiência – alternativas – para que elas tenham características distintivas mais marcadas entre si e menos subjetivas; o processo de correção por pares, em que duas pessoas avaliam a mesma amostra e, no caso de discrepâncias, convida-se um terceiro avaliador; o treinamento dos avaliadores, dentre outros. No entanto, apesar da eficácia e da importância desses métodos, é válido observar que eles ainda geram alguns problemas insatisfatórios para o processo

decisório. No caso da escala, uma redução muito drástica da quantidade de níveis de proficiência pode gerar, por exemplo, uma generalização exacerbada da linguagem avaliada, deixando-se de compreender e de considerar características importantes ou, ainda, fazendo-se com que a tentativa de representatividade de se adequar à realidade das amostras encontradas seja prejudicada. Na avaliação por pares, caso não haja concordância, não há garantias de que o terceiro, ou o quarto corretor irão concordar com uma das avaliações iniciais, o que pode aumentar ainda mais a discrepância e o dispêndio de recursos. No caso dos treinamentos, muitas vezes, o processo de avaliação não é esclarecido, seu construto e suas dimensões acabam sendo colocadas de maneira implícita e dependente da experiência dos avaliadores, o que também pode aumentar tanto a subjetividade quanto as discordâncias.

Com relação à severidade para atribuição de notas e conceitos, o estudo de Myford e Wolfe (2000), utilizando-se do modelo multifacetado de Rasch¹³, mais especificamente do modelo politômico de crédito parcial, analisou, dentre outros aspectos, as medidas de severidade dos avaliadores do antigo *Test of Spoken English* (TSE). Os resultados do estudo revelaram que as médias de proficiência dos candidatos eram cinco vezes maiores do que a extensão da severidade avaliativa, indicando, portanto, que o impacto das diferenças individuais de severidade tinha menor importância geral, o que tornaria o processo avaliativo menos tendencioso. As discrepâncias encontradas no estudo foram originadas por avaliações dicotômicas de avaliadores mais severos e de avaliadores mais tolerantes, erros aleatórios de avaliação, diferenças significativas de desempenho por parte dos candidatos. No caso do primeiro tipo de discrepância, os autores sugerem que ajustes à medida de severidade dos avaliadores sobre a pontuação bruta dos candidatos sejam realizados de modo a poupar a necessidade da presença de um terceiro avaliador no processo.

Stassenko, Skopinskaja e Liiv (2010), por sua vez, em um estudo sobre o desempenho oral de alunos no exame nacional de língua inglesa da Estônia, investigaram a presença de variabilidades culturais que interferem no julgamento avaliativo. No que se refere à severidade, mais especificamente, os autores encontraram uma correlação significativa entre experiência com ensino de línguas e severidade. De maneira geral, os avaliadores menos experientes –

¹³ O modelo de Rasch consiste em um modelo cujos resultados exibem o que seria esperado em respostas a itens mensurados em detrimento da capacidade do candidato e do nível de dificuldade da resposta. O modelo de Rasch básico é empregado para itens de respostas dicotômicas, ou seja, quando há apenas duas categorias de respostas envolvidas (certo ou errado, por exemplo). O modelo multifacetado de Rasch é empregado para itens de respostas politômicas, ou seja, quando existem variáveis (facetadas) que podem contribuir para a ocorrência de erros nas mensurações (TOFFOLI; SIMON, 2018).

aqueles com menos de cinco anos de experiência docente – tendiam a ser mais severos com relação à avaliação de fluência e de pronúncia dos candidatos do que os demais.

Borger (2014), por sua vez, destaca a ausência de consideração do efeito da severidade no processo avaliativo, de modo que apenas questões de concordância são levadas em consideração. Para o cálculo dos índices de severidade de avaliadores suecos e avaliadores externos, acostumados a empregar o CEFRL, da Finlândia e da Espanha no teste sueco nacional de inglês, foram utilizados testes de correlação. O teste de correlação de postos de Spearman e a correlação por postos de Kendall foram empregados no cálculo da confiabilidade entre avaliadores, cujos resultados se mostraram razoáveis. O coeficiente alfa de Cronbach mediu a consistência interna do grupo de avaliadores, revelando a existência de variabilidade entre os avaliadores e diferenças de consistência. O estudo sugere o emprego de avaliações em duplas e o uso do modelo multifacetado de Rasch para tratar as discrepâncias.

Yan (2014) estudou o desempenho de avaliadores em um teste americano de abrangência local de proficiência em inglês. No que concerne à severidade, o estudo revelou que diferenças influenciaram no resultado final apenas em níveis específicos de proficiência – as discrepâncias entre avaliadores aumentam conforme o nível diminui –, sem impactar, contudo, toda a avaliação. Similarmente, Youn (2018) observou consistência de concordância entre avaliadores, mas níveis de severidade variados. O estudo de Youn (2018), utilizando-se do modelo multifacetado de Rasch, confirma que os avaliadores tendem a ser mais tendenciosos em avaliações de candidatos com menores níveis de proficiência, variando seus níveis de severidade para estes candidatos.

McNamara (2006) ressalta a importância da confiabilidade do procedimento avaliativo. Assim, na avaliação de itens discretos, como gramática, vocabulário, compreensão oral ou escrita, o autor destaca que a aferição da confiabilidade pode ser feita de maneira simples por meio da avaliação da consistência do desempenho em cada item, de maneira discriminada, e na sequência pela agregação desses índices. No entanto, em testes cujo objetivo é a avaliação da capacidade de produção oral ou escrita, pelo fato de a avaliação envolver julgamentos de ordem subjetiva, sugere-se que a estimativa de confiabilidade seja realizada por meio dos índices de concordância entre os avaliadores.

Nesse sentido, Borger (2014) ressalta que, de acordo com a Teoria Clássica do Testes, a confiabilidade dos resultados aumenta conforme mais de um avaliador é empregado no procedimento avaliativo. Desse modo, pode-se realizar tanto uma média das notas fornecidas pelos avaliadores ou empregar um terceiro avaliador na presença de discrepâncias. No entanto, é válido destacar que este método não é de todo não problemático, uma vez que dois avaliadores

podem cometer erros no processo, super ou subestimando o desempenho do candidato em seus julgamentos e, conseqüentemente, distanciando-os de uma aproximação apropriada da representação da real capacidade do candidato. Concordância entre avaliadores não é, portanto, sinônimo de confiabilidade.

O referido cenário deixa evidente a possibilidade de tratamento interdisciplinar entre a Linguística Aplicada, especialmente a avaliação de proficiência, e a ciência de Teoria da Decisão diante do entrelaçamento de conceitos e objetivos existente entre ambas. O entendimento da decisão enquanto uma atividade de posicionamento em relação ao futuro, estabelecido em Gomes, Gomes e Almeida (2009), alinha-se à definição de avaliação de proficiência proposta em McNamara (2000), segundo a qual, a testagem consiste na coleta e análise de informações relevantes por meio das quais inferências sobre as capacidades futuras de uso da língua podem ser realizadas. Desse modo, conforme expresso em Kane (2002), o ato de avaliar é equivalente ao ato de tomar decisões.

Além disso, há semelhança nos problemas enfrentados pelas áreas, como o tratamento da complexidade dos objetos, da incerteza sobre as informações, dos múltiplos objetivos da decisão e dos pontos de vista conflitantes por parte dos decisores. No caso da proficiência do professor de línguas, a complexidade é expressa por meio da dinamicidade interacional dos componentes nos quais o uso da língua – objeto de análise – pode ser fragmentado para fins de compreensão e análise; a incerteza, por sua vez, ocorre em consequência da complexidade do objeto e da subjetividade inerente a informações difusas e de difícil compreensão, como, por exemplo, o estabelecimento de padrões de classificação envolvendo a pronúncia dos candidatos; os múltiplos objetivos aparecem na necessidade de tratamento de duas esferas comunicativas, a de uso de língua para finalidades gerais e a de uso para finalidades específicas; e o conflito, na presença de inúmeros usuários (*stakeholders*) de avaliações com potencial alto impacto.

Tanto na Linguística Aplicada quanto na Teoria da Decisão, a procura por maneiras mais eficazes de estruturação de problemas decisórios é resultado da busca por decisões mais confiáveis, que gerem menos impactos sociais negativos e que considerem a opinião de um maior número de usuários no tratamento do problema. Nesse sentido, o cenário de avaliação de línguas atual mostra-se propício ao emprego de métodos de estruturação de problemas e de métodos multicritério de tomada de decisão, apesar do uso ainda escasso. À medida em que tais métodos permitem a análise de critérios em relação às alternativas, nesse caso, às faixas de proficiência, tem-se que eles garantem uma orientação do processo avaliativo similar à usualmente empregada por examinadores. Diferentemente do tradicional, contudo, tais métodos

possibilitam a consideração das opiniões dos usuários (*stakeholders*) de maneira agregada, ou seja, tratam os diferentes níveis de leniência e severidade inerentes ao processo, minimizando a ocorrência de discrepâncias e enviesamentos. O emprego dos métodos pode garantir, assim, a praticidade das avaliações, pois ele favorece a poupança de recursos diversos ao eliminar a necessidade de disponibilidade de um amplo grupo de examinadores, melhorando o tempo de correção e fornecimento de resultados aos candidatos por não incorrer em discrepâncias. Do mesmo modo, o emprego dos métodos pode garantir a validade dos resultados, uma vez que propicia uma maior compreensão do problema de análise. Além disso, os resultados fornecidos por tais métodos não se deixam influenciar pela presença de valores extremos, como ocorre no emprego de medidas de tendência central, como a média, usualmente empregada em cálculos avaliativos.

Para avaliação de proficiência, relata-se o uso de um método multicritério de análise da decisão no estudo de Taskin, Üstün e Deliktas (2013). Uma abordagem de lógica *fuzzy* – ou seja, que lida com imprecisão e vagueza – foi aplicada ao AHP – *Analytic Hierarchy Process* para avaliar os candidatos a intercâmbio pelo programa Erasmus no exame oral. De acordo com os autores, o exame avalia os candidatos com base em critérios subjetivos, como capacidade de compreensão oral, fluência, vocabulário, pronúncia, gramática e auto-expressão, determinados pelos coordenadores e pelo comitê do programa por meio de brainstorming. O estudo concluiu que o ranqueamento dos candidatos obtido por meio do AHP foi mais satisfatório na opinião dos decisores do que a rubrica de avaliação empregada até então, pois o método possibilitava mudança nos pesos aplicados a cada decisor.

2.5 O capítulo em resumo

Nesta seção foram apresentadas as teorias que embasam e orientam a investigação. Inicialmente, foram tecidas considerações acerca dos modelos de avaliação e de desenvolvimento e operacionalização de avaliações de proficiência linguística. Na sequência, foram apresentados os conceitos e elementos integrantes da proficiência do professor de línguas para ensinar em relação aos conceitos de validade e confiabilidade, bem como seus princípios avaliativos. O capítulo foi finalizado com a introdução do conceito de análise da decisão, por meio da apresentação dos métodos de estruturação de problemas e de análise da decisão multicritério, e com considerações acerca da possibilidade de tratamento interdisciplinar da área de estudos da decisão e de avaliação linguística. O capítulo a seguir trata da metodologia empregada na pesquisa.

3 CAPÍTULO II: METODOLOGIA DA PESQUISA

No presente capítulo apresenta-se a metodologia empregada para o desenvolvimento da investigação. O capítulo inicia-se pelo relato de informações referentes à natureza e ao contexto da pesquisa e é finalizado com informações sobre os instrumentos e procedimentos de geração, coleta, análise e triangulação de dados empregados.

3.1 Natureza da pesquisa

Este trabalho segue uma abordagem híbrida (ELLIS, 1990) de natureza aplicada (RIAZI, 2016), pois utiliza-se de procedimentos qualitativos e quantitativos para orientar o tratamento da subjetividade e da dinamicidade interacional dos componentes em uma abordagem avaliativa de um exame de proficiência de professores de línguas em fase de implementação. De acordo com Ellis (1990), a abordagem híbrida define-se pelo emprego complementar de técnicas qualitativas e quantitativas de análise de dados, o que, segundo Riazzi (2016), beneficia o desenho do problema de análise por minimizar a presença de sobreposição de lacunas de ambas as abordagens ao combinar suas forças. Nesta pesquisa, o aspecto qualitativo garante-se por meio do emprego de técnicas de coleta e análise de dados envolvendo questionários, consulta e descrição de dados secundários, e descrição do objeto; e o aspecto quantitativo, por sua vez, faz-se presente em virtude do emprego de métodos multicritério para a análise da decisão.

Riazzi (2016) determina que pesquisas de natureza aplicada se diferenciam das pesquisas básicas por preocuparem-se com a busca por soluções para problemas práticos imediatos em contextos determinados. Desse modo, por incorporarem as circunstâncias nas quais um problema se insere, os resultados deste tipo de pesquisa não têm força suficiente para fornecer generalizações a outros contextos. Neste trabalho, a natureza aplicada justifica-se pelo tratamento da avaliação do uso da língua para finalidades gerais e pedagógicas em um exame de proficiência para professores de línguas.

O tipo de pesquisa empregado consiste no exploratório-descritivo, cuja combinação, de acordo com Riazzi (2016), intensifica a abrangência da compreensão do fenômeno investigado. Na definição do autor, a pesquisa exploratória consiste naquela que se ocupa de objetos que ainda não foram exaustivamente investigados, ou seja, cuja compreensão ainda experimenta limitações. Neste estudo, o caráter exploratório explica-se pelo surgimento recente do interesse

por estudos acerca da caracterização e da avaliação da proficiência do professor de línguas e, por extensão, da carência de estudos sobre o tema. A pesquisa descritiva, por sua vez, ocupa-se da descrição do relacionamento das variáveis de um determinado fenômeno social de uma população, fornecendo, como resultados, uma compreensão detalhada e abrangente sobre o objeto. Neste trabalho, o caráter descritivo explica-se pelo propósito de oportunizar maior entendimento sobre a avaliação e sobre o conceito de proficiência do professor de línguas.

O trabalho consiste, por fim, em um estudo de caso (CASANAVE, 2010; RIAZI, 2016), pois visa investigar o processo avaliativo de um objeto bem delimitado, o exame EPPLE. Casanave (2010) e Riazi (2016) determinam como estudos de caso pesquisas que se ocupam de um único objeto e cuja investigação é realizada de forma detalhada e holística e orientada por perguntas de pesquisa.

3.2 Contexto e participantes da pesquisa

Este estudo presta-se à investigação da abordagem avaliativa do exame EPPLE e visa propor uma abordagem que agregue a opinião dos usuários (*stakeholders*) do exame e que permita uma consideração mais prática do julgamento dos avaliadores. Para tanto, realiza uma análise descritiva das tarefas e dos dados de fala que compõem o banco de dados de aplicações do teste oral do EPPLE nos anos de 2015 e 2017 a formandos em Letras de uma universidade pública paulista e duas mineiras, bem como dos conceitos atribuídos a esses candidatos. Na sequência, procede-se à aplicação de questionários aos usuários (*stakeholders*) do exame, de modo a delimitar suas opiniões sobre os elementos que integram a proficiência do professor de línguas e sua avaliação; à elaboração de uma escala de mensuração e de uma escala de proficiência ao exame; e à reavaliação dos dados de fala com base nas novas escalas e na nova abordagem. Consistem em participantes do estudo, portanto, os candidatos ao exame nas aplicações de 2015 e 2017 e os usuários (*stakeholders*) especialistas, envolvidos direta ou indiretamente com o processo avaliativo.

3.2.1 O teste oral do exame EPPLE

O EPPLE consiste em um exame de proficiência que avalia a proficiência do professor de língua estrangeira em suas modalidades oral e escrita (CONSOLO, 2008; CONSOLO, 2011a, 2011b). O exame surgiu no ano de 2008 no formato de entrevista em pares face-a-face, da qual participavam dois examinadores – um responsável pela interação e outro examinador-observador. Em 2011, evoluiu para o formato eletrônico. Neste formato, a interação dá-se

exclusivamente com o computador por meio de perguntas e respostas e o processo de avaliação passa a ser realizado *a posteriori*, na qual dois avaliadores analisam a fala gravada dos candidatos – na presença de discrepância sobre o resultado final, um terceiro avaliador é solicitado e, assim, sucessivamente até a obtenção de uma faixa em consenso para cada candidato.

O exame trabalha com duas escalas de proficiência: uma exclusiva para o teste oral e outra exclusiva do teste escrito. Neste trabalho, analisa-se apenas a escala do teste oral apresentada em Consolo e Teixeira da Silva (2014) – Anexo III. A referida escala foi proposta, com base nas reflexões de Consolo (2004), Consolo e Teixeira da Silva (2007b) e Baffi-Bonvino (2010), e tem sido utilizada desde 2011. Ela consiste em uma escala holística com descritores analíticos, dividida em cinco faixas de proficiência (A – mais alta, B, C, D e E – mais baixa), cada uma contendo seis descritores do tipo *can-do*. Ressalta-se que, atualmente, não há no EPPLÉ uma faixa de proficiência mínima que o candidato deva atingir para estar ou não apto a ministrar aulas, ou seja, as faixas do exame têm caráter descritivo, cabendo a seus usuários (*stakeholders*) tomarem decisões com base nos resultados fornecidos.

As versões empregadas nos anos de 2015 e 2017 apresentam algumas diferenças no que se refere à tarefa de uso de língua para finalidade pedagógica/metalinguística. Em 2015 e na versão preencial de 2017, o teste oral continha, no total, dez questões divididas em três partes. As cinco primeiras questões solicitavam que o candidato fornecesse informações de cunho pessoal, acadêmico e profissional, incluindo experiências e planos relacionados ao ensino e à aprendizagem de língua inglesa. As quatro questões sequenciais solicitavam que o candidato assistisse a dois vídeos sobre pronúncia de inglês como língua estrangeira. Os candidatos deveriam discorrer sobre o assunto abordado, o público ao qual o vídeo era destinado e relacionar o assunto a aulas de língua estrangeira. A última questão solicitava que o candidato escolhesse uma de duas situações que representavam dúvidas comumente trazidas por alunos de língua estrangeira – uma relacionada ao emprego de presente simples em oposição ao emprego de presente perfeito e outra relativa aos usos de *will* e *going to* para expressão de futuro – e discorresse de modo a solucionar a dúvida do aluno.

Já a versão eletrônica aplicada em 2017 era composta por doze questões e dividida em quatro partes. Na primeira, informações de cunho pessoal, profissional e acadêmico eram solicitadas, assim como na versão de 2015. A segunda parte, requeria que o candidato descrevesse e comparasse duas situações de sala de aula apresentadas por meio de figuras. Na terceira, ele deveria assistir a apenas um vídeo sobre pronúncia de inglês como língua estrangeira e discorrer sobre o assunto abordado, o provável público ao qual o vídeo era

destinado e relacionar o assunto a aulas de língua estrangeira. A respeito desse mesmo vídeo, o candidato deveria, ainda, responder a quatro questões de múltipla-escolha que tinham por finalidade avaliar sua capacidade de compreensão oral. A questão sobre emprego de linguagem específica era similar à de 2015, havendo alteração apenas no conteúdo solicitado – emprego de modais para tratar de possibilidade.

Tanto no ano de 2015 quanto em 2017, os conceitos (A, B, C, D ou E) eram atribuídos com base na escala de proficiência oral por meio do modelo de avaliação por pares, com a presença de um terceiro examinador, o revisor, no caso da ocorrência de discrepâncias. No EPPLÉ, para ser classificado em uma faixa de proficiência é necessário que o desempenho do candidato represente todos os descritores daquele nível, caso contrário, ele deve ser classificado na faixa imediatamente inferior. Tal abordagem de classificação pode prejudicar o candidato, uma vez que ele fica sujeito a receber informações inferiores a seu desempenho real. A impossibilidade de uma escala de proficiência descrever fielmente os candidatos é relatada, inclusive, no volume de apresentação de novos descritores do CEFR (COE, 2018). No entanto, o texto defende o caráter estimativo das faixas, o que nos leva a questionar o caráter penalizante da abordagem de avaliação atualmente empregada pelo EPPLÉ.

A linguagem pedagógica configura-se na característica mais marcante que difere o exame EPPLÉ de outros exames de proficiência, inclusive do TKT. Isso porque, no EPPLÉ, a competência de uso desta linguagem é abordada por meio de princípios comunicativos, ou seja, o exame é organizado de modo a não se preocupar apenas com a estrutura linguística, mas também com as funções comunicativas enunciadas por meio da fala.

É importante ressaltar que o emprego de dois formatos de interação distintos – eletrônico e presencial – como objetos desta investigação não inviabiliza sua condução, uma vez que, em ambos os formatos, o objetivo da avaliação proposta pelo exame é o mesmo. Deste modo, este estudo pode contribuir para a geração de uma abordagem padronizada de inferências sobre as capacidades de uso da língua, tratando as especificidades dos modos como as evidências empíricas foram solicitadas.

3.2.2 Os candidatos ao exame EPPLÉ

Os candidatos, cujos dados de fala consistem em objetos de investigação deste estudo, correspondem a formandos em Letras de uma universidade pública paulista e de duas universidades públicas mineiras, responsáveis pela oferta de cursos de excelência segundo os resultados do Enade de 2017 – respectivamente, notas 5 para a universidade paulista e 4 para as universidades mineiras. Os dados de fala do ano de 2015 compreendem aplicações do exame

realizadas no formato eletrônico e os de 2017, a aplicações realizadas em formato de entrevista em duplas face-a-face.

No ano de 2015, a aplicação contou com a presença de 9 participantes da universidade paulista e 31 participantes de uma universidade pública mineira, dos quais apenas 27 dados de gravação de fala tinham qualidade suficiente para análise. Em 2017, participaram 13 candidatos por parte da universidade mineira e 11 da universidade paulista, sendo que, deste último grupo, apenas 9 dados garantiam qualidade de análise. No total, a amostra empregada neste estudo contou com 64 participantes, dos quais apenas 58 compõem o corpus de análise deste estudo em virtude de problemas de gravação.

O Quadro 8 condensa informações a respeito do ano de aplicação dos dados selecionados para análise, das nomenclaturas adotadas para referência aos candidatos, aos estados a que pertencem, às universidades que os participantes integram, ao formato da interação, à duração das gravações e à faixa de proficiência a eles atribuída.

Quadro 8. Quadro-resumo dos dados de fala.

Ano	Candidato	Estado	Interação	Duração	Faixa
2015	HEL	SP	Eletrônica	4min 48s	C
2015	KAR	SP	Eletrônica	5min 58s	C
2015	BRU	SP	Eletrônica	4min 56s	C
2015	DEM	SP	Eletrônica	6min 63s	C
2015	DES	SP	Eletrônica	5min 55s	B
2015	LUC	SP	Eletrônica	3min 86s	B
2015	MAC	SP	Eletrônica	4min 96s	D
2015	SUE	SP	Eletrônica	5min 00s	B
2015	MAR	SP	Eletrônica	5min 01s	B
2015	RIT	MG	Eletrônica	08min 14s	D
2015	ALE	MG	Eletrônica	n/a	C
2015	BAS	MG	Eletrônica	n/a	C
2015	CAH	MG	Eletrônica	06min 42s	B
2015	CLC	MG	Eletrônica	07min 33s	B
2015	CLO	MG	Eletrônica	05min 39s	B
2015	DAL	MG	Eletrônica	07min 19s	B
2015	ANC	MG	Eletrônica	n/a	C
2015	BAA	MG	Eletrônica	07min 34s	B
2015	FRA	MG	Eletrônica	n/a	n/a
2015	LAC	MG	Eletrônica	08min 14s	A
2015	CAS	MG	Eletrônica	07min 57s	B
2015	CLA	MG	Eletrônica	07min 05s	A
2015	LUI	MG	Eletrônica	06min 33s	B
2015	MAA	MG	Eletrônica	02min 18s	C
2015	GUS	MG	Eletrônica	07min 42s	B
2015	MIR	MG	Eletrônica	04min 04s	C
2015	PAT	MG	Eletrônica	05min 07s	B
2015	JEL	MG	Eletrônica	07min 14s	D
2015	JEQ	MG	Eletrônica	06min 00s	B
2015	JOA	MG	Eletrônica	08min 48s	B
2015	ROB	MG	Eletrônica	07min 34s	E

2015	LID	MG	Eletrônica	05min 58s	B
2015	LIG	MG	Eletrônica	07min 27s	C
2015	SAB	MG	Eletrônica	05min 53s	C
2015	TAS	MG	Eletrônica	06min 26s	C
2015	TAC	MG	Eletrônica	06min 03s	D
2015	PAB	MG	Eletrônica	05min 08s	C
2015	WAN	MG	Eletrônica	06min 46s	D
2015	RIC	MG	Eletrônica	05min 15s	E
2015	THI	MG	Eletrônica	05min 10s	B
2015	THY	MG	Eletrônica	06min 00s	D
2017	REN	MG	Par 1	25min 27s	B
2017	SUL	MG	Par 1	25min 27s	B
2017	LOR	MG	Par 2	20min 17s	B
2017	ALC	MG	Par 2	20min 17s	C
2017	LAI	MG	Par 3	20min 17s	C
2017	MAT	MG	Par 3	20min 17s	B
2017	LCA	MG	Par 4	37min 21s	B
2017	JOS	MG	Par 4	37min 21s	C
2017	MAE	MG	Par 5	27min 58s	C
2017	ALN	MG	Par 5	27min 58s	B
2017	GAB	MG	Par 6	20min 53s	B
2017	HIL	MG	Par 6	20min 53s	B
2017	ISA	MG	Individual	14min 25s	B
2017	REG	SP	Par 7	20min 26s	C
2017	PAM	SP	Par 7	20min 26s	B
2017	MAI	SP	Par 8	18min 13s	C
2017	SAR	SP	Par 8	18min 13s	B
2017	LUS	SP	Par 9	24min 38s	D
2017	JUL	SP	Par 9	24min 38s	C
2017	BAR	SP	Par 10	02min 30s	C
2017	FER	SP	Par 10	02min 30s	B
2017	RAI	SP	Par 11	24min 26s	C
2017	NAD	SP	Par 11	24min 26s	B
2017	NAS	SP	Individual	13min 21s	B

Fonte: Elaborado pela autora.

Ressalta-se que as diferenças na quantidade de tempo exibida pelas gravações no Quadro 8 devem-se, primeiramente, ao fato de que, na versão eletrônica, são gravadas apenas as respostas fornecidas pelos candidatos, e não o insumo para interação, como na versão presencial. Ambas as versões têm durações aproximadas de 20 minutos.

3.2.3 Os usuários (*stakeholders*) especialistas do exame

Os participantes desta pesquisa foram selecionados com base nas orientações de Elder *et al.* (2017) e McNamara (2006). Desse modo, consistem em (i.) especialistas envolvidos com a elaboração e operacionalização do exame EPPL e (ii.) especialistas genéricos, ou seja, usuários (*stakeholders*) do exame não envolvidos com as fases de elaboração e operacionalização, como professores, diretores de escolas, selecionadores de professores,

treinadores de professores, aplicadores de exames de proficiência, alunos, proprietários de escolas, dentre outros.

O grupo de especialistas diretamente envolvido com o exame, é composto por seis representantes, na primeira etapa da pesquisa, e três, na segunda, das regiões Norte, Centro-Oeste e Sudeste, cujo tempo de envolvimento varia de dois a mais de 16 anos. O grupo de especialistas envolvidos indiretamente é composto por 58 representantes, na primeira etapa, e 17, na segunda, de todas as regiões brasileiras, os quais atuam em instituições de ensino nos âmbitos federal, estadual e particular dos ensinos fundamental e médio; de cursos pré-vestibulares, preparatórios e de idiomas; além de cursos de graduação e de pós-graduação.

3.3 Procedimentos e instrumentos de coleta e geração de dados

Os procedimentos e instrumentos de coleta e geração de dados empregados nesta pesquisa consistem em transcrições de gravações de áudio do desempenho de fala de candidatos ao teste oral do exame EPPLE e em questionários aplicados aos usuários (*stakeholders*) do exame.

3.3.1 Transcrições das gravações em áudio do desempenho de fala de candidatos ao teste oral do exame EPPLE

Os dados de fala que integram este estudo foram extraídos do corpus do exame EPPLE e transcritos utilizando-se do *software* CLAN – *Computerized Language Analysis*¹⁴, em virtude de sua disponibilização gratuita e da automaticidade de cálculo de alguns índices de capacidades linguísticas propiciado pelas funções incorporadas ao programa. CLAN foi desenvolvido por Leonid Spektor da Universidade Carnegie Mellon da Pensilvânia, nos Estados Unidos, e é empregado em estudos de interação conversacional, aprendizado de línguas e de distúrbios de fala. O *software* emprega padrões de transcrições no formato CHAT, que compreende uma série de codificações de análise conversacional desenvolvidas por Sacks, Schegloff e Jefferson (1974) e internacionalmente empregadas em estudos de conversação (MCWHITNEY, 2019). O Quadro 9 condensa os códigos empregados nas transcrições realizadas neste estudo.

¹⁴ Disponível em: <http://dali.talkbank.org/clan/>. Acesso em 10 Mar. 2019.

Quadro 9. Códigos empregados nas transcrições.

CÓDIGO	CARACTERÍSTICAS DE FALA
*ABC: bla	Fala do interagente
bla .	Declaração completa
bla +...	Declaração incompleta (sem interrupção)
bla +/.	Declaração interrompida
bla !	Exclamação completa
bla ?	Pergunta completa
bla +..?	Pergunta incompleta (sem interrupção)
bla +/?	Pergunta interrompida
+, bla	Auto-preenchimento (após interrupção)
++ bla	Preenchimento do enunciado de outro falante
"bla"	Citação de pequenos trechos
bla +"/.	Citação longa iniciada na próxima trilha (<i>tier</i>)
+ " bla	Início de citação longa
goed [*] [: went]	Erro estrutural seguido de correção pela transcrição
piece [:: peach] [*]	Erro de escolha lexical seguido de correção pela transcrição
<bla> [/] bla	Repetição de palavras ou trechos de fala
<bla> [/] blabla	Reconstituição de palavras ou trechos de fala
<bla> [-] blabla	Falso início sem reconstituição
<bla> [//] cha	Reformulação completa da mensagem
bla [*] [/] bla	Erro no trecho de fala a ser reparado
bla [/] [*] bla	Erro no trecho de fala de reparo
0bla	Palavra omitida
bl(a)	Encurtamento fonológico
bla:bla	Alongamento de som silábico
blabla:	Alongamento de palavra
&bl	Fragmentos fonológicos
&-um	Pausas preenchidas (preenchedores não lexicais)
bla^bla	Pausa dentro de uma palavra
(.)	Pausa silenciosa (sem duração)
(0.6)	Pausa silenciosa (com duração em segundos)
xxx	Discurso incompreensível
bla [=! tosse]	Vocalizações não verbais (ocorrência paralinguística) com ocorrência linguística na trilha (<i>tier</i>)
0 [=! tosse]	Vocalizações não verbais (ocorrência paralinguística) sem ocorrência linguística na trilha (<i>tier</i>)
bla ↑	Entonação ascendente
bla ↓	Entonação descendente
bla &*ABC: bla bla	Sobreposição de discurso (comentários curtos)
<bla> [>]	Sobreposição de discurso (falas simultâneas)
<bla> [<]	
bla@s:por	Palavra em língua portuguesa
bla@q	Uso de metalinguagem
%pho: bla	Desvio de pronúncia
bla'bla	Sílabas enfatizadas
bla <bla bla> [!]	Trecho enfatizado
%com: bla	Comentários
bla ‡ bla	Agrupamento de marcadores conversacionais (posição inicial): vocativos
bla ,, bla	Agrupamento de marcadores conversacionais (posição final): vocativos, perguntas curtas
bla [= bla]	Identidade dêitica
bla +.	Pausa na transcrição
bla [^c]	Delimita orações dentro de sentenças complexas

Fonte: Elaborado pela autora com base em McWhitney (2019).

3.3.2 Questionários para usuários (*stakeholders*) especialistas do exame EPPLE

Os questionários foram elaborados seguindo-se as orientações da etapa qualitativa do método de estruturação de problemas *Value-Focused Thinking* proposto por Keeney (1992) e encontram-se disponibilizados nos apêndices II e III. Os dados foram coletados, inicialmente, por meio de um questionário de respostas abertas e, em sua segunda etapa, por meio de um questionário que mesclava respostas de cunho fechado e aberto (BROWN, 2009), em uma adaptação ao formato de entrevistas sugeridas pelo VFT (KENNEY, 1992, 1996) devido à diversidade de perfis e quantidade de participantes da pesquisa. O contato com os usuários (*stakeholders*) foi realizado por meio do envio de mensagens eletrônicas a programas de pós-graduação reconhecidos com foco em Linguística Aplicada e a associações de Linguística Aplicada. Além disso, divulgações por meio de redes sociais foram realizadas por colegas de profissão visando-se ao alcance de um público amplo e diversificado de participantes.

Para cada grupo, foi aplicado um questionário composto por duas etapas. Na primeira, foram realizadas questões com o objetivo de caracterizar o público participante da pesquisa, solicitando informações como: experiência profissional, experiência acadêmica, contato/experiência com exames de proficiência para professores, dentre outros. Na segunda, foram realizadas dez questões que visavam elicitar as opiniões desses participantes com relação à proficiência do professor de línguas e a exames que visassem à avaliação dessa proficiência, como o EPPLE. O formato das perguntas da segunda etapa seguiu os direcionamentos sugeridos por Keeney (1996) e os critérios abordados resultaram da compilação das respostas do primeiro questionário e a padrões tradicionais de mensuração de proficiência oral explicitados pela literatura, como os apontados em Brown, Iwashita e McNamara (2005) e em Jamieson e Poonpon (2013).

A escolha pelo VFT deu-se em virtude da necessidade apontada por Widdowson (2001), de que a competência comunicativa deve ser avaliada por meio de aspectos que tenham relevância para sua ocorrência, podendo esses aspectos serem dependentes ou derivados uns dos outros. Assim, o VFT é empregado, neste trabalho, em uma tentativa de estruturar o construto comunicacional da situação pedagógica, compreendendo-o mais adequadamente e, a partir daí, propondo modos de mensurá-lo.

3.4 Procedimentos e instrumentos de análise e triangulação de dados

A análise de dados, neste estudo, iniciou-se pela leitura das respostas fornecidas aos questionários, buscando-se categorizar padrões de preferências por critérios para avaliação da proficiência do professor de línguas na opinião dos usuários (*stakeholders*) do exame EPPLE.

A identificação dos padrões orientou-se pelo critério semântico em relação aos componentes da proficiência linguística para finalidades gerais e para finalidades específicas de ensino delimitados pelo CEFR (COE, 2018) e pelo CEFT (CAE, 2018), tendo em vista o modelo de competência linguística proposto por Bachman (1990) – uma vez que o modelo foi proposto especificamente para a avaliação de línguas – e o de competência linguístico-comunicativa proposto por Almeida Filho (1992). A contagem das preferências dos critérios foi realizada por meio do cálculo de frequências de apontamentos.

Na sequência, a investigação voltou-se à análise do teste oral do EPPLE, mais especificamente, de sua escala de proficiência e de suas tarefas, em uma tentativa de clarificação de seu construto. Seguindo-se as orientações do VFT (KEENEY, 1992, 1994, 1996) foi realizada uma hierarquização dos objetivos da avaliação da proficiência na escala empregada pelo exame e, paralelamente, conduziu-se uma análise dos tipos de tarefas presentes no teste, extraindo-se informações a respeito das funções comunicativas por elas obrigatória e potencialmente mobilizadas. A análise da relação existente entre a escala e os critérios efetivamente elicitados pelas tarefas apontaram para a existência de discrepâncias que justificam a necessidade de reformulação da escala empregada pelo exame.

Seguiu-se, então, para a análise do desempenho dos dados de fala dos candidatos, primeiramente, por meio do levantamento das características de fala e, posteriormente, por meio da comparação desses desempenhos aos conceitos atribuídos a eles pelos examinadores. A análise das características de fala foi realizada congregando-se as funções comunicativas elicitadas pelas tarefas, os critérios levantados pelos usuários (*stakeholders*) do exame nas respostas aos questionários e as unidades de medida tradicionalmente empregadas na literatura sobre avaliação de proficiência. A análise da qualidade da fala produzida pelos candidatos no exame EPPLE foi realizada por meio de contagens manuais e de uma combinação de programas de análises automáticas, como funcionalidades do *software* CLAN, o programa de análise de fala PRAAT¹⁵ e o programa de análise de frequência vocabular RANGE¹⁶.

O resultado da análise para cada critério foi submetido a uma análise multivariada de conglomerados, para a qual foi empregado o método de agregação Ward com distância euclidiana e estandardização de variáveis, a fim de se definir hierarquicamente os grupos de desempenho elicitados pelos dados. A análise do dendrograma gerado baseou-se na busca por quatro conglomerados ($k = 4$) – aceitando-se três quando os dados não apontavam para a possibilidade de definição de quatro grupos – em virtude de ser essa a quantidade de faixas de

¹⁵ Disponível em: <http://www.fon.hum.uva.nl/praat/>. Acesso em 10 Mar. 2019.

¹⁶ Disponível em: <https://www.victoria.ac.nz/lals/resources/range>. Acesso em 10 Mar. 2019.

variação de desempenho em proficiência definida pelo CETF (CAE, 2018). Considerando-se esses mesmos critérios, procedeu-se ao cálculo da mediana das características de fala dos candidatos em relação às faixas originalmente atribuídas pelos examinadores com a finalidade de se investigar a medida com que o referido critério se encontrava representado no resultado da avaliação.

A análise dos dados obtidos e das informações da literatura a respeito dos quadros de referência, especialmente o CEFR e o CETF, possibilitou a elaboração de uma escala analítica de mensuração para orientar a avaliação dos desempenhos de fala no exame. Os critérios considerados na escala foram definidos com base na relação existente entre o potencial de eles informarem aspectos holísticos da proficiência e o potencial distintivo desses critérios encontrados na análise de desempenho dos candidatos. As variações de desempenho da escala foram definidas com base na escala de números difusos e variáveis linguísticas propostas por Chen (2000). Paralelamente, procedeu-se à definição de uma escala de proficiência ao exame que comportasse descritores possíveis de serem inferidos por meio dos critérios delimitados na escala de mensuração. Seguindo-se a orientação de elaboração de descritores comunicativos do COE (2018), a escala foi desenvolvida focando-se nas possíveis inferências a respeito das capacidades de uso da língua demonstradas pelo desempenho dos candidatos, e não em suas deficiências, tendo, portanto, uma orientação positiva. As faixas da escala foram apresentadas em uma disposição vertical e decrescente, respeitando-se a disposição originalmente empregada pela escala de proficiência do EPPL.

A etapa de reavaliação dos dados de fala do teste oral do exame utilizando-se do método *Fuzzy* TOPSIS (CHEN, 2000) consiste no momento exato de triangulação dos dados da pesquisa, uma vez que integra a escala de mensuração elaborada com base na análise do desempenho dos candidatos e dos valores dos usuários (*stakeholders*) do exame ao emprego da abordagem de análise proposta, cujos resultados fomentam o refinamento da escala de proficiência sugerida. O método foi aplicado por meio da planilha para o Microsoft Excel® desenvolvida por Bahabadi (2018). A inserção do insumo na planilha divide-se em duas etapas. Inicialmente, conforme apresentado na Figura 8, devem ser inseridos o número de critérios, de alternativas e de decisores a serem considerados, juntamente com os pesos atribuídos a cada critério. Na segunda etapa, exibida pela Figura 9, informa-se a avaliação de cada critério em relação às alternativas propostas. A planilha apresenta, então, a classificação sugerida, conforme exposto pela Figura 10.

Figura 8. Definição dos pesos agregados e do número de critérios e de decisores.

number of experts	number of criteria	number of alternatives	guide to determine kind of criteria	
			profit	cost
			kind of criteria	
			1	-1

criteria	kind	weight												
1st			6th			11th			16th			21th		
2nd			7th			12th			17th			22th		
3rd			8th			13th			18th			23th		
4th			9th			14th			19th			24th		
5th			10th			15th			20th			25th		

Data of criteria | Data of 1st to 20th criteria | Data of 21th to 40th criteria | Chen (2000)

Fonte: Bahabadi (2018).

Figura 9. Avaliação dos critérios em relação às alternativas.

1st criteria													2nd criteria												
VP	P	RP	F	RG	G	VG	VP	P	RP	F	RG	G	VG												
1st Alternativa							1st Alternativa																		
2nd Alternativa							2nd Alternativa																		
3rd Alternativa							3rd Alternativa																		
4th Alternativa							4th Alternativa																		
5th Alternativa							5th Alternativa																		
6th Alternativa							6th Alternativa																		
7th Alternativa							7th Alternativa																		
8th Alternativa							8th Alternativa																		
9th Alternativa							9th Alternativa																		
10th Alternativa							10th Alternativa																		
11th Alternativa							11th Alternativa																		
12th Alternativa							12th Alternativa																		
13th Alternativa							13th Alternativa																		
14th Alternativa							14th Alternativa																		
15th Alternativa							15th Alternativa																		
16th Alternativa							16th Alternativa																		
17th Alternativa							17th Alternativa																		
18th Alternativa							18th Alternativa																		
19th Alternativa							19th Alternativa																		
20th Alternativa							20th Alternativa																		
21th Alternativa							21th Alternativa																		
22th Alternativa							22th Alternativa																		
23th Alternativa							23th Alternativa																		
24th Alternativa							24th Alternativa																		
25th Alternativa							25th Alternativa																		
26th Alternativa							26th Alternativa																		
27th Alternativa							27th Alternativa																		
28th Alternativa							28th Alternativa																		
29th Alternativa							29th Alternativa																		
30th Alternativa							30th Alternativa																		
31th Alternativa							31th Alternativa																		
32th Alternativa							32th Alternativa																		
33th Alternativa							33th Alternativa																		
34th Alternativa							34th Alternativa																		
35th Alternativa							35th Alternativa																		
36th Alternativa							36th Alternativa																		
37th Alternativa							37th Alternativa																		

Very poor (VP)	(0,0,1)
Poor (P)	(0,1,3)
Medium poor (MP)	(1,3,5)
Fair (F)	(3,5,7)
Medium good (MG)	(5,7,9)
Good (G)	(7,9,10)
Very good (VG)	(9,10,10)

Data of criteria | Data of 1st to 20th criteria | Data of 21th to 40th criteria | Chen (2000)

Fonte: Bahabadi (2018).

Figura 10. Resultado da classificação sugerida.

AP	AP	AP	AR	AS	AT	AU	AV	AW	AX	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BH	A34
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A35
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A36
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A37
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A38
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A39
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A40

Rank of Alternatives			
A1	0.3243	0.3268	2
A2	0.7151	0.1651	1
A3	0.2031	0.2031	3
A4	0.0520	0.0520	4
A5	0.0520	0.0520	4
A6	0		0
A7	0		0
A8	0		0
A9	0		0
A10	0		0
A11	0		0
A12	0		0
A13	0		0
A14	0		0
A15	0		0
A16	0		0
A17	0		0
A18	0		0

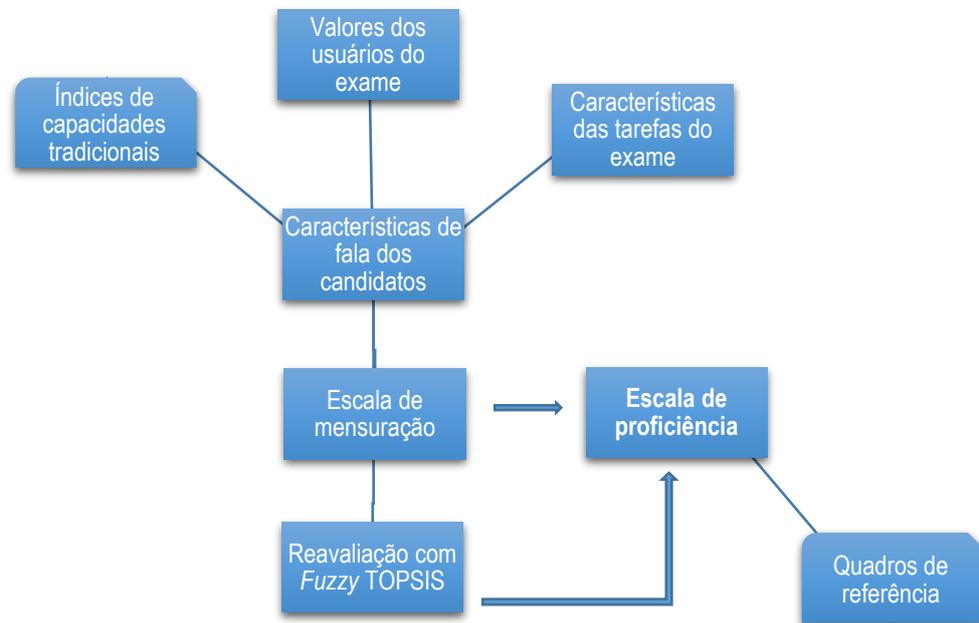
Data of criteria | Data of 1st to 20th criteria | Data of 21th to 40th criteria | Chen (2000)

Fonte: Bahabadi (2018).

Apesar da existência de *softwares* com *layouts* mais elaborados para aplicação do *Fuzzy TOPSIS*, a planilha foi utilizada em virtude de sua gratuidade de acesso. A definição dos pesos dos critérios que compõem a escala foi realizada considerando-se a média aritmética da importância a eles atribuída pelos usuários (*stakeholders*) do exame, pela literatura e pela criterialidade definida pela análise das medianas da atribuição de conceitos, cálculo esse possibilitado pela categorização dos dados pelo critério semântico e pela frequência de menções. Os resultados da reavaliação foram, por fim, analisados em relação à escala de proficiência sugerida, buscando-se verificar a força da representatividade dos resultados da nova abordagem avaliativa.

É importante ressaltar que, neste estudo, não se trabalha com análise estatística paramétrica para a elaboração de parâmetros de resultados de proficiência, uma vez que essas técnicas apresentam limitações com relação à especificação e modelagem da variável dependente, bem como de disponibilidade, estabilidade e outras exigências hipotéticas com relação aos dados, o que restringe a capacidade de predição (ZOPOUNIDIS; PENTARAKI; DOUMPOS, 1998). No entanto, análises matemáticas foram empregadas com a finalidade de validar os resultados obtidos. Para todas as análises empregadas neste estudo, foi utilizado o *software* BioEstat 5.0 (AYRES; AYRES JR.; AYRES; SANTOS, 2007).

A Figura 11 resume, por fim, a interação dos dados gerados e analisados neste estudo.

Figura 11. Triangulação dos dados do estudo.

Fonte: Elaborado pela autora.

3.5 O capítulo em resumo

Neste capítulo foram apresentadas informações acerca dos procedimentos metodológicos empregados na investigação. O estudo proposto utiliza-se de uma abordagem híbrida de caráter exploratório-descritivo para realizar um estudo de caso acerca da abordagem avaliativa do exame EPPL e propor adequações em sua estruturação. Assim, inicialmente, trabalha com base nos princípios do método de estruturação de problemas *Value-Focused Thinking*, que permite a elicitação das opiniões que embasam a compreensão do objeto por parte dos usuários (*stakeholders*) e da escala de proficiência oral do exame. Após a fase de estruturação, analisa-se o desempenho dos candidatos e as características das tarefas e dá-se início à elaboração de uma escala de mensuração e de uma escala de proficiência para o exame que comporte as necessidades apontadas pelos usuários (*stakeholders*). Na sequência, o estudo procede à reavaliação de dados de fala do EPPL com base no modelo de decisão multicritério aplicado por meio do método multicritério de análise da decisão *Fuzzy TOPSIS*. O estudo é finalizado por meio de uma discussão acerca da comparação dos resultados fornecidos pela nova metodologia e a escala de proficiência proposta para o exame. O capítulo a seguir aborda as discussões e análises dos dados conduzidas.

4 CAPÍTULO III: DISCUSSÃO E ANÁLISE DE DADOS

Este capítulo tem por objetivo apresentar a análise e discussão dos dados gerados e coletados em consonância com a metodologia e a fundamentação teórica apresentadas nos capítulos anteriores, buscando respostas para a pergunta e para as perguntas de pesquisa propostas e apresentadas no texto introdutório deste trabalho.

4.1 Em busca de critérios para avaliação da proficiência oral do professor de línguas: os interesses dos usuários (*stakeholders*) do exame

Nesta seção, são apresentadas as análises das respostas fornecidas aos questionários enviados aos usuários (*stakeholders*) do exame nos anos de 2017 e 2018. Inicialmente, apresenta-se a análise das respostas obtidas ao questionário enviado no ano de 2017, que tinha por objetivo realizar um levantamento de critérios relevantes para avaliação do professor de línguas na opinião dos participantes. Na sequência, procede-se à análise dos dados obtidos por meio do questionário enviado em 2018, o qual consistia em um apanhado de técnicas de mensuração de proficiência com base no levantamento propiciado pelo primeiro questionário e por técnicas de avaliação da proficiência empregada em exames consagrados, como os da ETS (BROWN; IWASHITA; MCNAMARA, 2009). A subseção é finalizada com considerações que resumem os dados observados.

4.1.1 As respostas fornecidas ao questionário de levantamento de critérios

A aplicação do primeiro questionário aos usuários (*stakeholders*) não envolvidos diretamente com o exame no primeiro semestre de 2017 resultou em 58 respostas, com participação de público das cinco regiões brasileiras: Norte (Roraima e Tocantins), Nordeste (Bahia, Piauí e Ceará), Centro-Oeste (Mato Grosso e Mato Grosso do Sul), Sul (Santa Catarina e Rio Grande do Sul), Sudeste (São Paulo, Minas Gerais, Espírito Santo e Rio de Janeiro) e Distrito Federal. Estes participantes encontravam-se envolvidos com atividades de ensino nos âmbitos estadual, federal e particular – apenas o âmbito municipal não foi contemplado na amostra. Além disso, de todos os níveis de ensino – ensino fundamental, ensino médio, ensino superior (Licenciatura em Letras, outras licenciaturas), pós-graduação (Linguística Aplicada, outras), cursos de línguas para diversos níveis, cursos preparatórios, dentre outros –, não houve representantes apenas para a educação infantil.

Questionados a respeito da exigência de nível de proficiência mínimo para contratação docente, 56,9% (33) afirmaram já ter passado pela experiência. Destes, 39,4% (13) declararam ter sido solicitado nível de proficiência igual ou superior àquele para o qual lecionam e 27,3% (9) afirmaram que o nível intermediário superior – B2 do CEFR – seria o mais adequado e requerido para ensino de línguas estrangeiras no Brasil. Um participante, contudo, apontou acreditar ser ideal que o docente esteja apenas um nível acima daquele para o qual leciona, de modo que um conhecimento muito amplo da língua não interfira em sua capacidade de adaptação, contrariando, assim, a perspectiva de Richards *et al.* (2013), os quais atestam a existência de uma relação direta entre o nível de proficiência genérica do candidato e sua capacidade de atuação enquanto professor de línguas.

Com relação a exames que tenham por finalidade avaliar a proficiência de professores de línguas, 56,25% (18) dos participantes demonstraram um possível equívoco com relação ao entendimento dos objetivos deste tipo de avaliação. Para eles, avaliações genéricas, como exames de proficiência consagrados – TOEFL, TOEIC, Michigan e os exames de Cambridge, dos quais mais da metade dos candidatos afirmou já ter participado –, testes de nivelamento e provas de concursos públicos, cumprem essa função. 37,5% (12) dos participantes negaram ter participado de um exame específico e os 6,25% (2) que afirmaram participação não informaram o nome do exame.

Quando inquiridos a respeito do exame EPPL, mais especificamente, houve um relativo equilíbrio entre a quantidade de participantes que conheciam e que não conheciam o exame, conforme exibido na Tabela 1. Dentre os participantes que apontaram como tomaram conhecimento do exame, destaca-se a participação em eventos científicos. Um participante que alegou desconhecimento reclamou da falta de informações disponíveis na internet e outro disse estar interessado em saber mais sobre o exame.

Tabela 1. Tabela de contingência de conhecimento e participação no exame EPPL

	CONHECIMENTO DO EXAME EPPL					
	PARTICIPOU		NÃO PARTICIPOU		TOTAL	
	f	%	f	%	f	%
CONHECE	1	3,44	11	37,93	12	41,37
NÃO CONHECE	0	0	17	58,62	17	58,62
TOTAL	1	3,44	28	96,55	29	100

Fonte: Elaborado pela autora.

Dos participantes que afirmaram desconhecimento, 94,44% (54) informaram ter interesse em participar desse tipo de avaliação. Diversos foram os motivos apontados para participação no EPPL, dentre os quais se destacam a possibilidade de conhecimento de potencialidades e limitações individuais para orientar a busca por aprimoramento linguístico e profissional; a obtenção de uma qualificação profissional; a possibilidade de emprego do exame como uma forma de seleção unificada; e a geração de um provável aprimoramento no ensino de línguas, graças à seleção de profissionais mais qualificados – pensando-se, assim, no possível efeito retroativo a ser gerado pelo exame. Os excertos 01 e 02 ilustram algumas das respostas fornecidas pelos participantes.

Excerto 01:

Sim. Acredito que tal exame certificaria e legitimaria um bom profissional de línguas. Não acredito que a proficiência apenas faça um bom professor, mas um bom professor deve, necessariamente, ter uma boa proficiência.

Participante SNE28. Questionário 1, 2017.

Excerto 02:

Sim, participaria. Considero de total importância tal avaliação para enriquecimento profissional, a partir do *feedback* recebido, e educacional como um todo. Acredito que um professor proficiente na L2 incentivaria à (sic) proficiência dos alunos, já que ele é um modelo.

Participante SNE31. Questionário 1, 2017.

No Excerto 01, é possível notar a existência de uma crença segundo a qual a competência linguística não é, isoladamente, a única competência necessária a um professor de línguas – conforme aponta Almeida Filho (1993); contudo, na opinião do participante, um nível de proficiência satisfatório na língua é essencial para uma atuação eficaz desse profissional – conforme ressaltam Richards *et al.* (2013). No Excerto 02, por sua vez, a crença apresentada corrobora o círculo vicioso de Consolo (2008) ao atribuir um possível impacto positivo no ensino e na aprendizagem de línguas por meio da seleção adequada de profissionais de ensino.

Dos participantes, 13,72% (7) impuseram algumas condições para participação. Dentre elas, encontra-se a necessidade de que o exame se diferencie dos exames internacionais, abordando as necessidades do contexto brasileiro e alinhando-se com a noção de inglês como língua franca; seja de baixo custo; seja aplicado a professores recém-formados – não a professores em serviço –; e que tenha finalidade informativa, e não seletiva. Os candidatos demonstram, ainda, preocupação com a possibilidade de o exame não conseguir atingir níveis de aceitação como os exames de Cambridge. As razões apresentadas demonstram, assim, a

existência de uma preocupação com a qualidade do exame, bem como com a necessidade de haver um cuidado no tratamento da qualidade da proficiência dos professores em serviço, caso o exame seja aceito em larga escala, de modo a não causar prejuízos à categoria. Os excertos 03 e 04 ilustram essas preocupações.

Excerto 03:

Depende de como esse exame fosse elaborado, se levasse ou não em conta o contexto brasileiro. Se fosse nos moldes de outros exames internacionais, não vejo relevância. Além disso, tem a questão da língua. A língua inglesa, por exemplo vem tendo suas variantes questionadas, em favor de um inglês como língua franca, que não obedece as (sic) mesmas normas de variantes consideradas padrão.

Participante SNE18. Questionário 1, 2017.

Excerto 04:

Até poderia participar, mas não vejo como essa prova poderia ser mais renomada ou bem elaborada do que as provas CAE e CPE, logo, não sei se teria alguma utilidade profissional.

Participante SNE37. Questionário 1, 2017.

Apenas 5,55% (3) dos respondentes negaram categoricamente a participação nesse tipo de avaliação. Os argumentos utilizados por eles para justificar a não participação no exame foram a agressividade que esse tipo de avaliação impõe aos cursos de formação de professores, a não utilidade desse tipo de avaliação frente à existência de outras abordagens avaliativas ou, ainda, à relevância da competência linguística em detrimento das demais competências necessárias à atuação do profissional de línguas, demonstrando um desconhecimento, por parte deste público, das especificidades da proficiência desse professor. Os excertos 05, 06 e 07 ilustram algumas dessas opiniões.

Excerto 05:

Não. Acho desnecessário e redutor, pois pressupõe um entendimento de que para ser professor de línguas bastaria ter proficiência linguística. Ou seja, reforça um importante mito contra o qual nós, pesquisadores envolvidos na educação linguística em língua adicional, temos de combater com firmeza.

Participante SNE16. Questionário 1, 2017.

Excerto 06:

Não, pois tenho comprovação de proficiência anterior ao meu ingresso na universidade como docente.

Participante SNE57. Questionário 1, 2017.

Excerto 07:

Sou contra a existência desse exame, pois funciona como uma declaração de que nosso sistema educacional não dá conta de formar professores com proficiência adequada ao ensino do inglês no contexto da comunicação internacional. Ainda que nosso sistema educacional enfrente desafios diversos, a composição de um exame de proficiência se alinha com a reprodução do mito segundo o qual não

aprenderíamos inglês a contento na escola, sendo que, na verdade, para longe de ser uma constatação, este mito produz e sustenta um mercado em que o setor privado se interessa e explora.

Participante SNE28. Questionário 1, 2017.

No que se refere aos critérios delimitados por esses usuários para avaliação da proficiência de professores de línguas, tem-se que eles apontam para uma configuração que atrela a competência linguística à didática. Os critérios foram levantados por meio de questionamentos que envolviam a opinião dos participantes a respeito (i.) dos aspectos que caracterizam um bom professor de línguas, (ii.) das características da proficiência do professor valorizadas pelos empregadores e, mais especificamente, (iii.) dos elementos linguísticos, comunicativos e pedagógicos que devem ser avaliados por um exame de proficiência para professores de línguas. A relação desses critérios foi estabelecida seguindo-se as orientações propostas pelo método de estruturação de problemas VFT, analisando-se a proximidade semântica dos vocábulos com as competências docentes propostas em Almeida Filho (1993) e pelas competências comunicativas propostas pelo CEFR (COE, 2018).

A Figura 12 exhibe as competências e os critérios destacados pelos participantes como relevantes na atuação do professor de línguas, os quais, por extensão, deveriam ser avaliadas por um exame como o EPPL. Com relação aos critérios listados pelos participantes, percebe-se a possibilidade de existência de um alinhamento entre o que os eles consideram essenciais e as competências linguístico-comunicativo-pedagógicas listadas no CEFR e no CETF. O Quadro 10 a seguir ilustra as relações.

Quadro 10. Relação entre as competências e os critérios do CEFR e do CETF e a opinião dos participantes.

REFERÊNCIA	COMPETÊNCIAS	CRITÉRIOS	CONSIDERADOS
CEFR	Estratégica	Produção	x
		Compreensão	x
		Mediação	x
		Interação	x
	Linguística	Precisão lexical	x
		Precisão gramatical	x
		Precisão fonológica	x
		Complexidade lexical	x
	Pragmático-discursiva	Complexidade sintática	x
		Desenvolvimento temático	x
		Tomada de turno	-
		Coerência e coesão	x
	Pragmático-funcional	Fluência	x
		Flexibilidade	x
Pragmático (esquema)	Precisão proposicional	x	

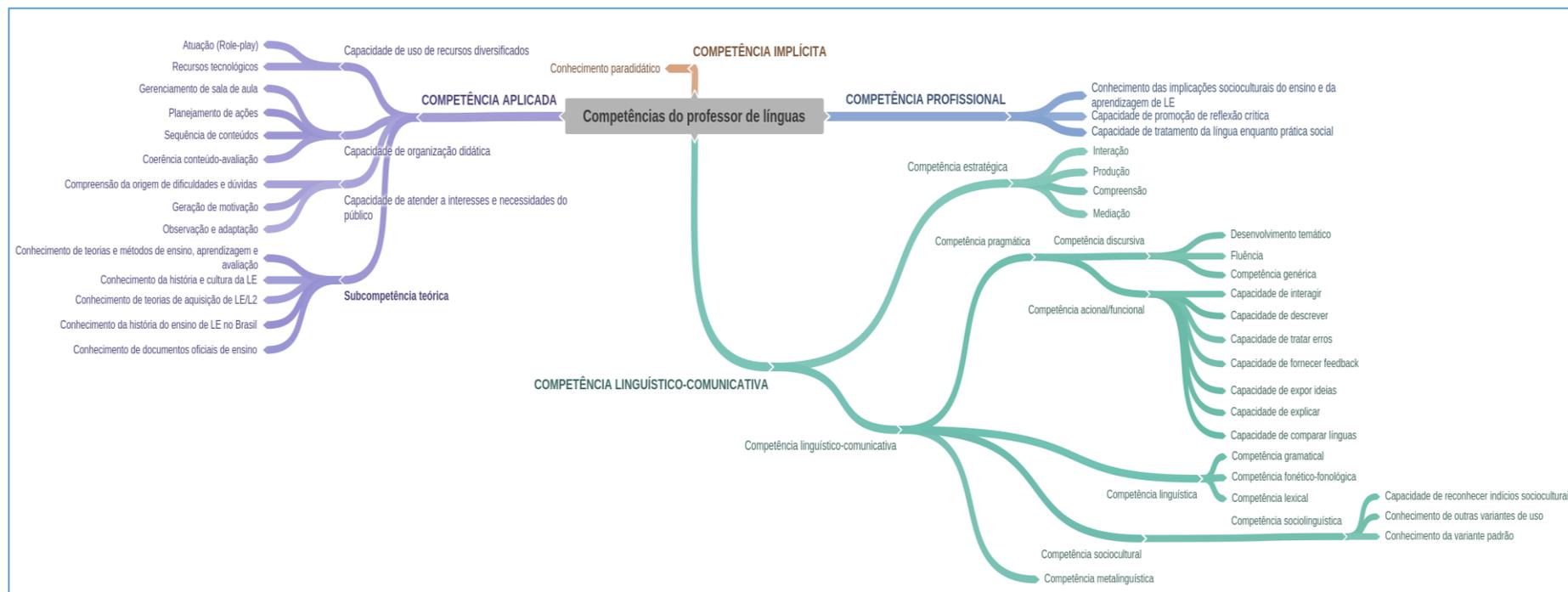
	Sociolinguística	Nível de formalidade	-
		Uso das funções comunicativas	x
		Socialização	-
		Identificação de pistas socioculturais	x
		Propriedade de registro	x
CETF	Capacidade linguística	Linguagem de sala de aula	x
		Modelo de língua	x
		Reconhecimento de erros	x
		Comunicação com outros profissionais	-
	Conhecimento linguístico	Consciência linguística	x
		Terminologia para descrição de língua	x
		Materiais de referência	-

Fonte: Elaborado pela autora.

O fato de algumas respostas abordarem conceitos referentes a competências de caráter didático que não a linguístico-comunicativa confirma um possível choque de interesses e compreensão acerca dos propósitos de um exame com as finalidades do EPPL. Quando questionados especificamente sobre quais deveriam ser os objetivos desse tipo de avaliação, 73,21% (41) dos participantes informaram que o objetivo deveria ser avaliar aspectos da competência linguístico-comunicativa, contra 26,78% (15) que acreditavam que, além dos aspectos linguísticos, deveriam ser incluídos aspectos de avaliação da competência didática.

Os respondentes também apontaram algumas características que eles acreditavam ser essenciais para a atuação de um bom professor e, conseqüentemente, de uma boa aula de línguas, como dinamicidade, empatia, responsabilidade, paciência, tolerância, carisma, simpatia, curiosidade, disposição para aprender e experiência profissional; bem como as condições que devem se fazer presentes no ambiente de ensino-aprendizagem para que o processo seja bem sucedido, como flexibilidade e abertura por parte do docente pra preparar aulas, condições para pesquisa e disponibilidade para atendimento ao aluno. No entanto, não se pode estabelecer uma relação dessas características e condições com o conceito de proficiência linguístico-comunicativa-pedagógica, ou seja, tais aspectos não são relevantes para a definição do conceito e, portanto, não foram consideradas na análise.

Figura 12. Competências e critérios relevantes para a atuação do professor de línguas segundo usuários do exame.



Fonte: Elaborado pela autora.

No âmbito dos usuários especialistas diretamente envolvidos com o exame, em 2017, foram obtidas cinco respostas de representantes do Norte (Acre), Centro-Oeste (Distrito Federal), Sul (Paraná) e Sudeste (São Paulo), dos quais quatro atuavam como professores de línguas e um como aplicador de exames de proficiência. Dos professores, 50% afirmaram que o nível avançado de proficiência na língua fora exigido para a contratação por meio de concurso público e 50% afirmaram não haver exigência de nível mínimo. Todos os especialistas encontravam-se envolvidos com o desenvolvimento e/ou a operacionalização do exame EPPLE.

Dentre os critérios elencados como importantes componentes da proficiência linguístico-comunicativo-pedagógica, encontram-se os listados na Figura 13. As respostas foram obtidas com base nas opiniões dos participantes com relação (i.) aos descritores fundamentais de uma escala de um exame como o EPPLE, (ii.) aos modos de mensuração destes descritores, (iii.) às características distintivas essenciais aos níveis de proficiência de faixas para o EPPLE. O Quadro 11 evidencia a relação entre os critérios propostos pelos participantes e os considerados no CEFR e no CETF.

Quadro 11. Relação entre as competências e os critérios do CEFR e do CETF e a opinião dos participantes.

REFERÊNCIA	COMPETÊNCIAS	CRITÉRIOS	CONSIDERADOS
CEFR	Estratégica	Produção	x
		Compreensão	x
		Mediação	x
		Interação	x
	Linguística	Precisão lexical	x
		Precisão gramatical	x
		Precisão fonológica	-
		Complexidade lexical	-
	Pragmático-discursiva	Complexidade sintática	-
		Desenvolvimento temático	x
		Tomada de turno	-
		Coerência e coesão	x
	Pragmático-funcional	Fluência	x
		Flexibilidade	x
	Pragmático (esquema)	Precisão proposicional	x
		Sociolinguística	Nível de formalidade
Uso das funções comunicativas	x		
Socialização	-		
Identificação de pistas socioculturais	-		
CETF	Capacidade linguística	Propriedade de registro	x
		Linguagem de sala de aula	x
		Modelo de língua	x
		Reconhecimento de erros	x

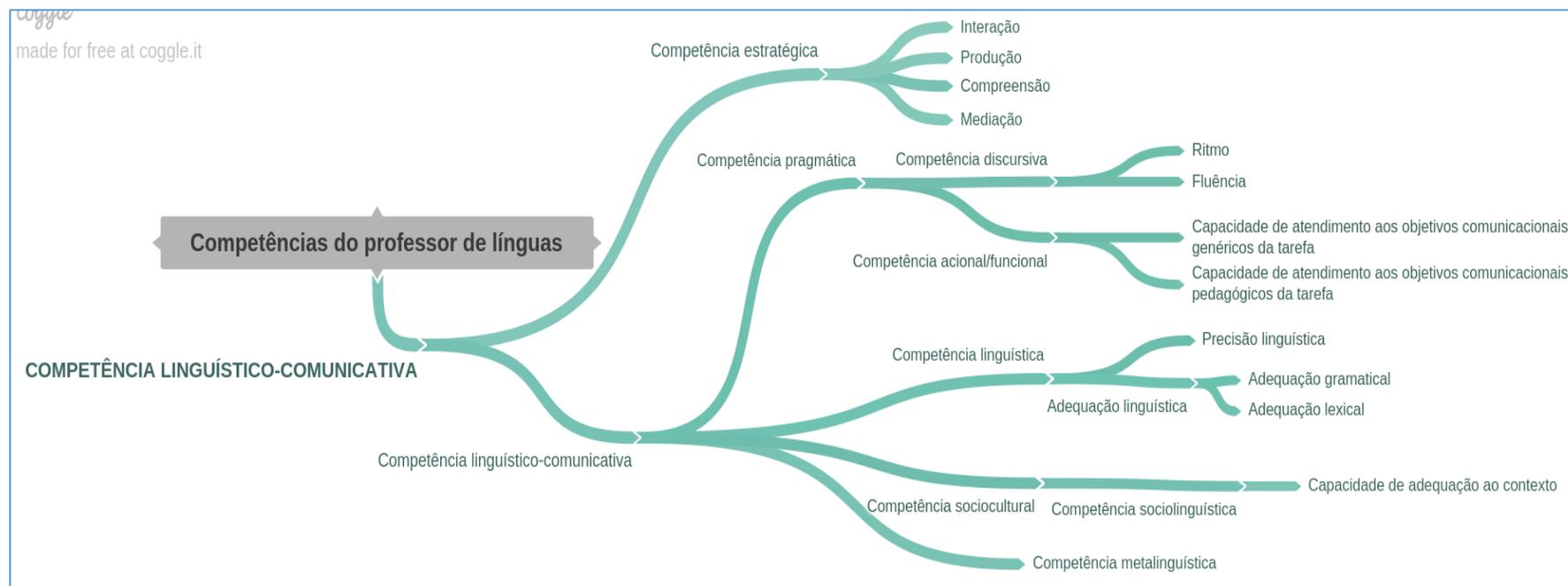
		Comunicação com outros profissionais	-
	Conhecimento linguístico	Consciência linguística	x
		Terminologia para descrição de língua	x
		Materiais de referência	-

Fonte: Elaborado pela autora.

Com relação ao modelo de correção do exame, os candidatos apontaram problemas na quantidade de avaliadores disponíveis; no modo como esses avaliadores são selecionados, fazendo com que a correção fique centrada em poucas pessoas; e na eficácia da organização da correção e da divulgação dos resultados. Sobre a escala, os participantes mostraram-se satisfeitos com a diversidade de aspectos linguísticos contemplados e com a autenticidade situacional dos critérios avaliados. No entanto, exibiram preocupação com relação à ausência de escalas holísticas e analíticas independentes, o que, segundo eles, pode confundir os examinadores no momento da avaliação; e de coerência entre os objetivos comunicacionais efetivamente angariados pelas tarefas e os abordados na escala.

Dentre os possíveis impactos positivos gerados pelo exame, os participantes listaram melhorias nas práticas de ensino de língua inglesa, graças a efeitos retroativos gerados na formação docente, bem como na contratação de professores graças à certificação. Tais considerações alinham-se com os efeitos positivos apontados pelos participantes não-especialistas, demonstrando a existência de expectativas sociais majoritariamente positivas com relação à aplicação do exame.

Figura 13. Competências e critérios relevantes para a atuação do professor de línguas segundo os envolvidos com a elaboração do exame.



Fonte: Elaborado pela autora.

4.1.2 As respostas fornecidas aos questionário sobre mensuração da proficiência

O segundo questionário, aplicado aos usuários não envolvidos diretamente com o exame no ano de 2018, contou com 17 participantes. Inicialmente, o questionário abordou como opções de critérios de avaliação da proficiência oral do professor de línguas aqueles mais recorrentemente apontados pelos participantes no primeiro questionário. Desse modo, a avaliação contemplava a capacidade de compreensão oral, a capacidade de uso da língua para finalidades gerais e pedagógicas, e aspectos da competência linguístico-comunicativa. Do total de participantes, 82,4% (14) consideraram todos os referidos critérios suficientemente adequados e igualmente importantes para a avaliação, contra 17,6% (3), que discordaram.

Dos participantes que consideraram os critérios adequados, 93,3% (14) disseram acreditar que todos os critérios eram passíveis, subjetiva ou objetivamente, de mensuração. Destes, 100% concordaram que a avaliação da proficiência deve ocorrer por meio de uma abordagem *top-down*, ou seja, dos conteúdos mais genéricos para os mais específicos, iniciando-se pela avaliação da capacidade de compreensão oral do candidato, passando para a avaliação da competência funcional em relação ao uso da língua para finalidades genéricas, seguindo para a avaliação de elementos linguístico-comunicativos e, por fim, focando-se na avaliação da competência funcional em relação ao uso da língua para finalidades pedagógicas.

O Quadro 12 resume os critérios e as mensurações consideradas importantes na opinião de mais de 50% dos participantes.

Quadro 12. Resumo da importância dos critérios na opinião dos *stakeholders* não-especialistas

CRITÉRIO	MODO DE MEDIÇÃO	UNIDADE DE MEDIDA
Compreensão oral (14 respostas)	100% Cumprimento da tarefa	78,6% Grau de completude da resposta
Contribuição com a interação (14 respostas)	92,9% Propriedade de ideias	85,7% Grau de relevância das ideias
		85,7% Grau de relação com o tema
		57,1% Quantidade de ideias
	57,1% Organização discursiva	64,3% Grau de clareza das ideias
		71,4% Grau de progressão textual
Uso de vocabulário (14 respostas)	71,4% Precisão	71,4% Grau de adequação ao significado
	92,9% Adequação	92,9% Grau de adequação ao contexto
	85,7% Variedade	57,1% Quantidade de <i>tokens</i> e <i>types</i>
Uso de gramática (14 respostas)	92,9% Acurácia	71,4% Quantidade de erros estruturais (nível da palavra/frase)
		71,4% Grau de variedade de erros estruturais (nível da palavra/frase)
	64,3% Complexidade	57,1% Produto da divisão da quantidade total de orações isoladas e de <i>C-units</i>
Uso de elementos fonético-fonológicos	69,2% Tom	84,6% Grau de variedade
	69,2% Acento	84,6% Grau de tipicidade

(14 respostas)	61,5% Ritmo	76,9% Grau de tipicidade	
	53,8% Velocidade de fala	0% Quantidade de sílabas/segundo*	
	100% Fluidez discursiva		85,7% Motivo de hesitações
			57,1% Frequência e duração de pausas (preenchidas e silenciosas)
			57,1% Frequência de repetição
			57,1% Frequência de reparo
	92,3% Inteligibilidade	92,3% Grau de efetividade da compreensão	
69,2% Compreensão	76,9% Grau de dificuldade de compreensão		
Conhecimento de tópico (12 respostas)	100% Cumprimento da tarefa	91,7% Grau de adequação da resposta	
		50% Grau de completude da resposta	
Adequação da linguagem docente (13 respostas)	84,6% Substituição/Reformulação	0% Frequência*	
	76,9% Checagem de compreensão/ <i>Feedback</i>	69,2% Grau de adequação contextual	
Fornecimento de instruções (10 respostas)	70% Variedade vocabular	60% Quantidade de <i>tokens</i> e <i>types</i>	
	60% Precisão vocabular	70% Grau de adequação ao significado	
Uso de terminologia específica (13 respostas)	92,3% Precisão vocabular	84,6% Grau de adequação ao significado	
Fornecimento de explicação (13 respostas)	61,5% Descrição do uso	61,5% Grau de adequação da resposta	
	69,2% Exemplificação	76,9% Grau de adequação da resposta	
	69,2% Elicitação de participação	61,5% Frequência	
	61,5% Fornecimento de correção	61,5% Grau de adequação	

Fonte: Elaborado pela autora.

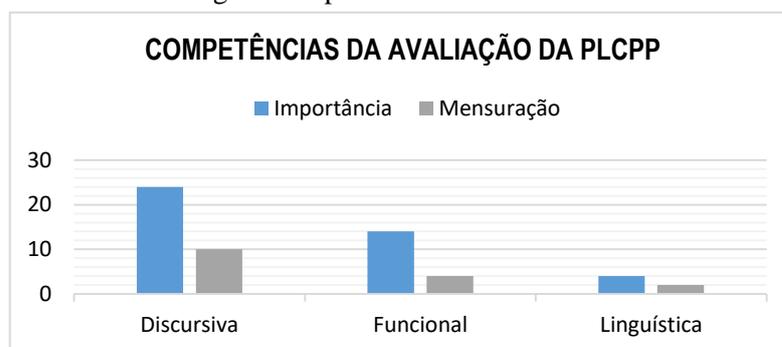
*Não foram apontadas alternativas de mensuração pelos participantes.

Os participantes demonstraram, ainda, algumas preocupações com relação aos critérios. Sobre a fluidez, um participante demonstrou preocupação com os limiares de mensuração do critério em virtude da carga emocional angariada pela própria situação de testagem, o que poderia levar o candidato a cometer um maior número de rupturas e hesitações na fala, conforme descrito em Weir (2005b). No que concerne à complexidade sintática, a preocupação foi relatada em virtude dos possíveis limiares para avaliação da frequência de subordinação na fala dos candidatos, uma vez que este recurso é mais característico de textos escritos do que falados. Do mesmo modo, um participante sugeriu que os limiares evitassem o emprego de variáveis linguísticas não modalizadas, ou seja, variáveis que descrevessem desempenhos com características extremas e binárias, como “adequado/não adequado” ou “completo/não completo”. Um participante sugeriu, por fim, que ao critério “fornecimento de instruções” fossem acrescentados modos de mensuração de base discursiva, e não apenas lexical.

Os participantes que não consideraram os critérios listados suficientemente adequados apontaram como elementos relevantes para avaliação da proficiência do professor de línguas componentes das competências pragmático-discursiva e pragmático-funcional, como adequação ao contexto, clareza e relevância de ideias, precisão proposicional e fluidez discursiva. No quesito linguístico, apenas a acurácia gramatical e o uso de metalinguagem foram considerados. A importância atribuída aos critérios mostra uma preocupação maior com

os aspectos da competência pragmática em detrimento dos da linguístico-comunicativa, conforme exibido no Gráfico 1.

Gráfico 1. Competências da avaliação da proficiência do professor de línguas na opinião dos usuários do exame.



Fonte: Elaborado pela autora.

A análise das respostas dos questionários enviados aos usuários elaboradores, por sua vez, contou com apenas três participações e possibilitou a definição de um perfil no qual os elementos considerados conscientemente na avaliação são limitados em número. Quando solicitados a avaliarem o grau de suficiência da proficiência demonstrada em dados de fala de cinco candidatos ao EPPLE, os critérios empregados nas justificativas foram, em ordem de importância, fluência, representada pela presença ou ausência de interrupções; complexidade linguística e adequação da linguagem docente; precisão linguística; e inteligibilidade.

Dos elementos apresentados no Quadro 12, os participantes, em concordância com os não-especialistas, consideram como a quantidade e o grau de variação de *AS-units* inapropriados para mensuração da acurácia gramatical e a contagem de sílabas por segundo inapropriada para avaliação da velocidade de fala. Em discordância com os não-especialistas, por sua vez, os participantes não acreditam na relevância do produto da divisão da quantidade total de orações isoladas e de *C-units* como meio de mensuração da complexidade sintática, da variedade vocabular para avaliação da capacidade de fornecimento de instruções e da elicitación de participação para a avaliação da capacidade de fornecimento de explicações. Ressalta-se, contudo, que os participantes não apresentaram alternativas de mensuração para nenhum dos critérios.

4.1.3 A estruturação do problema: considerações gerais sobre a opinião dos usuários

Em resumo, as respostas dos candidatos permitem perceber que, apesar de os usuários (*stakeholders*) não diretamente envolvidos com a elaboração do EPPL, em um primeiro momento, tenderem a valorizar elementos de ordem da competência pragmática em detrimento dos componentes da competência linguística, a porcentagem de participantes que concorda com a avaliação dos elementos linguísticos, quando informados a respeito dos modos de mensuração, é igualmente alta. Do mesmo modo, as respostas também evidenciam uma aceitação de modos de mensuração de caráter mais objetivo, ou seja, numérico, além da aceitação da existência de limitação da capacidade cognitiva de processamento de um grande número de critérios por parte dos avaliadores humanos no momento da avaliação, o que propicia a proposição de abordagens avaliativas mais automatizadas.

Deve-se considerar, contudo, as preocupações apresentadas por alguns participantes em relação à definição das variáveis linguísticas de mensuração, às características da linguagem oral e à influência de questões de ordem emocional no desempenho dos candidatos, além do controle de impactos negativos possivelmente gerados pela abordagem avaliativa do exame na vida profissional dos candidatos.

Assim, os questionários aplicados, seguindo-se as orientações do VFT, resultaram no Quadro 13.

Quadro 13. Hierarquização dos objetivos do exame EPPL

Objetivo estratégico:	Maximizar a avaliação da proficiência linguístico-comunicativo-pedagógica
Objetivo fundamental:	Maximizar a avaliação da compreensibilidade da fala do professor de línguas
Objetivos meios:	<p>1 Avaliar o uso da competência pragmática para finalidades gerais e pedagógicas</p> <hr/> <p>Avaliar a flexibilidade do uso das funções comunicativas</p> <p>Avaliar o desenvolvimento temático</p> <p>Avaliar a coerência das ideias</p> <p>Avaliar a precisão proposicional</p> <p>Avaliar a fluência</p> <hr/> <p><i>Velocidade de fala</i></p> <hr/> <p><i>Rupturas</i></p> <hr/> <p><i>Reparos</i></p> <hr/> <p>2 Avaliar o uso da competência linguístico-comunicativa</p>

Avaliar a inteligibilidade da fala
<i>Divergência fonêmica</i>
<i>Acento</i>
<i>Entoação</i>
Avaliar a precisão do uso dos elementos
<i>Gramaticais</i>
<i>Lexicais</i>
<i>Sociolinguísticos</i>
<i>Metalinguísticos</i>
Avaliar a complexidade de uso dos elementos
<i>Sintáticos</i>
<i>Lexicais</i>

Fonte: Elaborado pelo autor.

4.2 Avaliação da proficiência do professor de línguas pelo EPPL

A avaliação da proficiência oral do professor de línguas pelo EPPL foi investigada neste estudo levando-se em consideração as características da escala de proficiência oral empregada pelo exame, as características das tarefas do teste oral e o desempenho de fala dos candidatos, tanto em relação a seus perfis quanto aos conceitos a eles atribuídos.

4.2.1 O construto informado e representado do exame: hierarquização dos objetivos da escala de proficiência oral

O construto do exame EPPL indica que ele tem por objetivo avaliar a proficiência de uso da competência linguística do candidato para tarefas que a exijam em contexto de ensino de línguas na escola regular de Ensino Fundamental, Médio e Superior e nos cursos de idiomas. Conforme apontado em Consolo (2008, 2011a, 2011b), a competência linguística do professor de línguas engloba não apenas aspectos de ordem puramente estrutural, mas, sim, de ordem comunicativa e, especificamente, pedagógica, conforme representados na escala de proficiência oral do EPPL.

Os critérios de avaliação que compõem a escala foram definidos com base em pesquisas que investigavam contextos representativos da realidade de cursos de Licenciatura em Letras

no Brasil. Na faixa mais alta (Faixa A), os critérios são apresentados por meio de capacidades, como: atingir, apresentar, exibir, fornecer, utilizar, narrar, descrever, levantar, estabelecer, sugerir, expressar, selecionar, esclarecer, explicar, fazer uso, apresentar e ser compreendido. Percebe-se, pela escolha lexical das capacidades, que, na faixa A, o desempenho do candidato é representado com base naquilo que ele se mostra capaz de fazer. O excerto 08 exemplifica os descritores dessa faixa.

Excerto 08:

- A1) Atinge plenamente os objetivos de comunicação verbal, apresentando fluência em 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, sem 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.

Fonte: Extraído de Consolo e Teixeira da Silva (2014).

No entanto, conforme os níveis das faixas progressivamente diminuem, os verbos passam a ser acompanhados de palavras de sentido negativo, como “sem”, “não”, “distinto”, “interferência”, “limitado”, “parcial”, “porém”, “dificuldade”, “simples”, “pouco”, “nenhum”, “prejudicar”, “prejuízo”, “falta” e “erro”. O excerto 09 exemplifica os descritores da Faixa E.

Excerto 09:

- 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.

Fonte: Extraído de Consolo e Teixeira da Silva (2014).

Assim, tem-se que o desempenho do candidato passa a ser descrito por meio de suas incompetências, ou seja, por meio do que o candidato não é capaz de fazer, em uma orientação negativa. Nota-se, assim, que as características distintivas (JIN; MARK, 2012) entre uma faixa e outra na escala do exame são progressivamente marcadas pelas gradações daquilo que o candidato perde competência para realizar à medida em que o nível da proficiência diminui. No Apêndice I é possível encontrar uma comparação das características distintivas que determinam as faixas cujos limiares são bastante próximos. No referido apêndice, é possível perceber que as faixas contemplam três critérios de avaliação da proficiência bastante genéricos, ou seja, critérios que, por sua amplitude, acabam por perpassar a avaliação de desempenho dos candidatos em todas as tarefas do teste oral, não se restringindo a nenhuma tarefa em específico, sendo eles: (i.) a capacidade de atingir os objetivos da comunicação verbal, (ii.) a qualidade da pronúncia, e (iii.) a capacidade de fazer uso de estruturas linguísticas.

Nota-se que a escala do teste oral assume que o primeiro critério - atingir os objetivos da comunicação verbal – deve ser avaliado por meio da fluência, a qual é introduzida na escala por meio de seus meios de mensuração: velocidade e ritmo de fala. No entanto, tais meios são expressos apenas na primeira e na segunda faixa da escala – nesta última de maneira incompleta, pois não há menção à velocidade. Ressalta-se, ainda, que, nas duas últimas faixas (D e E), não são elencados elementos diferenciadores para o critério “fluência”. Do mesmo modo, o primeiro critério não é diferenciado em candidatos aos quais sejam atribuídas as duas primeiras faixas, ocorrendo a diferenciação apenas a partir da terceira, conforme se observa no excerto 10.

Excerto 10:

- A1) Atinge plenamente os objetivos de comunicação verbal, apresentando fluência em velocidade e ritmo de fala.
- 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 autocorriger.
- 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.
- D1) Atinge os objetivos de comunicação verbal com limitações. Exibe falta de fluência e de competência no uso de estruturas linguísticas.
- E1) Não atinge satisfatoriamente os objetivos de comunicação verbal, apresentando falta de fluência e de competência na produção oral.

Fonte: Extraído de Consolo e Teixeira da Silva (2014).

Já o segundo critério – qualidade de pronúncia – é introduzido na escala por meio da influência de padrões da língua materna do candidato e dos conceitos de inteligibilidade e compreensibilidade. O descritor de número 2 tem caráter genérico, uma vez que não se atrela apenas a uma tarefa do exame.

O falante ideal adotado como padrão distancia-se da figura do falante nativo, mas permite um grau de subjetividade amplo ao prender-se à figura de “falantes competentes da língua estrangeira”. Assim, embora o construto do exame não defina quais características de pronúncia apresenta o sujeito “falante competente” e limite a comparação da produção do candidato à fala desse sujeito, bastante amplo, percebe-se aqui um respeito aos ‘World Englishes’ (RAJAGOPALAN, 2009, 2012), sem a marca do idealizado, e controverso, falante nativo.

Nas faixas A, B e E, a inteligibilidade não é explicitamente mencionada, prevalecendo apenas a proximidade e a interferência da língua materna na produção. A progressividade das faixas permite-nos inferir, contudo, que em A e B a inteligibilidade é preservada, uma vez que é apresentada como tal em faixas inferiores, como C e D; mas que, em E, ela acaba sendo prejudicada pelo caráter destoante da produção do candidato à fala de falantes competentes da língua estrangeira. Nesse mesmo sentido, é válido mencionar que as faixas C, D e E substituem a figura dos “falantes competentes da língua estrangeira” para um sujeito ainda mais genérico: os “falantes da língua estrangeira”, reforçando a característica descritiva de incompetências por parte da escala conforme os níveis das faixas progressivamente diminuem.

Percebe-se, ainda, que a influência da língua materna não é apresentada como característica distintiva entre as faixas A e B, para as quais não se admitem influências marcantes, nem para as faixas C e D, para as quais a habilidade da pronúncia não é mencionada.

A singularidade atribuída ao descritor “pronúncia” pela escala do exame aponta para a importância a ele conferida pelo teste. No entanto, nota-se que a avaliação da pronúncia não leva em consideração características como acento e propriedade de ritmo (SAWAKI, 2007) - tendo sido este último avaliado na habilidade “fluência” -, mas, sim, foca-se na compreensibilidade, englobando aspectos como a identidade de fala e, por extensão, a entoação e os sons, preocupando-se majoritariamente e novamente, portanto, com o alcance dos objetivos da comunicação verbal. No entanto, percebe-se que a compreensibilidade é abordada apenas nas faixas inferiores (D e E), sendo retomada novamente apenas na faixa A, o que pode indicar uma crença por parte dos elaboradores da escala segundo a qual essa seria uma característica de falantes com nível de proficiência inferior. No entanto, conforme ressalta Thomson (2018), a compreensibilidade é afetada, dentre outros elementos, pelo sotaque, o qual não indica, necessariamente, níveis inferiores de proficiência oral.

O excerto 11 exhibe o segundo descritor.

Excerto 11:

- A2) Exibe padrões de pronúncia bastante semelhante a falantes competentes da língua estrangeira, sem influências marcantes dos padrões de sua língua materna, sem causar qualquer desconforto ou incompreensão por parte do interlocutor.
- 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.
- C2) Exibe pronúncia compreensível, porém com alguns desvios com relação aos padrões de falantes competentes da língua estrangeira.
- 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.
- 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.

Fonte: Extraído de Consolo e Teixeira da Silva (2014).

O descritor de número três consiste em um descritor específico, atestando o uso de estruturas linguísticas para o atingimento de funções comunicativas específicas. O terceiro critério - uso de estruturas linguísticas - aparece interligado à capacidade do aluno de realizar autocorreções na segunda faixa do teste (B). Na primeira faixa (A), contudo, ressalta-se que o descritor não é mencionado. No entanto, pela correlação demarcada na faixa B entre o uso de estruturas linguísticas e a capacidade de autocorreção do candidato, pressupõe-se que, na faixa A, o candidato seja capaz de fazer uso de estruturas linguísticas de modo a não cometer deslizes quanto ao emprego dos elementos estruturais da língua, consistindo, assim, em uma faixa com um nível maior de idealização, conforme exibido no excerto 12. O CEFR, contudo, considera a flexibilidade e a realização de monitoramento e reparo da fala como capacidades representativas, inclusive, de níveis de proficiência superiores (COE, 2018).

Excerto 12:

- 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.
- 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.
- C3) Fornece informações sobre experiências presentes e passadas utilizando estruturas linguísticas simples e pouca variedade lexical.
- D3) Fornece informações sobre experiências presentes e passadas utilizando estruturas linguísticas simples, e pouca variedade lexical.
- 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.

Fonte: Extraído de Consolo e Teixeira da Silva (2014).

A clareza no fornecimento de informações é um elemento distintivo da faixa A em relação a todas as outras faixas do exame – para as quais o elemento distintivo se perde. As

faixas A e B diferenciam-se, por extensão, apenas pela questão da clareza, uma vez que não há elemento distintivo entre o uso de estruturas linguísticas e o uso de vocabulário e itens lexicais. Não há qualquer elemento diferenciador entre as faixas C e D, sendo a faixa E individualizada em relação a estas apenas por meio do elemento “estruturas linguísticas”, o qual, na última faixa, além de simples, deve implicar em prejuízo para a comunicação.

O fato de a escala recuperar os elementos estruturais e introduzir os itens lexicais como elementos de medição da capacidade de fornecer informações destaca o caráter comunicativo do teste, uma vez que os elementos propriamente linguísticos são empregados com uma finalidade maior: comunicar experiências. Além disso, percebe-se que, no terceiro descritor, o critério aparece interligado a funções comunicativas bastante específicas, intimamente ligadas às tarefas do exame, ou seja, tem caráter fortemente descritivo, o que prejudica seu potencial inferencial.

A combinação das habilidades de “atingir os objetivos de comunicação verbal”, “pronúncia” e “utilizar estruturas linguísticas” delimita a concepção de competência linguístico-comunicativa do exame, uma vez que condensa, em um único descritor, elementos de ordem linguística e comunicativa em uma hierarquia na qual o primeiro é empregado como elemento que engloba os demais. A esse respeito, é válido mencionar, contudo, a existência de uma generalização da competência de produção oral do aluno na faixa E, uma vez que o espaço destinado à descrição da habilidade de uso de estruturas linguísticas aparece por meio da expressão “produção oral”, demonstrando a importância do elemento estrutural para o teste. Também contribui para essa interpretação a inexistência de elementos distintivos entre as primeiras faixas da escala no quesito comunicacional, mas a marcação de tais elementos nos dois critérios de ordem linguística.

No descritor de número quatro, são apresentadas as capacidades de narrar e descrever uma cena de vídeo, levantar hipóteses a respeito da cena, estabelecer associações com outras experiências e sugerir contribuições para o ensino e aprendizagem de língua estrangeira. Todas as habilidades apresentam como elemento comum a cena de vídeo em questão e, portanto, consistem em descritores de ordem restrita, atrelando-se a tarefas específicas do exame, conforme se observa no excerto 13.

Excerto 13:

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.

- 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.
- 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.
- 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.
- 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.

Fonte: Extraído de Consolo e Teixeira da Silva (2014).

No excerto 13, percebe-se que, nas faixas A e B, os elementos diferenciadores ficam por conta do elemento narração, que não é apresentado na faixa B, e, também, por conta de certa dificuldade por parte do candidato no estabelecimento de associações e na sugestão de contribuições ao ensino-aprendizagem, também na faixa B. Não há elementos diferenciadores entre as faixas C e D, marcadas pela habilidade de descrição e pela dificuldade em desempenhar as demais capacidades. A faixa E, por fim, diferencia-se das faixas C e D apenas pela dificuldade no fornecimento de descrição. Percebe que as faixas para este descritor levam em consideração o nível de abstração de cada capacidade. Assim, quanto maior o nível de abstração requerido do candidato, maior a faixa que representa sua proficiência.

No entanto, é válido destacar que, a partir deste descritor, a escala passa a requer capacidades de ordem pedagógica de maneira mais explícita. Na faixa B, é possível perceber que a incapacidade de desempenhar as capacidades pedagógicas 3 e 4 é o elemento diferenciador da faixa A, uma vez que as habilidades 1 e 2 são positivamente preservadas em B. Assim, de acordo com a escala de proficiência oral do EPPLE, para que a um candidato do exame seja atribuída a faixa mais alta, é necessário que ele apresente domínio sobre os elementos de ordem linguística e comunicativa igualmente aos elementos pedagógicos, ou seja, apenas o conhecimento de duas dimensões não é suficiente para tal.

O descritor de número cinco é exemplo da preocupação por parte do exame, desde seus primórdios, com o elemento pedagógico, abordando de maneira explícita e individualizada o critério “metalinguagem”. Embora tenha deixado de ser o único elemento mensurador da competência específica do professor de línguas e tenha sido incorporada à escala geral do teste oral, a metalinguagem ainda ocupa posição de destaque, sendo mensurada em termos de conhecimento de regras de uso da língua, seleção de informações para esclarecer dúvidas,

explicação de regras linguísticas, uso de terminologia específica e compreensibilidade por parte de alunos, conforme exhibe o excerto 14.

Excerto 14:

- 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.
- 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.
- 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.
- 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.
- 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.

Fonte: Extraído de Consolo e Teixeira da Silva (2014).

Em continuidade ao descritor anterior, as características distintivas da avaliação da metalinguagem excluem da faixa A candidatos que apresentem dificuldades em selecionar informações para esclarecer dúvidas, explicar regras linguísticas, fazer uso de terminologia específica e se fazer compreensível aos alunos, ressaltando a importância conferida à dimensão linguístico-comunicativa e, mais especificamente, à dimensão pedagógica por parte do teste. Ressalta-se que o descritor de número cinco é o descritor do teste oral que mais apresenta características distintivas entre uma faixa e outra.

O último descritor do teste oral do EPPLLE refere-se à capacidade do candidato de compreender a fala em LE em ritmo normal e, diante dessa compreensão, se são ocasionados problemas em sua produção verbal. O fato de a habilidade de compreensão oral ser avaliada em um descritor específico acaba por permitir a inferência de que os demais descritores da escala se referem diretamente à produção oral. No entanto, o modo como a compreensão é apresentada no último descritor, em relação com a produção, acaba por demonstrar a visão das habilidades empregada pelo teste: uma visão integradora. Assim, percebe-se que o critério “compreensão”, embora tenha um descritor específico e embora haja uma tarefa específica de compreensão oral no teste, perpassa todas as tarefas, uma vez que é apresentado como elemento influenciador indissociável do desempenho verbal do candidato – já que, se o candidato não for capaz de

compreender as tarefas oralmente, ele não conseguirá realizá-las. O excerto 15 exemplifica o descritor de número seis.

Excerto 15:

- A6) Não apresenta dificuldades de compreensão da fala na língua estrangeira em ritmo normal.
 B6) Não apresenta dificuldades de compreensão da fala na língua estrangeira em ritmo normal.
 C6) Pode apresentar dificuldades de compreensão da fala em língua estrangeira mas não a ponto de prejudicarem seu desempenho verbal.
 D6) Apresenta dificuldades de compreensão da fala na língua estrangeira em ritmo normal, as quais podem, ocasionalmente, prejudicar o desenvolvimento de seu desempenho verbal.
 E6) Apresenta dificuldades de compreensão da fala em língua estrangeira em ritmo normal, as quais prejudicam seu desempenho verbal.

Fonte: Extraído de Consolo e Teixeira da Silva (2014).

Conforme se observa no excerto 15, não há características distintivas entre as faixas A e B. Na faixa C, o elemento diferenciador é a ocasionalidade da dificuldade de compreensão, e, nas faixas D e E, diz respeito ao grau de interferência no desempenho verbal do candidato. A dificuldade está presente no descritor C, mas não há interferência no desempenho verbal.

De um modo geral, percebe-se que a escala de proficiência oral contempla descritores tanto de ordem genérica, ou seja, que perpassam todas as tarefas do exame, quanto de ordem específica, ou seja, descritores que são úteis para a avaliação e descrição da proficiência do candidato na realização de algumas tarefas específicas do teste oral.

Os descritores de ordem genérica são compostos por capacidades com foco na linguagem e na comunicação, como o descritor 1, que compreende a capacidade de atingir os objetivos da comunicação verbal, por meio de fluência e do emprego de estruturas linguísticas; o descritor 2, representado pela pronúncia; e o descritor 6, representado pela habilidade de compreensão de fala. Os descritores de ordem específica, por sua vez, estão mais voltados a tarefas de ordem pedagógica, como o descritor 4, em que o candidato deve tecer considerações sobre uma cena de vídeo; e o descritor 5, que trata da capacidade de uso da metalinguagem. O descritor 3, sobre a capacidade do candidato de fornecer informações sobre experiências presentes e passadas, consiste em um descritor de ordem mista, pois perpassa as três dimensões consideradas pelo teste: comunicativa, linguística e pedagógica.

Embora seja possível notar nuances de enfoques distintos nos descritores das faixas de proficiência oral do exame, as habilidades contempladas na escala permitem a consideração das três dimensões da proficiência do professor de línguas de maneira integrada. A integração ocorre por meio da repetição de alguns critérios em descritores diferentes, como, por exemplo, no caso dos descritores 1 e 3, em que o elemento “estruturas linguísticas” aparece como um

critério de mensuração das habilidades de “atingir os objetivos da comunicação verbal” e de “fornecer informações”. A integração entre as dimensões propiciadas pela escala de proficiência oral do EPPLÉ é condizente com a concepção de proficiência do professor de línguas proposta em Consolo (2007), em que é descrita a existência de um domínio de intersecção entre uma competência de ordem geral e uma competência de ordem específica.

Apesar da integração, é possível notar uma orientação de ordem comunicativo-pedagógica, em detrimento do linguístico no teste oral, uma vez que há uma prevalência de habilidades comunicativas e pedagógicas na hierarquização dos domínios, ficando o domínio estrutural, com exceção do descritor 2 - em que a pronúncia é avaliada de maneira isolada - sempre dependente dos demais, ou melhor, como critério de avaliação das habilidades comunicativas e/ou pedagógicas. Tal organização remete os descritores da proficiência do teste oral do EPPLÉ à visão de linguagem proposta pela abordagem sobre competência linguístico-comunicativa (ALMEIDA FILHO, 1992), em que os elementos linguísticos têm sua importância preservada, mas não como um fim em si mesmos, e, sim, com a finalidade de garantir o estabelecimento da comunicação em sala de aula.

Contudo, deve-se ressaltar que a abordagem avaliativa do EPPLÉ orienta que o não alcance de apenas um descritor de faixa por parte do candidato seja suficiente para atribuir-lhe uma faixa imediatamente inferior, ou seja, se o candidato cumprir apenas cinco dos seis requisitos para ser enquadrado em uma determinada faixa de proficiência oral do EPPLÉ - independentemente de qual descritor não foi atingido -, o avaliador não pode lhe designar a referida faixa, mas, sim, a faixa imediatamente inferior. Tal abordagem avaliativa pode ser entendida como uma tentativa de atribuir pesos idênticos aos três domínios da proficiência do professor de línguas por parte do exame; contudo, há de se verificar se tal opção não implica falha de estruturação e, por extensão, não prejudica a representação do construto.

O Apêndice I permite perceber a valorização conferida aos elementos de ordem pedagógica em detrimento dos demais como elementos distintivos das duas faixas mais altas do teste oral (A e B). Conforme o nível de proficiência diminui, contudo, percebe-se que os elementos de ordem linguística vão se tornando características distintivas cada vez mais frequentes - em detrimento do desaparecimento de elementos de ordem pedagógica na descrição das habilidades -, o que permite perceber que o EPPLÉ considera, em consonância com o estudo realizado por Richards *et al.* (2013), que níveis maiores de proficiência linguística são necessários para o desenvolvimento de competências de ordem pedagógica na LE.

Assim, pode-se entender que a nomenclatura “linguístico-comunicativo-pedagógica” conferida à proficiência do professor de línguas é apresentada em gradação. Desse modo, o

elemento linguístico é empregado como auxiliar do elemento comunicativo, e o elemento linguístico-comunicativo, por sua vez, como base para o elemento pedagógico.

No entanto, a descrição de um grande número de faixas do EPPLÉ por meio de incompetências acaba por dificultar uma delimitação mais clara do conceito, pois aumenta o nível de imprecisão dos descritores em alguns níveis da escala. Isso porque, a avaliação por meio da negatividade, ou seja, por meio do que o aluno deixou de produzir permite a inferência de incompetência no desempenho de capacidades que as tarefas do teste, por exemplo, podem não ter elicitado. Além disso, é necessário que os resultados de um exame de proficiência sejam apresentados com clareza para que os usuários desses resultados possam fazê-lo com o máximo de “acuidade” (BACHMAN; PALMER, 1996), ou seja, é importante que uma descrição dos critérios, bem como dos medidores desses critérios estejam presentes nas faixas de proficiência do EPPLÉ.

4.2.2 Definição dos critérios de avaliação do desempenho oral

De acordo com as faixas de proficiência que compõem a escala do exame EPPLÉ, foi possível delimitar a existência de dois descritores relativos à clareza na expressão da fala sobre experiências, três descritores para avaliação da capacidade de uso da língua para a mesma finalidade, quatro descritores para avaliar o conhecimento do candidato sobre as regras da língua e outros quatro descritores para avaliar a capacidade de uso da metalinguagem com finalidade pedagógica, os quais se encontram resumidos no Quadro 14.

Quadro 14. Variáveis de uso e conhecimento sobre regras linguísticas da escala do exame EPPLÉ

Faixas	FINALIDADE GERAL		FINALIDADE ESPECÍFICA	
	Falar sobre experiências	Utilizar estruturas linguísticas	Demonstrar conhecimento sobre regras	Empregar metalinguagem
A	2- Fala com clareza 1- Fala sobre experiências	3- Uso de estruturas adequadas e específicas	4- Demonstra conhecimento	4- Explicação clara + uso de terminologia específica
B				3- Dificuldade na explicação
C		2- Uso de estruturas simples	3- Demonstra conhecimento parcial 2- Demonstra conhecimento limitado	2- Não explica de modo claro
D				
E		1- Uso de estruturas simples + Comete erros gramaticais	1- Demonstra pouco/nenhum conhecimento	1- Não consegue explicar

Fonte: Elaborado pela autora, com base na escala do EPPLÉ em Consolo e Teixeira da Silva (2014).

Desse modo, refletindo-se a respeito da relação existente entre os elementos que compõem esses descritores e a avaliação da capacidade de uso da língua para o ensino de

línguas com base nas informações trazidas pela escala de proficiência do EPPLE, obteve-se um cenário em que os critérios que compunham a avaliação da capacidade de uso de língua geral foram subdivididos em avaliação do atingimento da comunicação verbal – por meio do fornecimento de experiências, de narração e de descrição – e do uso de estruturas linguísticas, cuja composição foi especificada por meio de fluência, pronúncia e estruturas gramaticais e lexicais. Já a capacidade específica obteve quatro instâncias de avaliação: a expressão do conhecimento sobre regras linguísticas, a seleção de informações para esclarecer dúvidas linguísticas, o fornecimento de explicação sobre essas regras e o uso, propriamente dito, de terminologia específica. As informações encontram-se dispostas no Quadro 15.

Quadro 15. Hierarquização dos objetivos fundamentais segundo os princípios do VFT

Avaliar a capacidade de uso da língua para finalidades gerais e específicas	(objetivo estratégico)
1 Maximizar o entendimento da avaliação da capacidade de uso de língua geral	
1.1 Avaliar o atingimento da comunicação verbal	
1.2 Avaliar o uso de estruturas linguísticas	
1.2.1 Fluência	
1.2.2 Pronúncia	
1.2.3 Estruturas gramaticais	
1.2.4 Estruturas lexicais	
2 Maximizar o entendimento da avaliação da capacidade de uso de língua específica	
2.1 Avaliar a expressão de conhecimento sobre regras linguísticas	
2.2 Avaliar a seleção de informações para esclarecer dúvidas linguísticas	
2.3 Avaliar o fornecimento de explicação sobre regras linguísticas	
2.4 Avaliar o uso de terminologia específica	

Fonte: Elaborado pela autora, com base na escala do EPPLE em Consolo e Teixeira da Silva (2014).

Todos os objetivos avaliativos definidos para o exame consistem em atributos categóricos, ou seja, características de ordem qualitativa que podem ou não ser ordenadas para mensuração. No Quadro 16, encontram-se expressos os atributos empregados para mensuração de cada objetivo no exame, os quais foram, novamente, extraídos da escala de proficiência.

Quadro 16. Definição dos atributos/critérios de mensuração

OBJETIVOS	TIPO DE ATRIBUTO	ATRIBUTO
1 Maximizar o entendimento da avaliação da capacidade de uso de língua geral		
1.1 Avaliar o atingimento da comunicação verbal	EXP	Categoria
1.2 Avaliar o uso de estruturas linguísticas		
1.2.1 Fluência	FLU	Categoria
1.2.2 Pronúncia	PRO	Categoria
1.2.3 Estruturas gramaticais	GRA	Categoria
1.2.4 Estruturas lexicais	LEX	Categoria

2 Maximizar o entendimento da avaliação da capacidade de uso de língua específica			
2.1 Avaliar a expressão de conhecimento sobre regras linguísticas	CON	Categoria	Adequação
2.2 Avaliar a seleção de informações para esclarecer dúvidas linguísticas	SEL	Categoria	Clareza
2.3 Avaliar o fornecimento de explicação sobre regras linguísticas	REG	Categoria	Adequação Compreensibilidade ¹⁷
2.4 Avaliar o uso de terminologia específica	TER	Categoria	Adequação

Fonte: Elaborado pela autora, com base na escala do EPPLE em Consolo e Teixeira da Silva (2014).

Com base nos objetivos e atributos especificados, os limites superiores e inferiores de mensuração de cada critério foram definidos e encontram-se expressos nos Quadros 17, 18, 19, 20, 21, 22, 23, 24 e 25.

Quadro 17. Limite superior e inferior do atributo “clareza” para o objetivo “avaliar o fornecimento de informações sobre experiências”

CATEGORIA	DESCRIÇÃO
0 EXP	inexistente: candidato não responde à questão.
1 EXP	menos claro: candidato fornece informações sobre experiências pessoais.
2 EXP	mais claro: candidato fornece informações sobre experiências pessoais com clareza.

Fonte: Elaborado pela autora com base na escala do EPPLE em Consolo e Teixeira da Silva (2014).

Quadro 18. Limite superior e inferior dos atributos “velocidade” e “ritmo” para o objetivo “avaliar a fluência da fala”

CATEGORIA	DESCRIÇÃO
0 FLU	inexistente: candidato não responde à questão.
1 FLU	menos adequado: candidato exibe falta de fluência.
2 FLU	simples: candidato exibe fluência com limitações.
3 FLU	mais adequado: exibe fluência.

Fonte: Elaborado pela autora com base na escala do EPPLE em Consolo e Teixeira da Silva (2014).

Quadro 19. Limite superior e inferior dos atributos “compreensibilidade”, “inteligibilidade” e “proximidade” para o objetivo “avaliar a pronúncia”

CATEGORIA	DESCRIÇÃO
0 PRO	inexistente: candidato não responde à questão.
1 PRO	menos compreensível: pronúncia nitidamente distinta, podendo gerar interferências e desconfortos.
2 PRO	compreensível: pronúncia distinta, mas compreensível; podendo gerar interferências e desconfortos.
3 PRO	compreensível e distante de padrões: pronúncia distinta, mas compreensível.
4 PRO	compreensível e próxima a padrões: pronúncia próxima a padrões, não gera incompreensão.
5 PRO	compreensível e semelhante a padrões: pronúncia semelhante a padrões, não gera desconforto/incompreensão.

Fonte: Elaborado pela autora com base na escala do EPPLE em Consolo e Teixeira da Silva (2014).

Quadro 20. Limite superior e inferior dos atributos “acurácia” e “complexidade” para o objetivo “avaliar o uso de estruturas gramaticais”

CATEGORIA	DESCRIÇÃO
0 GRA	inexistente: candidato não responde à questão.
1 GRA	menos adequado: candidato utiliza estruturas simples e comete erros, com prejuízo para a expressão oral.

¹⁷ Compreensibilidade, neste estudo, refere-se ao julgamento do ouvinte acerca da dificuldade de se compreender uma produção oral e não engloba aspectos de inteligibilidade (MARTINS; FARIA, 2016).

2 GRA	simples: candidato utiliza estruturas adequadas, mas simples.
3 GRA	mais adequado: utiliza estruturas adequadas e mais complexas.

Fonte: Elaborado pela autora com base na escala do EPPL em Consolo e Teixeira da Silva (2014).

Quadro 21. Limite superior e inferior dos atributos “acurácia” e “variedade” para o objetivo “avaliar o uso de estruturas lexicais”

CATEGORIA	DESCRIÇÃO
0 LEX	inexistente: candidato não responde à questão.
1 LEX	menos específicas: candidato utiliza estruturas pouco variadas e comete erros, com prejuízo para expressão oral.
2 LEX	pouco variadas: candidato utiliza estruturas específicas, mas pouco variadas.
3 LEX	mais específicas: candidato utiliza estruturas específicas e bastante variadas.

Fonte: Elaborado pela autora com base na escala do EPPL em Consolo e Teixeira da Silva (2014).

Quadro 22. Limite superior e inferior do atributo “adequação” para o objetivo “avaliar a expressão de conhecimento sobre regras linguísticas”

CATEGORIA	DESCRIÇÃO
0 COM	inexistente: candidato não responde à questão ou não expressa conhecimento sobre regras linguísticas.
1 COM	pouca: candidato expressa pouco conhecimento sobre regras linguísticas.
2 COM	limitada: candidato expressa conhecimento limitado sobre regras linguísticas.
3 COM	parcial: candidato expressa conhecimento parcial sobre regras linguísticas.
4 COM	adequada: candidato expressa conhecimento adequado sobre regras linguísticas.

Fonte: Elaborado pela autora com base na escala do EPPL em Consolo e Teixeira da Silva (2014).

Quadro 23. Limite superior e inferior do atributo “clareza” para o objetivo “avaliar a seleção de informações para esclarecer dúvidas linguísticas”

CATEGORIA	DESCRIÇÃO
0 SEL	inexistente: candidato não responde à questão ou expressa incapacidade de selecionar informações.
1 SEL	com dificuldades: candidato demonstra dificuldade para selecionar informações adequadas.
2 SEL	adequada: candidato demonstra ser capaz de selecionar informações adequadas.

Fonte: Elaborado pela autora com base na escala do EPPL em Consolo e Teixeira da Silva (2014).

Quadro 24. Limite superior e inferior dos atributos “adequação” e “compreensibilidade” para o objetivo “avaliar o fornecimento de explicação sobre regras linguísticas”

CATEGORIA	DESCRIÇÃO
0 REG	inexistente: candidato não responde à questão ou expressa incapacidade de fornecer explicações adequadas e compreensíveis.
1 REG	com maior dificuldade: candidato demonstra dificuldade no fornecimento de explicações, que podem ser de difícil compreensão
2 REG	com menor dificuldade: candidato demonstra dificuldade no fornecimento de explicações, que podem ser apenas parcialmente compreendidas.
3 REG	com clareza: candidato fornece explicações de modo claro e plenamente compreensível.

Fonte: Elaborado pela autora com base na escala do EPPL em Consolo e Teixeira da Silva (2014).

Quadro 25. Limite superior e inferior do atributo “adequação” para o objetivo “avaliar o uso de terminologia específica”

CATEGORIA	DESCRIÇÃO
0 TER	inexistente: candidato não responde à questão ou não utiliza terminologia específica.
1 TER	limitado: candidato faz uso limitado de terminologia específica.
2 TER	adequado: candidato faz uso adequado de terminologia específica.

Fonte: Elaborado pela autora com base na escala do EPPL em Consolo e Teixeira da Silva (2014).

Ressalta-se, por fim, que a delimitação dos limites superiores e inferiores de mensuração, dado que a escala de proficiência do exame não foi elaborada por meio dos princípios do VFT e, portanto, encontrava-se pronta, nesta etapa foi realizada com base exclusivamente no documento da escala.

4.2.3 Avaliação da proficiência oral no EPPL: o desempenho dos candidatos nas tarefas de produção oral

Esta subseção trata da análise do desempenho de candidatos nas tarefas de produção oral no exame EPPL. O desempenho é analisado levando-se em consideração os níveis atribuídos pelos examinadores, a escala de proficiência oral do exame, a abordagem avaliativa empregada no exame e os critérios considerados relevantes para avaliação da proficiência do professor de línguas apontados pelos usuários do exame, os quais se encontram condensados no Quadro 13. A análise tem por objetivo determinar o caráter discriminatório dos critérios levando em consideração a avaliação dos examinadores e as possibilidades fornecidas pela tarefa.

A seção encontra-se dividida em três partes. Na primeira, investiga-se o desempenho dos candidatos em tarefas de avaliação do uso da língua para finalidades gerais; na segunda, em tarefas de uso da língua para finalidades pedagógicas; e, na terceira, investiga-se o desempenho de uso da competência linguístico-comunicativa. Em todas as partes, a análise de desempenho é realizada tarefa a tarefa, iniciando-se pela descrição e apresentação das mesmas e seguindo-se para as capacidades que elas evocam, pelos elementos passíveis de mensuração por meio da escala do exame e pelas características da linguagem produzida pelos candidatos.

A compreensão oral não foi considerada neste trabalho em virtude da ausência de dados para avaliação e comparação na Tarefa 3B da versão 3.1.

4.2.3.1 Avaliação do uso da linguagem com finalidade geral no exame EPPL

As tarefas que embasam a avaliação da linguagem com finalidade de uso geral no exame são apresentadas de modo a partir do fornecimento de informações de cunho bastante genérico – como informações pessoais e experiências e planos profissionais e educacionais – para informações mais voltadas a questões relativas ao ensino-aprendizagem de línguas – como a descrição de figuras que ilustram o ambiente da sala de aula ou a discussão de tópicos didáticos. O modo de organização das tarefas – do mais genérico para o mais específico – auxilia, assim,

na preparação do candidato para as tarefas de linguagem pedagógica que se apresentam na sequência.

O grau de autenticidade situacional dessas tarefas pode ser considerado alto, pois elas simulam situações típicas de contextos de apresentações pessoais e de formação, desenvolvimento ou atuação docente. Ademais, as tarefas não requerem níveis altos de interação com interlocutores ou de imprevisibilidade de retorno por parte desses interlocutores, uma vez que elas se focam em opiniões ou informações de cunho pessoal, preservando, assim, a autenticidade interacional das mesmas.

É importante ressaltar, ainda, a independência funcional e estatística das tarefas. Independência funcional porque o fornecimento de uma resposta incorreta em uma tarefa não acarreta, necessariamente, uma resposta errada em outra questão; e estatística porque o fornecimento de resposta correta em uma tarefa não ocasiona o levantamento de informações relevantes para outras – com exceção da Tarefa 3B. No teste oral do exame EPPLE, as tarefas que requerem o uso da linguagem para finalidades gerais variam do nível B1 ao B2 do CEFR, como se verá no decorrer das seções subsequentes. No entanto, de maneira geral, as instruções apresentadas pelas tarefas apresentam níveis linguísticos inferiores ao requerido para produção das respostas, o que acaba por minimizar possíveis interferências da capacidade de uso de habilidades receptivas do candidato em sua produção oral. O Quadro 26 condensa as principais características das tarefas que compõem o teste oral do exame EPPLE.

Quadro 26. Especificações das tarefas do teste oral do exame EPPLE (Versão 3.0 e 3.1)

INFORMAÇÕES DAS TAREFAS QUE COMPÕEM O TESTE ORAL DO EPPLE					
Tarefa	1	2	3	4	5
Versões	3.0, 3.1	3.1	3.0, 3.1	3.1	3.0, 3.1
CARACTERÍSTICAS GERAIS DAS TAREFAS					
Tópico da tarefa	Apresentação pessoal	Descrição de figuras	Discussão de tópico	Uso de linguagem pedagógica	Uso de metalinguagem
Língua das instruções	Língua-alvo do teste: Inglês não-nativo				
Duração da tarefa	Total:	5 minutos			
	Preparação:	1 minuto			
O desempenho é gravado?	Sim, apenas em áudio				
Controle da rubrica	Rigidamente controlada				
Controle do interlocutor	Parcialmente controlado				
Especificação de conteúdo	Apresentação pessoal Fornecimento de experiências Planos futuros	Descrição e comparação de figuras Especulação sobre figuras Relacionamento de figura a experiências	Descrição de conceito Fornecimento de opinião	Instruções para realização de tarefa	Versão 3.0: Emprego de Presente Simples e Presente Perfeito (Contínuo) Emprego de “will” e “going to” para expressão de futuro Versão 3.1: Emprego de condicional para expressar possibilidade futura e passada Emprego de modais (“might” e “must”) para expressar possibilidade
Canal das instruções	Falado (gravação em áudio) e texto escrito				
Nível linguístico das instruções	Abaixo do nível-alvo do teste				
Apresentação dos recursos	Falado (gravação em áudio) Texto escrito	Falado (gravação em áudio) Texto escrito Imagens	Falado (gravação em áudio) Texto escrito	Falado (gravação em áudio) Texto escrito	Falado (gravação em áudio) Texto escrito
Interação requerida	Monólogo				
Público real	<i>Eletrônico:</i> nenhum (computador) <i>Presencial:</i> examinador-interlocutor, examinador-avaliador e par entrevistado (candidato 2)				
Público imaginado	Interlocutor genérico	Interlocutor genérico	Colega de profissão	Alunos de LE (nível básico)	Alunos de LE (nível básico)
Cenário imaginado	Contexto social	Local de trabalho	Local de trabalho	Aula de língua estrangeira	Aula de língua estrangeira
Formato da resposta	Falado (gravação em áudio)				
Tipo de fala requerida	Referencial	Referencial	Referencial	Conativa	Metalinguística
Tópico CEFR	Identificação pessoal/Educação/Trabalho	Educação/Trabalho	Educação/Trabalho	Educação/Trabalho	Educação/Trabalho

RESPOSTAS ESPERADAS					
Tipo de resposta	Monólogo estendido (discurso neutro-formal)				
Tempo esperado de resposta	1 minuto (mínimo) - 2 minutos (máximo)				
Integração de habilidades	Leitura e/ou compreensão oral (não integram a avaliação)				
Propósito comunicativo	Referencial	Referencial	Referencial	Conativo	Metalinguístico
Funções retóricas esperadas	Apresentação Narração	Descrição (figuras) Comparação Narração Especulação	Descrição Fornecimento de opinião	Instrução	Explicação
Registro esperado	Neutro-Formal				
Domínio CEFR	Público	Educacional/Profissional	Educacional/Profissional	Profissional	Profissional
Nível de competência gramatical	Emprego majoritário de estruturas simples	Emprego variado de estruturas complexas	Emprego variado de estruturas complexas	Emprego majoritário de estruturas simples	Ampla variedade de estruturas gramaticais complexas
Nível de competência lexical	Emprego majoritário de vocabulário simples	Emprego variado de vocabulário extenso	Emprego variado de vocabulário extenso	Emprego majoritário de vocabulário simples	Ampla variedade de vocabulário avançado e especializado
Características discursivas	Uso competente	Uso competente	Uso competente	Uso competente	Uso avançado
Autenticidade situacional	Alta	Alta	Alta	Alta	Alta
Autenticidade interacional	Alta	Alta	Alta	Média	Média
Processamento cognitivo	Reprodução de ideias conhecidas	Reprodução de ideias conhecidas	Reprodução de ideias conhecidas	Reprodução de ideias conhecidas	Reprodução de ideias conhecidas
Conhecimento de conteúdo	Áreas de conhecimento especializadas	Áreas de conhecimento especializadas	Áreas de conhecimento especializadas	Áreas de conhecimento especializadas	Áreas de conhecimento especializadas
Nível esperado da resposta	B1	B2	B2	C1	C1

Fonte: Autor com base nas grades para tarefas de produção oral do CEFR (COE, 2009, 2014).

4.2.3.1.1 A tarefa de apresentação pessoal do teste oral do exame EPPLE

A tarefa de fornecimento de informações pessoais e de experiências encontra-se presente tanto na versão 3.0 (eletrônica e presencial) quanto na versão 3.1 (eletrônica) do exame investigadas neste trabalho. A tarefa requer respostas abertas, por meio das quais o candidato deve se apresentar, e fornece sugestões de tópicos a serem abordados, sendo eles: o nome do candidato; sua origem; há quanto tempo estuda inglês; se já trabalhou como professor de inglês e, em caso afirmativo, solicita o fornecimento de alguns detalhes sobre a experiência como professor da língua; e os planos futuros com relação à área de ensino e aprendizagem de inglês. A figura 14 ilustra a rubrica disponibilizada ao candidato na Tarefa 1 da versão 3.0.

Figura 14. Tarefa de apresentação pessoal da versão 3.0 do teste oral do exame EPPLE.

PART 1: Who are you ?

Tell us some information about yourself, for example,

- your name,
- where you are from,
- how long you have been learning English,
- if you have already worked as an English teacher (any details about your experience as a teacher?), and
- your future plans concerning the area of ELT (English Language Teaching).




Fonte: EPPLE (2017).

A figura 15 ilustra a rubrica disponibilizada ao candidato na Tarefa 1 da versão 3.1.

Figura 15. Tarefa de apresentação pessoal da versão 3.1 do teste oral do exame EPPLE.

TASK 1			
TASK TYPE	PROMPT MODE	PREPARATION TIME	PRODUCTION TIME
<i>Giving Personal Information</i>	<i>Written and Spoken</i>	<i>1 minute</i>	<i>2 minutes</i>
SCRIPT TO BE RECORDED			
DIRECTIONS	PROMPT		

 <p>In this part of the test you will speak about your personal, educational and professional experiences.</p>	 <ul style="list-style-type: none"> • Tell us some information about yourself. For example: your name, where you are from, how long you have been learning English, if you have any experience teaching English, your future plans concerning the area of ELT (English Language Teaching). • Please record your speech. Speak for about one to two minutes.
---	---

Fonte: EPPL (2015).

Por trabalhar com um tópico bastante genérico de uso da língua, a apresentação, a tarefa não requer níveis mínimos muito altos de proficiência linguística, especialmente porque as instruções fornecem uma lista de sugestões – e não tópicos obrigatórios – sobre os quais o candidato pode discorrer. Das sugestões apresentadas, a apresentação pessoal exige o uso da língua nos níveis A1-A2 do CEFR, segundo os quais, respectivamente, o candidato “é capaz de apresentar-se ou apresentar alguém e colocar questões ao seu interlocutor sobre assuntos como, por exemplo, o local onde vive, as suas relações, o que lhe pertence, etc.” (CEFR A1, tradução oficial) ou, ainda, “é capaz de compreender frases isoladas e expressões de uso frequente relacionadas com assuntos de prioridade imediata (por exemplo, informações pessoais e familiares simples, compras, meio envolvente, trabalho)” (CEFR A2, tradução oficial).

O fornecimento de experiências, por sua vez, relaciona-se com os níveis A2 e B1 do CEFR, segundo os quais o candidato “é capaz de descrever com meios simples a sua formação, o seu meio envolvente e referir assuntos que correspondam a necessidades imediatas” (CEFR A2, tradução oficial) ou “é capaz de relatar acontecimentos, experiências ou um sonho, expressar um desejo ou uma ambição e justificar, de forma breve, as razões de um projecto ou de uma ideia” (CEFR B1, tradução oficial).

O fornecimento de informação sobre planos futuros, por sua vez, alinha-se ao nível B1 do quadro, pelos mesmos motivos apresentados no excerto sobre o nível B1 anteriormente. Delimita-se, portanto, a seguinte estrutura esquemática da resposta mínima solicitada:

<p>[Introdução]</p> <p>Apresentação pessoal [nome, idade, profissão, ...]</p> <p>[Experiências]: [aprendizado de inglês, ensino de inglês, ...]</p> <p>[Planos]: [ensino de inglês, aprendizagem de inglês]</p> <p>[Conclusão]</p>
--

*Nota: Colchetes denotam movimentos opcionais

4.2.3.1.2 Avaliação do desempenho na apresentação pessoal no teste oral do exame EPPL

Atualmente, a apresentação pessoal é avaliada pelo descritor de número 3 da escala de proficiência oral do exame. Os critérios que integram essa avaliação combinam elementos pragmáticos e linguístico-comunicativos e resumem-se em (i.) fornecimento de informações sobre experiências presentes e passadas – o descritor desconsidera os planos futuros – e (ii.) precisão no emprego de estruturas linguísticas, com vistas a garantir a compreensibilidade da produção oral. O excerto 16 ilustra esses descritores.

Excerto 16. Descritores de apresentação pessoal da escala de proficiência oral do EPPL

- 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.
- 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.
- C3) Fornece informações sobre experiências presentes e passadas utilizando estruturas linguísticas simples e pouca variedade lexical.
- D3) Fornece informações sobre experiências presentes e passadas utilizando estruturas linguísticas simples, e pouca variedade lexical.
- 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.

Fonte: Consolo e Teixeira da Silva (2014).

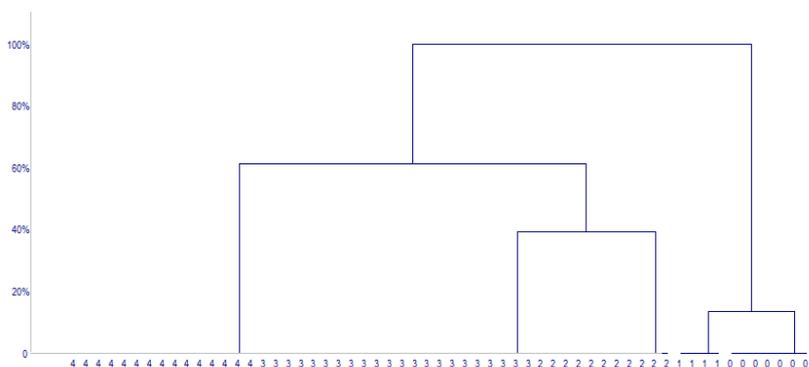
Tem-se, assim, que a Tarefa 1, no exame, é avaliada por meio de critérios que envolvem a avaliação de desenvolvimento temático, coerência e coesão, e precisão proposicional. Desse modo, nesta subseção, o desempenho na tarefa de apresentação pessoal será avaliado por meio do tipo de informação fornecida – pessoal, experiência educacional, experiência profissional ou plano futuro – e pela expansão e apresentação do tópico selecionado – apresentação em listas

De maneira geral, dos quatro tópicos analisados, informações pessoais, como nome, origem e profissão estiveram presentes em 100% (59) das ocorrências. Os tópicos de experiência educacional e de informações sobre planos futuros estiveram presentes em 94,91% (56) das ocorrências cada. O fornecimento de experiência profissional foi o que menos se fez presente com 89,83% (53) de presença.

Também não foram identificados fuga ao tema/tópico de discussão proposto ou problemas de organização textual ou de relevância das ideias. Os critérios de grau de construção original ou de reprodução de ideias, de sofisticação de ideias e de suficiência não foram empregados na análise por não serem possibilitados pela instrução da tarefa. No entanto, a análise revela que o emprego das funções comunicativas requeridas pela tarefa varia em relação à extensão da discussão e detalhamento do tópico, ou seja, em relação ao uso do tempo disponível para a completude da mesma.

A análise do agrupamento do emprego mediano de tópicos expandidos, ou seja, do fornecimento de informações com detalhe e precisão, utilizando-se de argumentos ou exemplificações, por outro lado, revelou a existência de três padrões de emprego, como se vê no Gráfico 3.

Gráfico 3. Dendrograma da quantidade de emprego de tópicos expandidos na tarefa de apresentação pessoal.



Nota: O eixo das abcissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

O primeiro grupo diferencia-se dos demais em número de candidatos e é composto pelo uso de quatro funções comunicativas expandidas. O segundo, varia o emprego entre três e duas funções, e o terceiro, emprega apenas 1 ou nenhuma função em detalhe. A análise mediana dos desempenhos em relação ao conceito atribuído revelou a existência de uma relação direta entre a extensão da fala e o nível de proficiência atribuído ao candidato, conforme se observa na Tabela 3.

Tabela 3. Resumo descritivo do emprego mediano de tópicos expandidos

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	3.5	0.5000	3.2500	3.7500
B	3.0	1.0000	2.0000	3.0000
C	3.0	2.0000	2.0000	4.0000
D	2.0	2.5000	0.5000	3.0000
E	2.5	0.5000	2.2500	2.7500

Fonte: Elaborado pela autora.

Assim, o tópico de fornecimento de informações de cunho pessoal mostrou-se menos propício para expansão, que apareceu em apenas 52,54% (31) das ocorrências, provavelmente por favorecer respostas mais diretas. O tópico de fornecimento de experiências educacionais foi expandido em 64,40% (38) das ocorrências, o de experiências profissionais em 66,10% (39) e o de planos futuros em 67,79% (40). Os excertos 17 e 18 ilustram alguns dos desempenhos analisados.

Excerto 17:

*RIC: <i>my name is ((nome do candidato)) .</i>	Informações pessoais
*RIC: <i>I (a)m from ((cidade do candidato)) .</i>	
*RIC: <i>I (ha)ve never (.) taught English before .</i>	Experiência profissional
*RIC: <i>&-um but I want to learn more and more about the structure of the language .</i>	Expansão do tópico
*RIC: <i>so I (a)m specially (.) interested in linguistics and phonology .</i>	
*RIC: <i>I do n(o)t have a (.) future plan to be a teacher .</i>	Planos futuros
*RIC: <i>but I want to develop my skills in English .</i>	Expansão do tópico

Candidato RIC. Faixa E, Tarefa 1, EPPLE, Versão 3.1, 2015.

Excerto 18:

*LAC: <i>so (.) my name is ((nome do candidato)) .</i>	Informações pessoais
*LAC: <i>&-um my first name is Greek and my last name is: misspelled Italian .</i>	Expansão do tópico
*LAC: <i>I am Brazilian .</i>	Informações pessoais
*LAC: <i>I (ha)ve been learning English since I was eleven years old .</i>	Experiência educacional
*LAC: <i>and I started teaching it when I was nineteen .</i>	Experiência profissional
*LAC: <i>&-um initially I was going to Law school .</i>	Expansão do tópico educacional
*LAC: <i>but then I decided to change my major into English .</i>	
*LAC: <i>because that (i)s what I like doing .</i>	
*LAC: <i>currently I (.) run my own business .</i>	Expansão do tópico profissional
*LAC: <i>and I also work at the university teaching English .</i>	

*LAC: <i>bu:t I would like to become a conference interpreter in the future.</i>	Planejamento futuro
*LAC: <i>so I (a)m also doing the course in order for that to be (.) &-um: possible in the future .</i>	Expansão do tópico de planejamento
*LAC: <i>I have lived abroad once .</i>	Expansão do tópico educacional
*LAC: <i>I lived in the United States .</i>	
*LAC: <i>an:d it was a very interesting experience .</i>	
*LAC: <i>because I think it (.) got me acquainted with a lot of (.) culture:</i>	
*LAC: <i>an:d it was very interesting for me .</i>	
*LAC: <i>and +/-.</i>	

Candidato LAC. Faixa A, Tarefa 1, EPPLÉ, Versão 3.1, 2015.

No excerto 17, além do fornecimento de informações diretas, com pouca expansão, nota-se o emprego de uma lista de sentenças, com emprego escasso de conectivos, os quais resumem-se a elementos linguísticos frequentes, como *and*, *but* e *so*. No excerto 18, por outro lado, nota-se o emprego de uma argumentação mais extensa e organizada em uma sequência linear de fatos.

De acordo com as escalas de desenvolvimento temático, coesão e coerência e precisão proposicional apresentadas pelo CEFR, nos níveis A1 e A2, os falantes apresentam restrições quanto à capacidade de variar tipos de informações e graus de detalhes fornecidos em suas descrições. Desse modo, nesses níveis, o tipo de informação que o candidato consegue discorrer a respeito restringe-se a informações de cunho pessoal e familiar e a uma quantidade pequena de detalhes. O texto é organizado por meio de palavras ou grupos de palavras conectados por meio de palavras simples ou organizados em sentenças simples apresentadas em lista. A fala do candidato RIC, apresentada no excerto 17, é representativa dessa capacidade. No nível B1, por sua vez, a variedade de tópicos de discussão é ampliada, bem como a capacidade de fornecimento de detalhes. Além disso, a organização do texto passa a comportar o emprego de conectivos para organizar as sentenças em sequências lineares. A fala do candidato LAC, apresentada no excerto 18, alinha-se a essa definição.

Percebe-se, assim, que a Tarefa 1 do exame permite a diferenciação do desempenho de fala dos candidatos em dois níveis do CEFR (A1/A2) e B1 – representados, respectivamente por desempenhos das faixas A, B e C, e por C, D e E – condizente com as medianas encontradas a respeito da capacidade de expansão do tópico da tarefa.

4.2.3.1.3 A tarefa de descrição de figuras do exame EPPLE

A tarefa de descrição de figuras encontra-se presente apenas na versão 3.1 (eletrônica) do exame EPPLE. A tarefa requer respostas abertas, nas quais o candidato deve descrever as figuras apresentadas, compará-las, especular a respeito do que elas representam e associá-las a experiências educacionais e profissionais. A figura 16 ilustra a rubrica disponibilizada ao candidato na Tarefa 2 da versão 3.0.

Figura 16. Tarefa descrição de figuras da versão 3.1 do teste oral do exame EPPLE.

TASK 2			
TASK TYPE	PROMPT MODE	PREPARATION TIME	PRODUCTION TIME
<i>Describing Pictures</i>	<i>Written and Spoken</i>	<i>1 minute</i>	<i>2 minutes</i>
SCRIPT TO BE RECORDED			
DIRECTIONS	PROMPT		
 <p>In this part of the test you will describe and compare two pictures. YOU WILL HAVE ONE MINUTE TO PREPARE YOUR SPEECH. YOU SHOULD SPEAK FOR ABOUT ONE TO TWO MINUTES.</p>	 <ul style="list-style-type: none"> • How would you describe the following pictures? What do they represent? How similar or different are they? Have you ever been in those situations before? • PLEASE RECORD YOUR SPEECH. SPEAK FOR ABOUT ONE TO TWO MINUTES. 		



Fonte: EPPLE (2015).

Tem-se, então, que a estrutura esquemática da resposta se configura da seguinte maneira:

<p>[Introdução]</p> <p>Descrição das figuras</p> <p>Opinião sobre a representação das figuras</p> <p>Comparação das figuras</p> <p>Associação a experiências</p> <p>[Conclusão]</p>

*Nota: Colchetes denotam movimentos opcionais

A tarefa de descrição do teste oral do EPPLE exige o fornecimento de respostas que variam do nível A2 ao B2 do CEFR. De acordo com a escala de produção oral geral, no nível A2, o candidato “é capaz de fornecer descrições ou apresentações simples de pessoas, condições de moradia e trabalho, rotinas, gostos, etc. em uma série de frases e sentenças curtas e simples elencadas em uma lista”¹⁸ (COE, 2018, p. 69). Em B1, o candidato “é capaz de sustentar uma descrição direta de um ou de uma variedade de assuntos de sua área de interesse com fluência razoável, apresentando-a em uma sequência linear de pontos”¹⁹ (COE, 2018, p. 69). No nível B2, por sua vez, o candidato “é capaz de fornecer descrições e apresentações claras e detalhadas sobre uma ampla gama de assuntos relacionados a sua área de interesse, expandindo e dando suporte a ideias com pontos subsidiários e exemplos relevantes”²⁰ (COE, 2018, p. 69).

¹⁸ *Can give a simple description or presentation of people, living or working conditions, daily routines, likes/dislikes etc. as a short series of simple phrases and sentences linked into a list.* (COE, 2018, p. 69)

¹⁹ *Can reasonably fluently sustain a straightforward description of one of a variety of subjects within his/her field of interest, presenting it as a linear sequence of points.* (COE, 2018, p. 69)

²⁰ *Can give clear, detailed descriptions and presentations on a wide range of subjects related to his/her field of interest, expanding and supporting ideas with subsidiary points and relevant examples.* (COE, 2018, p. 69)

O nível C1 não foi considerado na análise por tratar de “assuntos complexos”. Do mesmo modo, o nível A1 foi desconsiderado por restringir-se ao fornecimento de informações em frases simples e isoladas, o que não permitiria ao candidato completar a tarefa, especialmente com relação à realização de comparações.

4.2.3.1.4 Avaliação do desempenho na descrição de figuras no teste oral do exame EPPL

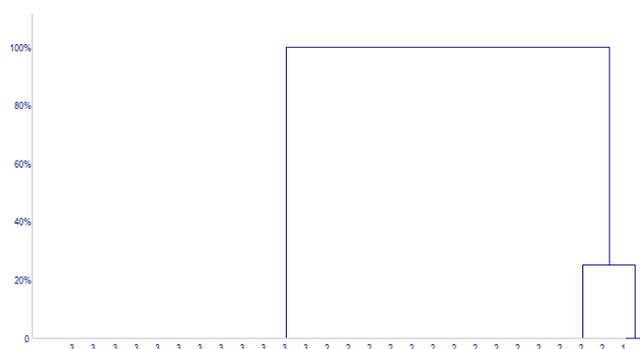
Atualmente, não há na escala de proficiência oral do exame EPPL um descritor específico para a tarefa de descrição de figuras. O descritor de número quatro, voltado à avaliação da descrição e narração de cenas de vídeo, é o que mais se aproxima do propósito comunicativo proposto pela tarefa. De acordo com o descritor, os critérios que integram essa avaliação são (i.) a qualidade da descrição, em termos de detalhamento, (ii.) a especulação a respeito da cena, (iii.) o estabelecimento de associações com experiências e (iv.) a sugestão de contribuições para o ensino e a aprendizagem de línguas. O excerto 19 ilustra estes descritores.

Excerto 19. Descritores de apresentação pessoal da escala de proficiência oral do EPPL

- 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.
- 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.
- 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.
- 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.
- 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.

Fonte: Consolo e Teixeira da Silva (2014).

Diante da não exigência de fornecimento de contribuições por parte das instruções da tarefa, este critério foi excluído da análise. Assim, a qualidade da descrição leva em consideração o grau de detalhamento fornecido e a presença ou ausência de comparação, de especulação e de associação a experiências passadas. No entanto, na análise de fala dos candidatos, não foi possível identificar um padrão distintivo com relação ao emprego de funções comunicativas, conforme exibido no Gráfico 4.

Gráfico 4. Dendrograma do emprego de funções comunicativas na tarefa de descrição de figuras.

Nota: O eixo das abscissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

A análise dos dados de fala dos candidatos corrobora o exibido pelo dendrograma, ou seja, não foi possível verificar a existência de uma relação direta forte entre o emprego mediano de funções comunicativas e os níveis de proficiência atribuídos, conforme se observa na Tabela 4.

Tabela 4. Resumo descritivo do emprego mediano de funções comunicativas para descrição

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	2.5	0.5000	2.2500	2.7000
B	3.0	1.0000	2.0000	3.0000
C	2.5	1.0000	2.0000	3.0000
D	2.0	0.0000	2.0000	2.0000
E	2.5	0.5000	2.2500	2.7000

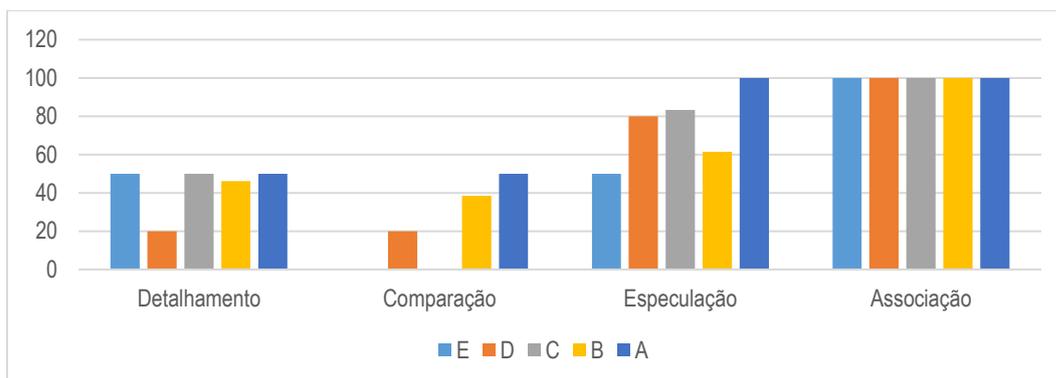
Fonte: Elaborado pela autora.

No entanto, dos recursos analisados, a associação das figuras com experiências esteve presente em 100% (28) das ocorrências. O segundo recurso mais frequente foi o da especulação, presente em 71,42% (20) das ocorrências. O detalhamento apareceu com uma frequência de 42,85% (12) e a comparação em apenas 25% (7) das ocorrências. Ressalta-se, contudo, que 58,33% (7) das ocorrências de detalhamento e das 85, 71% (6) das de comparação correspondiam à fala de candidatos nas faixas superiores A e B. Tal constatação indica que ambos os recursos, apesar do cunho de menor abstração se comparado aos outros dois, requerem uma maior competência de uso da língua por parte do falante, provavelmente, em virtude do caráter mais controlado de resposta que impõem. Isso porque eles exigem que o candidato

discorra a respeito da imagem que lhe acabou de ser apresentada, não abrindo espaço para tópicos de cunho mais familiar. A especulação e a associação com experiências, por outro lado, podem trazer mais conforto ao candidato com menor nível de proficiência por permitir que ele selecione informações mais previsíveis, para as quais ele tem maior domínio linguístico para discussão.

O Gráfico 5 exibe a frequência de emprego de função por nível de proficiência, considerando-se a escala de proficiência empregada pelo exame. Os resultados da análise informam que o emprego de comparação e de especulação apresentou uma relação direta com os níveis atribuídos, apontando para uma possível criterialidade. Ou seja, a análise do emprego de funções revelou que a extensão do tópico é, sim, capaz de diferenciar os níveis de proficiência. De maneira geral, a extensão dos tópicos de comparação e especulação esteve presente em níveis mais altos de proficiência ao contrário do foco predominante na extensão do tópico de associação.

Gráfico 5. Frequência do emprego de função comunicativa por nível de proficiência.



Fonte: Elaborado pela autora.

O grau de originalidade das ideias em relação ao grau de reprodutibilidade não foi considerado em virtude de ambas as possibilidades de discussão de tópicos serem inerentes às funções comunicativas elicitadas. Os excertos 20, 21 e 22 exibem alguns dos desempenhos investigados.

Excerto 20:

*LAC: <i>the first picture represents what all students would (.) like to do in class .</i>	Especação
*LAC: <i>they represent lack of discipline and lack of order an:d the freedom to wear whatever you want in class and to step on the tables and interact with your students in any way you want.</i>	Especação+Descrição
*LAC: <i>&-um sorry not with your students .</i>	
*LAC: <i>with your classmates in any way you want .</i>	
*LAC: <i>the second picture (.) shows us the perfect classroom (.) &-um as regards the teacher .</i>	
*LAC: <i>so (.) the second picture shows us everything as really very well organized .</i>	Descrição
*LAC: <i>an:d we can infer that the students are wearing a uniform .</i>	
*LAC: <i>and everyone is sat really quietly .</i>	
*LAC: <i>so it seems like this represents what a teacher would like a class to be like .</i>	Especação+Descrição
*LAC: <i>where [/] a class where all students are well behaved .</i>	
*LAC: <i>and they &ran raise their hands when they have a question or: (.) when they want to participate in class .</i>	
*LAC: <i>and you ha:ve (.) technology for you to use .</i>	
*LAC: <i>and so +/.</i>	
*LAC: <i>well I have been a teacher for almost ten years now .</i>	Associação a experiências
*LAC: <i>an:d I (ha)ve (.) studied for all my life .</i>	
*LAC: <i>so: (.) of course I have been to both situations in: both the position of a teacher or as a student .</i>	
*LAC: <i>&-um when I was a teenager I was not very well behaved .</i>	
*LAC: <i>and I was always the teenager who caused problems in class .</i>	
*LAC: <i>I did n(o)t wear uniform all the time .</i>	
*LAC: <i>and I really just made up mess in class .</i>	
*LAC: <i>so I think I identify myself a lot with the first picture (.) &-um when I think of my teenage years .</i>	
*LAC: <i>but (.) the second picture show:s (.) my everyday life <as a> [/] as a teacher .</i>	
*LAC: <i>I really like using technology in my classes .</i>	
*LAC: <i>an:d (.) I guess I can say that I (a)m pretty lucky .</i>	
*LAC: <i>because usually they are all very quiet .</i>	
*LAC: <i>and so I see tha:t (.) both pictures &repr +/.</i>	

Candidato LAC. Faixa A, Tarefa 2, EPPLÉ, Versão 3.1, 2015.

Excerto 21:

*LID: okay (.) so: (.) they represent a classroom .	Descrição
*LID: an:d (.) I see that both [/] (.) &-um both the classrooms have (.) students .	Comparação
*LID: but (.) one of them (.) has no teacher .	Associação a experiências
*LID: yes I have already been part of both (.) of (.) those situations .	
*LID: &-um the first one (.) when the teacher (.) is not in the classroom (.) the students are more comfortable .	Especulação
*LID: the second one when the teacher is in the classroom (.) we can (.) participate and (.) give our own opinions about the issue:s and (.) everything .	

Candidato LID. Faixa B, Tarefa 2, EPPLE, Versão 3.1, 2015.

Excerto 22:

*RIC: &-um the:y represent as I have &sa previously said (.) two: classes totally different one (.) from another .	Descrição
RIC: and as I have (.) said (.) also you do n(o)t can even (.) distinguish who is the teacher and who is the professor on [: in] [] picture A .	
RIC: otherwise [: on the other hand] [] in picture B (.) we can (.) clearly see the teacher using the blackboard that in this case (i)s a whiteboard .	
*RIC: an:d the: students &-um trying to answer a question that probably (.) the teacher (.) asked her [/] them .	Especulação
*RIC: oh yes (.) I: [/] I (ha)ve always attended classes (.) &-um with the same characteristics as picture B .	Associação a experiências
*RIC: I (ha)ve never been to a school (.) <or a> [///] (.) here (.) in the university &ha we have this situation of total (.) mess as we have in picture A .	
RIC: so I (.) have now (.) not experienced this [/] this kind of classroom .	
RIC: but in picture B (.) it (i)s possible to sees [: see] [] that if you have organization (.) if you have authority (.) your goal (.) will be (.) reachable .	Especulação
RIC: &-um: the students will learn (.) everything you taught [: teach] [] .	

Candidato RIC. Faixa E, Tarefa 2, EPPLE, Versão 3.1, 2015.

Os excertos exemplificam diferentes desempenhos correspondentes aos níveis B2, B1 e A2, respectivamente. No excerto 20, é possível verificar a presença de desenvolvimento temático por meio de exemplificações e detalhamento, além de apresentação de sentenças em uma sequência linear. A fala do candidato inclui as funções comunicativas de descrição, especulação e associação. No excerto 21, o candidato, apesar da pouca exposição oral, consegue sustentar a descrição de maneira direta e apresentar seus argumentos em uma lista. No excerto

22, por sua vez, o candidato RIC utiliza-se das mesmas funções empregadas por LAC, no entanto, não consegue providenciar detalhes relevantes para a descrição das figuras. A fala do candidato RIC gira em torno das funções de especulação e de associação a experiências. Além disso, o candidato utiliza-se de recursos coesivos simples e incorre de erro no emprego de um recurso mais complexo, como *otherwise*.

4.2.3.1.5 A tarefa de discussão de tópico do exame EPPLE

A tarefa de discussão de tópico encontra-se presente na versão 3.0 (presencial e eletrônica) e na versão 3.1 (eletrônica) do exame EPPLE. A versão 3.0 requer que o candidato assista a dois vídeos curtos, um que discute a relevância do sotaque na comunicação em língua estrangeira e outro que ilustra uma aula tradicional de inglês, em que uma professora tem por objetivo ajustar a pronúncia do aluno. A tarefa requer que o candidato identifique o assunto abordado, especule sobre o público-alvo do vídeo e opine relacionando o tópico abordado ao ensino e a aprendizagem de línguas.

A figura 17 ilustra a rubrica disponibilizada ao candidato na Tarefa 2 da versão 3.0.

Figura 17. Tarefa de discussão de tópico da versão 3.0 do teste oral do exame EPPLE.

PART 2: What do you think about ...?

You will watch a short video extract found in the YouTube website.
After watching the video extract once or twice, you will give us your opinions about the video, based on the following questions:

- 1) What is the video extract about?
- 2) Who do you think Mr. Duncan is speaking to?
- 3) What important piece(s) of information do you believe the video extract provides about pronunciation, for English language students and teachers?

Speak and interact with the other candidate for about 2:00 – 3:00 minutes.




Fonte: EPPLE (2015).

Na versão 3.1, por sua vez, a tarefa de discussão de tópico sucede a Tarefa 3A do exame. Na Tarefa 3A, o candidato deve assistir a um vídeo curto sobre a relevância da motivação na aprendizagem de línguas e responder a três questões de múltipla-escolha que avaliam sua capacidade de compreensão oral. A Tarefa 3B, por sua vez, utiliza o tópico discutido no vídeo

como insumo para a discussão. Desse modo, requer que o candidato descreva o conceito de motivação e opine sobre o grau de afetamento da motivação no aprendizado, de maneira geral, e, especificamente, no aprendizado de uma língua estrangeira, além do papel de professores e alunos na construção de motivação. De todas as tarefas que compõem o teste oral do exame, a Tarefa 3B é a única que apresenta dependência estatística. A figura 18 ilustra a rubrica disponibilizada ao candidato na Tarefa 3B da versão 3.1.

Figura 18. Tarefa de discussão de tópico da versão 3.1 do teste oral do exame EPPLE.

TASK 3B			
TASK TYPE	PROMPT MODE	PREPARATION TIME	PRODUCTION TIME
<i>Video Viewing & Production</i>	<i>Written and Spoken</i>	<i>1 minute</i>	<i>2 minutes</i>
SCRIPT TO BE RECORDED			
DIRECTIONS		PROMPT	
 <p>Based on information from the video extract and your own ideas, complete task 3B. YOU WILL HAVE ONE MINUTE TO PREPARE YOUR SPEECH. YOU SHOULD SPEAK FOR ABOUT ONE TO TWO MINUTES.</p>		 <ul style="list-style-type: none"> • How would you describe motivation? • Does motivation affect learning? • What about learning a foreign language? • What role do teachers play in motivating students? • What role do students play in motivating teachers? • PLEASE RECORD YOUR SPEECH. SPEAK FOR ABOUT ONE TO TWO MINUTES. 	

Fonte: EPPLE (2015).

A tarefa de discussão de tópico varia do nível B1 ao B2 do CEFR. De acordo com a escala de interação oral geral, no nível B1, o candidato

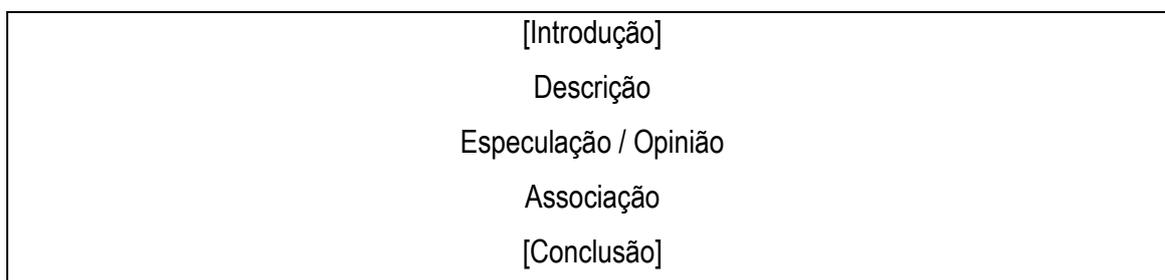
é capaz de entrar despreparado em uma conversação sobre tópicos familiares, **expressar opiniões pessoais** e trocar informações sobre tópicos que são familiares, de interesse pessoal ou pertinentes ao dia-a-dia (ex. família, hobbies, trabalho, viagem e eventos atuais)²¹ (COE, 2018, p. 135).

O quadro de características salientes de interação e produção do CEFR ainda estabelece que, para o nível B1, o candidato consegue “explicar por que algo é um problema, **resumir e opinar** sobre uma história, um artigo, uma palestra, uma entrevista de discussão ou documentário e responder questões adicionais sobre detalhes²²” (COE, 2018, p. 123).

No nível B2, por sua vez, o candidato

é capaz de destacar a significância pessoal de eventos e experiências, **explicar e sustentar opiniões** claramente, providenciando explicações e argumentos relevantes. É capaz de se comunicar com um pouco de confiança sobre **assuntos** familiares rotineiros e não rotineiros **relacionados a seus interesses e campo profissional**²³ (COE, 2018, p. 135).

A tarefa possibilita a seguinte estrutura esquemática:



*Nota: colchetes denotam movimentos opcionais.

4.2.3.1.6 Avaliação do desempenho na discussão de tópico no teste oral do exame EPPLE

O descritor de número quatro da escala de proficiência oral do exame EPPLE é voltado à avaliação da descrição e narração de cenas de vídeo e, portanto, é o que mais se aproxima do propósito comunicativo da tarefa. Os critérios que integram essa avaliação são (i.) a qualidade da descrição, em termos de detalhamento, (ii.) a especulação a respeito da cena, (iii.) o estabelecimento de associações com experiências e (iv.) a sugestão de contribuições para o

²¹ *Can enter unprepared into conversation of familiar topics, express personal opinions and exchange information on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).* (COE, 2018, p. 135)

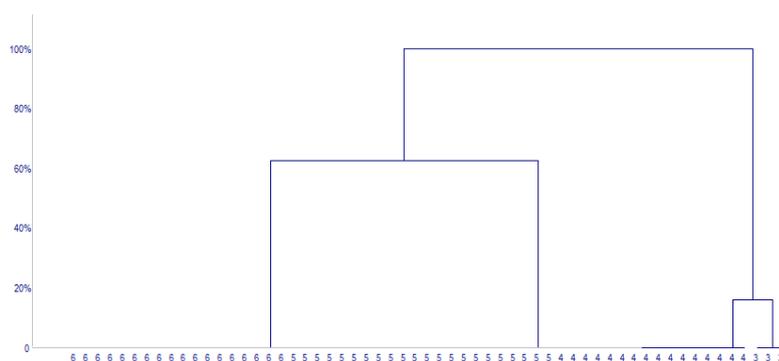
²² *[...] explain why something is a problem; summarise and give his or her opinion about a short story, article, talk, discussion interview, or documentary and answer further questions of detail [...].* (COE 2018, p. 123)

²³ *Can highlight the personal significance of events and experiences, account for and sustain views clearly by providing relevant explanations and arguments. Can communicate with some confidence on familiar routine and non-routine matters related to his/her interests and professional field.* (COE, 2018, p. 135)

ensino e a aprendizagem de línguas. No entanto, diante das exigências da tarefa, apenas os dois primeiros critérios mostram-se úteis para a avaliação.

Desse modo, a análise leva em consideração a capacidade de especulação a respeito do assunto do vídeo ou do conceito e a qualidade das opiniões fornecidas, mensuradas com base na presença ou não de explicações e de argumentos de sustentação. A análise de conglomerados revelou a existência de dois grupos, um no qual o emprego de funções variava de 6 a 5 e outro no qual o emprego variava de 4 a 3, conforme exibido pelo Gráfico 6.

Gráfico 6. Dendrograma do emprego de funções comunicativas na tarefa de discussão de tópico.



Nota: O eixo das abscissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

Contudo, conforme exibido pela Tabela 5, não houve uma relação direta entre o emprego mediano das funções comunicativas e a diferenciação entre os níveis de proficiência atribuídos.

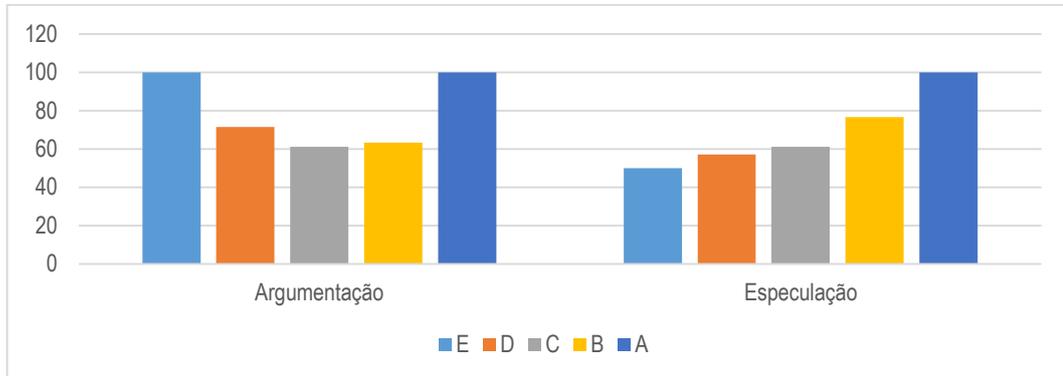
Tabela 5. Resumo descritivo do emprego mediano de funções comunicativas para opinião

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	5.5	0.5000	5.2500	5.7500
B	5.0	1.7500	4.2500	6.0000
C	5.0	2.0000	4.0000	6.0000
D	5.0	1.0000	4.0000	5.0000
E	4.0	1.0000	3.5000	4.5000

Fonte: Elaborado pela autora.

No entanto, a capacidade de extensão do tópico mostrou-se relevante, especialmente no que concerne à capacidade de fornecer razões para as especulações realizadas sobre a temática, conforme exibido no Gráfico 7.

Gráfico 7. Frequência de extensão de tópico na tarefa de discussão.



Fonte: Elaborado pela autora.

Os critérios coerência e coesão, e flexibilidade mostraram maior poder distintivo. Os excertos 23 e 24 exibem algumas das falas dos candidatos que, mesmo classificados na faixa D, em termos gerais de proficiência, apresentaram um desempenho satisfatório na tarefa em ambas as versões do exame.

Excerto 23:

*THY: &-um I believe that motivation is: something you feel (.) in [/] in [/] (.) <with you> [/] with yourself .	Descrição
*THY: and that (.) makes you (.) do things by yourself .	
*THY: you feel motivated (.) to do things .	
*THY: that is motivation for me .	
*THY: &-um of course because motivation is important when you (a)re learning .	Especulação
*THY: you can n(o)t learn things (.) without being motivated .	Argumentação
*THY: becau:se &-um when you are motivated you feel strong (.) to do things .	
*THY: you feel (.) more comfortable to do things .	
*THY: you believe in yourself .	
*THY: so you do things better .	
*THY: &-um teachers have an: [/] an important (.) role in motivating students .	Especulação
*THY: because they are their model .	Argumentação
*THY: they are the [/] the: (.) person: (.) who: have the: [/] (.) the knowledge (.) to: +...	
*THY: they have the power to teach .	
*THY: so motivation is something that they should work (.) in relation to students .	Resumo
*THY: because they are the model in class .	
*THY: the person the students are going to pay attention 0to and follow	

Candidato THY. Faixa D, Tarefa 3B, EPPLÉ, Versão 3.1, 2015.

Excerto 24:

MAC: <i>the first video talks [: is] [] about (.) &-um: what you should sound like when you are talking in English .</i>	Identificação do assunto
*MAC: <i>and &-um: (.) it says (.) that &any everyone should sound like themselves .</i>	Argumentação
*MAC: <i>they should n(o)t (.) &-um sound like somebody else just because they (a)re speaking (.) some other language .</i>	
*MAC: <i>and &-um: (3) <it> [/] it (i)s really about accent .</i>	
*MAC: <i><he (i)s talking about> [/] (.) he (i)s talking to (.) &-um: people that were born in (.) non English speaking countries .</i>	Especulação
*MAC: <i><the second> [/] (.) the second video is about how difficult it is (.) for people whose first language is not English (2) &-um: + to learn + English .</i>	Identificação do assunto
MAC: <i>the videos (.) &-um: (1) <are> [/] &-um make me think that (.) &giv give [: given] [] the information (.) that (.) it is very hard for someone (.) &-um: that was n(o)t born in a [: an] [*] &sp English speaking country (.) to learn English .</i>	Opinião
*MAC: <i>and &-um: (.) they also say (.) that &-um <as long as (.) you:> [/] as long as people understand you &-um: (.) your accent does n(o)t matter .</i>	Resumo
*MAC: <i>&th the way you sound does n(o)t matter as long <as> [/] (.) as people can understand (.) &yo &-um: what you (a)re talking (.) in English .</i>	

Candidato MAC. Faixa D, Tarefa 2, EPPLE, Versão 3.0, 2015.

Nos excertos, é possível perceber que os candidatos conseguem descrever, especular e argumentar a respeito do assunto, mobilizando as funções comunicativas de desenvolvimento temático, precisão proposicional que compõem a competência pragmática no nível B2. Em termos de flexibilidade, tanto THY quanto MAC conseguem reformular suas expressões para enfatizar ou esclarecer seus pontos de vista. As frases são ordenadas em uma sequência linear.

4.2.3.2 Avaliação do uso da linguagem com finalidade pedagógica no exame EPPLE

As tarefas que embasam a avaliação da linguagem pedagógica no exame requerem o uso da linguagem do professor para fornecimento de explicações sobre recursos linguísticos a determinados alunos e/ou de instruções para realização de atividades típicas de sala de aula a esses mesmos alunos. O grau de autenticidade situacional dessas tarefas é, portanto, alto, pois leva o candidato a executar uma função típica de professores em uma variedade de abordagens de ensino de língua estrangeira. No entanto, ambas as tarefas requerem níveis igualmente altos de interação com interlocutores – no caso, os alunos imaginários – e de imprevisibilidade de retorno por parte desses mesmos alunos, níveis estes que ficam especialmente prejudicados na versão eletrônica do exame. Desse modo, tem-se que a própria situação de testagem impõe um declínio à autenticidade interacional dessas tarefas, o qual deve ser considerado no estabelecimento de graus de exigência para a elaboração de escalas de proficiência.

É importante ressaltar, como se verá adiante, a independência funcional e estatística das tarefas. Independência funcional porque o fornecimento de uma resposta incorreta em uma tarefa não acarreta, necessariamente, uma resposta errada em outra questão; e estatística porque o fornecimento de resposta correta em uma tarefa não ocasiona o levantamento de informações relevantes para outras.

No teste oral do exame EPPLE, as tarefas que requerem o uso da linguagem pedagógica variam do nível B1 ao C1 ou superior do CEFR, como se verá no decorrer desta seção. No entanto, de maneira geral, as instruções apresentadas pelas tarefas apresentam níveis linguísticos inferiores ao requerido para produção das respostas, fazendo uso restrito de linguagem terminológica básica, o que acaba por minimizar possíveis interferências da capacidade de uso de habilidades receptivas do candidato em sua produção oral.

A tarefa de metalinguagem encontra-se presente tanto na versão 3.0 (presencial e eletrônica) do exame quanto na versão 3.1 (eletrônica); no entanto, a tarefa de instruções insere-se apenas na versão mais atual do exame, a 3.1 (eletrônica).

4.2.3.2.1 A tarefa de fornecimento de instruções no teste oral do exame EPPLE

A tarefa de instrução que compõe o teste oral do exame EPPLE consiste em tarefa de resposta aberta, para a qual o candidato deve colocar-se na situação de um professor de língua estrangeira de alunos de nível básico e instruí-los a elaborar uma história pessoal empregando figuras e utilizando-se do Passado Simples – tópico gramatical da aula. A figura 19 ilustra a rubrica disponibilizada ao candidato.

Figura 19. Rubrica da Tarefa 4 do teste oral do exame EPPLE (Versão 3.1).

TASK 4			
TASK TYPE	PROMPT MODE	PREPARATION TIME	PRODUCTION TIME
<i>Giving Instructions</i>	<i>Written and Spoken</i>	<i>1 minute</i>	<i>2 minutes</i>
SCRIPT TO BE RECORDED			
DIRECTIONS		PROMPT	

 <p>In this part of the test you will pretend to give students instructions for an activity to be carried out in the classroom. YOU WILL HAVE ONE MINUTE TO PREPARE YOUR SPEECH. YOU SHOULD SPEAK FOR ABOUT ONE TO TWO MINUTES.</p>	 <ul style="list-style-type: none"> • Suppose you have taught the Simple Past to a group of elementary level students and have used a set of pictures to talk about a person's past day. Now pretend you instruct your students to create their own stories using pictures. Do not forget to set up a time limit for them to conclude the activity. • PLEASE RECORD YOUR SPEECH. SPEAK FOR ABOUT ONE TO TWO MINUTES.
--	--

Fonte: Extraída de EPPL (2015).

A tarefa de fornecimento de instruções, por trabalhar com uma atividade cujo tópico é típico de aulas de nível B1 – mais especificamente, o uso de Passado Simples para descrever estados e ações habituais de maneira ordenada, no formato de narrativas – de acordo com o *English Grammar Profile*, alinha o uso de competências mobilizados aos apresentados no critério “capacidade linguística” do CETF (CAE, 2018). De acordo com o quadro, no nível *developing*, compatível ao B1 do CEFR, o professor “fornece exemplos precisos de tópicos linguísticos ensinados nos níveis A1, A2 e B1 do CEFR” (CAE, 2018, p. 08) e “emprega linguagem de sala de aula majoritariamente precisa” (CAE, 2018, p. 08). No nível *proficient*, por sua vez, compatível ao B2, o professor “fornece exemplos precisos de tópicos linguísticos ensinados nos níveis A1, A2, B1 e B2 do CEFR”²⁴ (CAE, 2018, p. 08). Além disso, o docente “usa linguagem de sala de aula consistentemente precisa ao longo de toda a aula” (CAE, 2018, p. 08). No nível *expert*, compatível ao C1 ou superior, por fim, o professor “fornece exemplos precisos de tópicos linguísticos ensinados do nível A1 ao C2 do CEFR” (CAE, 2018, p. 08) e, ainda, “usa uma variedade de linguagem de sala de aula consistentemente precisa ao longo de toda a aula” (CAE, 2018, p. 08).

A estrutura esquemática da tarefa consiste na apresentada a seguir:

²⁴ No original: “Provides accurate examples of the language points being taught at A1, A2 B1 and B2 levels of the CEFR.” (CAE, 2018, p. 8).

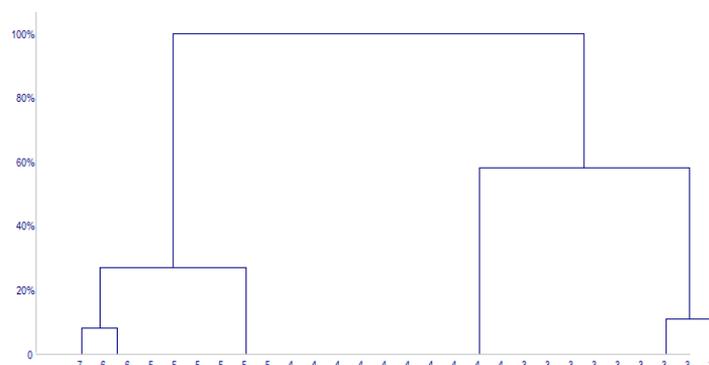
[Introdução]	[Gerenciamento de grupos]
Diretiva Principal	[Início da atividade]
Anúncio da Atividade	Municiamento linguístico
[Modelagem da Atividade]	[Digressão]
[Checagem de compreensão]	[Informações redundantes]
Estabelecimento de tempo	[Conclusão]

*Nota: colchetes denotam movimentos opcionais.

4.2.3.2.2 Avaliação do desempenho no fornecimento de instruções no teste oral do exame EPPLE

Atualmente, não há, na escala de proficiência oral do exame, descritores específicos voltados à avaliação da linguagem pedagógica que não sejam de caráter exclusivamente metalinguístico. Deste modo, para a avaliação das características da fala instrucional dos candidatos, levou-se em consideração a qualidade da organização do gênero textual instrutivo analisando-se as funções comunicativas propostas em Andreilino (2014), que divide o texto em elementos nucleares e elementos elaboradores. Os elementos nucleares, essenciais para a configuração do gênero, subdividem-se em elemento crucial, representado pelo fornecimento da diretiva principal; e elementos associativos, representados pelo anúncio da atividade, pela modelagem da atividade, pela checagem de compreensão, pelo estabelecimento de tempo, pelo gerenciamento de grupos e pelo início da atividade. Os elementos elaboradores, por fim, são representados pelo municiamento linguístico, pela presença de digressão e pelo fornecimento de informações redundantes.

Na análise dos dados de fala, verificou-se a presença de quatro grupos majoritários: o primeiro formado pelo emprego de 7 a 6 funções; o segundo, pelo emprego de 5; o terceiro pelo emprego de 4; e o quarto, pelo emprego de 3 ou 2 funções, conforme se vê no Gráfico 8.

Gráfico 8. Dendrograma do emprego de funções no fornecimento de instruções.

Nota: O eixo das abscissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

Na análise dos dados de fala dos candidatos, verificou-se que o emprego mediano de recursos de organização textual mantinha uma relação direta com o nível de proficiência atribuído ao desempenho de fala do candidato, conforme se observa na Tabela 6.

Tabela 6. Resumo descritivo do emprego mediano de organização do texto instrutivo

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	6.0	1.0000	5.5000	6.5000
B	3.0	3.0000	2.0000	5.0000
C	2.0	0.7500	2.0000	2.7500
D	3.0	0.0000	3.0000	3.0000
E	1.0	1.0000	0.5000	1.5000

Fonte: Elaborado pela autora.

De maneira geral, os recursos instrutivos mais empregados na tarefa pertenciam à classe dos elementos nucleares, representados pelo fornecimento de diretiva principal, presente em 96,43% (27) das ocorrências – apenas o candidato RIT, classificado na faixa D, não se utilizou da função; a elicitación de tempo, solicitada pela própria rubrica da tarefa, presente em 78,57% (22) das ocorrências; e o anúncio da atividade, presente em 75% (21) das ocorrências. Observou-se, ainda, que os candidatos classificados na faixa A empregaram recursos de organização textual mais diversificados. Na fala destes candidatos, além dos três recursos mais comuns, estiveram presentes elementos elaboradores, mais especificamente, digressão e municiação linguística, além das funções associativas de gerenciamento de grupo e

modelagem de atividade. No entanto, o emprego de digressões e de modelagens não foi exclusivo dos candidatos da faixa A, perpassando desempenhos nas faixas B, C e D. Os recursos de municiamento linguístico e gerenciamento de grupo, por sua vez, exibiram comportamentos de maior sofisticação em virtude de suas recorrências nas faixas superiores (A e B), com apenas uma ocorrência de municiamento na faixa C (candidato TAS) e uma de gerenciamento de na faixa D (candidato WAN). Os excertos 25 e 26 ilustram alguns desses desempenhos.

Excerto 25:

*TAS: *okay guys now (.) that we have learned the: simple past .*
 *TAS: *and I showed you the &exam some examples .*
 *TAS: *&-um: now it (i)s your turn to do it .*
 *TAS: *&-um you (a)re going to create (.) your own stories .*
 *TAS: *&-um you [/] you: (a)re going to talk about yourselves (.) using the simple past and using pictures .*
 *TAS: *so: (.) you are going to use for example .*
 *TAS: *I:[/] I ate breakfast last night [: morning] [*] .*
 *TAS: *and then you have a picture of (.) breakfast .*
 *TAS: *okay ?*
 *TAS: *an:d <you &ar> [/] I will give you: (.) twenty minutes (.) to: do this activity .*
 *TAS: *okay ?*
 *TAS: *any questions ?*
 *TAS: *you can ask me .*

Candidato TAS. Faixa C, Tarefa 4, EPPLE, Versão 3.1, 2015.

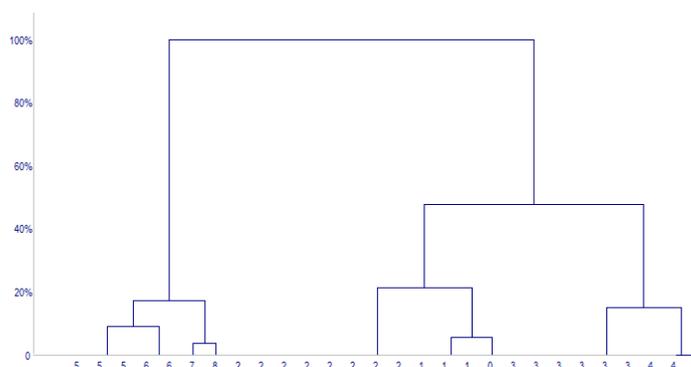
Excerto 26:

*TAC: *&st &-um <students you have> [//] &-um good morning everyone &-um please &-um now you are going to have about <five minutes to: look over> [//] the five to ten minutes to look over the pictures that you have on your desk .*
 *TAC: *and &-um you must create &yo &-um your own story about how &-um was:*
 *TAC: *take a look at those activities .*
 *TAC: *<pay attention> [//] focus on the activity .*
 *TAC: *and pretend (.) that it (i)s your doing those activities .*
 *TAC: *okay ?*
 *TAC: *so I want you to describe those activities for me .*
 *TAC: *imagine yourself doing (.) each one of those (.) actions .*
 *TAC: *and based on that please just create a whole story .*
 *TAC: *okay ?*
 *TAC: *<you are the main &st> [//] you are the main character of your own story .*
 *TAC: *you are the hero (.) just now &-um +/-*

Candidato TAC. Faixa D, Tarefa 4, EPPLE, Versão 3.1, 2015.

Com relação ao uso de terminologia, a análise apontou para a existência de três grandes padrões: o primeiro, variando de 8 a 5; o segundo, de 3 a 4; e o terceiro, de 2 ou menos, conforme exibido no Gráfico 9.

Gráfico 9. Dendrograma do emprego de recurso terminológico na tarefa de instrução.



Nota: O eixo das abscissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

Também se observou a ocorrência de uma relação direta entre o emprego mediano deste recurso linguístico e o nível de proficiência atribuído aos candidatos, conforme ilustrado na Tabela 7.

Tabela 7. Resumo descritivo do emprego mediano de variedade terminológica no texto instrutivo

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	6.0	1.0000	5.5000	6.5000
B	4.0	1.0000	4.0000	5.0000
C	4.0	0.0000	4.0000	4.0000
D	3.0	0.0000	3.0000	3.0000
E	2.5	0.5000	2.2500	2.7500

Fonte: Elaborado pela autora.

Por outro lado, a contagem de desvios no emprego de recursos terminológicos não se mostrou significativa para auxiliar na distinção dos níveis de proficiência dos candidatos, diante da inexistência de erros nos dados de fala alvos da análise. Observou-se, contudo, a presença de alguns vícios de linguagem pedagógica presentes tanto na fala de candidatos de faixas superiores, como a B, quanto de candidatos de faixas inferiores, como a D. São eles o emprego de *remember* por *recall* para referir-se a elementos linguísticos trabalhados em aulas anteriores,

de *doubt* por *question* para referir-se a dúvidas sobre as instruções, e de *small story* por *short story* em referência à extensão do texto solicitado pela atividade. O quadro 27 ilustra os recursos metalinguísticos mais empregados pelos candidatos do exame com base na lista de palavras que integra o glossário do TKT (CAE, 2015).

Quadro 27. Recursos terminológicos empregados no texto instrutivo

Fornecimento de Instruções (V. 3.1)		
Verbos	Expressões	Recursos linguísticos
Explain	Explanation	(Simple) Past
Revise	Question / Doubt	Present
Choose	Activity	(Ir)regular verb
Teach	Correct form	Subject
Demonstrate	Task	Sentences
Remember	Groups	Complement
Learn	Class	Auxiliary
Practice	(For) Example	Pronunciation
Check	Text / Paragraph	Base form
Discuss	Turn	Modals
Illustrate	Short story / Small story	Temporal Markers
Represent	-	-
Ask	-	-
Pretend	-	-

Fonte: Elaborado pela autora.

Em resumo, tem-se que o recurso de precisão terminológica indicado pelos usuários do exame como ferramenta necessária para a avaliação da linguagem pedagógica instrutiva não se mostrou útil para o corpus analisado neste trabalho. A análise revelou, contudo, que, apesar da inexistência de descritores específicos para o texto instrutivo na escala de proficiência oral do exame, há coerência entre a avaliação do desempenho dos candidatos e a quantidade de recursos de organização textual empregados de acordo com a proposta de Andreilino (2014), bem como entre a avaliação e a variedade terminológica utilizada.

4.2.3.2.3 A tarefa de explicação gramatical do EPPLE

As tarefas de explicação gramatical que compõem o teste oral do exame EPPLE nas duas versões analisadas neste trabalho consistem em tarefas de resposta aberta, para as quais o candidato deve selecionar uma, dentre duas opções de situações de dúvida linguística de alunos, e fornecer explicações. As tarefas simulam um domínio-alvo de uso da língua comum a aulas de inglês como língua estrangeira.

A tarefa da versão 3.0 do exame, empregada na aplicação eletrônica de 2015 e na presencial de 2017, ambas para a universidade pública paulista, tinham como situações de dúvida linguística as diferenças no emprego de Presente Simples e Presente Perfeito/Presente Perfeito Contínuo, e no emprego de “will” e “be going to” para expressão de futuro. A figura 20 ilustra a rubrica fornecida ao candidato.

Figura 20. Tarefa de metalinguagem da versão 3.0.1 do teste oral do exame EPPLÉ.



**PART 3: Some doubts I have as an English student...
Can you help me?**

In this part two situations will be presented. Each situation presents a doubt I have when I speak English.
Read the two situations and (a) decide about which you would like to help me (only one of the two situations). When you are ready, (b) speak with me and give me some explanation in order to solve my doubt.

Here are my doubts...

SITUATION 1
*Hi! My teacher always corrects me when I say:
“I live in São Paulo all my life”.
What’s the problem?*

SITUATION 2
*Hi! I’m an intermediate student and I would like to know
how to use GOING TO and WILL when I speak about
the future... I’m not sure there’s a difference ...*

Please give me your explanation. Speak for about 1:00 – 2:00 minutes.

©DAConsolo 

Fonte: Adaptado de Consolo (2017).

A tarefa da versão 3.1 do exame, por sua vez, empregada na aplicação eletrônica de 2015 para a universidade pública mineira, solicitava que o candidato discorresse a respeito do

uso da primeira e da terceira condicional ou sobre o uso de modais para expressão de possibilidade. A figura 21 ilustra a rubrica fornecida ao candidato.

Figura 21. Tarefa de metalinguagem da versão 3.1 do teste oral do exame EPPL.

TASK 5			
TASK TYPE	PROMPT MODE	PREPARATION TIME	PRODUCTION TIME
<i>Using Metalanguage</i>	<i>Written and Spoken</i>	<i>1 minute</i>	<i>2 minutes</i>
SCRIPT TO BE RECORDED			
DIRECTIONS		PROMPT	
 <p>In this part of the test you will use your metalinguistic knowledge to solve a language problem faced by a student. YOU WILL HAVE ONE MINUTE TO PREPARE YOUR SPEECH. YOU SHOULD SPEAK FOR ABOUT ONE TO TWO MINUTES.</p>		 <ul style="list-style-type: none"> You will be showed two different situations by students of English. Each situation presents a difficulty the student has when he or she speaks English. Read and listen to the two situations and decide about which you would like to help the student. You have to choose just one of the situations. When you are ready, speak with the student and give them some explanation. PLEASE RECORD YOUR SPEECH. SPEAK FOR ABOUT ONE TO TWO MINUTES. 	

SITUATION A

I don't know if these sentences are correct. The first is If I happen to run into her, I would tell her where you live. The second is If antibiotic drugs like peniciline were available during World War I, many more lives could be saved. Could you help me?

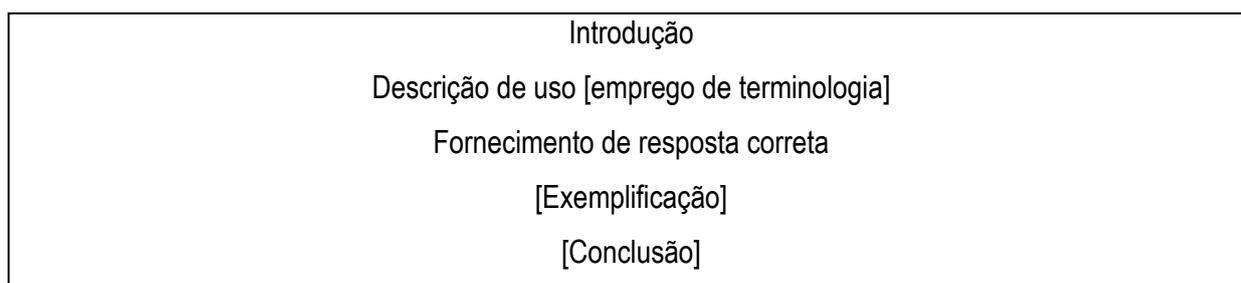
SITUATION B

These sentences are really confusing in this exercise. Which alternative best completes the sentence: I've written to Maria but she hasn't replied?
Letter A: She might be away. You never know with her. Letter B: She must be in Salvador. You never know. How can you explain that?

Fonte: EPPLE (2015).

A tarefa de metalinguagem consiste naquela com maior nível de exigência do teste oral do EPPLE, pois ela oferece ao candidato a possibilidade de identificar e discutir apenas uma situação de problema linguístico, não abrindo espaço para erros na identificação do problema. De acordo com o CETF (2018), o reconhecimento de todos os erros cometidos pelos alunos consiste em uma capacidade compatível com ao menos o nível C1 do CEFR, o que faz com que o desempenho mínimo esperado pela tarefa também se relacione ao nível C1.

A estrutura esquemática da tarefa encontra-se representada a seguir:



*Nota: colchetes denotam movimentos opcionais.

4.2.3.2.4 Avaliação do desempenho na explicação gramatical pela escala de proficiência oral do exame EPPLE

Atualmente, o modelo de avaliação da explicação gramatical no EPPLE é realizado por meio do descritor de número cinco da escala de proficiência oral. As instruções para realização da tarefa solicitam que o candidato interaja com o aluno imaginário, no caso do formato eletrônico, ou com o examinador-interlocutor na condição de aluno imaginário, no caso da aplicação presencial, e forneça explicações. Desse modo, o descritor relativo ao uso dessa competência na escala do exame espera que o candidato: expresse conhecimento sobre regras de uso da língua, selecione informações relevantes, explique as regras de modo claro e inteligível e empregue terminologia específica. Em resumo, o descritor prevê o fornecimento de explicações acuradas e o emprego de terminologia específica, como exibido pelo excerto 27.

Excerto 27. Descritores de metalinguagem da escala de proficiência oral do EPPLE

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.

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.

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.

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.

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.

Fonte: Consolo e Teixeira da Silva (2014).

Para análise da precisão da explicação, optou-se pela utilização da estrutura da explicação gramatical identificada por Fernandes (2016), composta por introdução, descrição de uso, exemplificação e fornecimento de resposta correta e conclusão. Nos dados de fala analisados, as funções mais empregadas foram a de introdução, descrição do uso e fornecimento de resposta correta. Os recursos de exemplificação ou de conclusão da fala foram utilizados poucas vezes. A exemplificação, mais especificamente, esteve maiormente presente em situações que não forneciam exemplos de sentenças para serem explicados, como a situação 2 da Tarefa 3 (Versão 3.0) sobre o emprego de *will* e *going to* para expressão de futuro, ao passo que o fornecimento de resposta correta, em oposição à exemplificação, se fez mais presente nas demais situações. Os excertos 28 e 29 exemplificam alguns desempenhos.

Excerto 28:

BRU: I: cho:se [: choose] [] (.) the first situation .

*BRU: and I would say to my student that (.) in English when we want to say that something happened in the past (.) and still happens in the present we use the Present Continuous instead of just the Present Tense .

*BRU: so: he student (.) has: to use "subject" plus "has" or "have" plus

“the main verb in (.) past participle” instead of just the Present Tense as: &he we use <to:> [/] to speak in Portuguese .

**BRU: so <we:> [//] <the correct> [/] the correct way to say that would be “I have lived in São_Paulo all my life” .*

Candidato BRU. Faixa C, Tarefa 5, EPPLE, Versão 3.0, 2015.

Excerto 29:

**LUC: &-um: going to is used (.) when you are talking about a situation in the future that you know (.) is going to happen .*

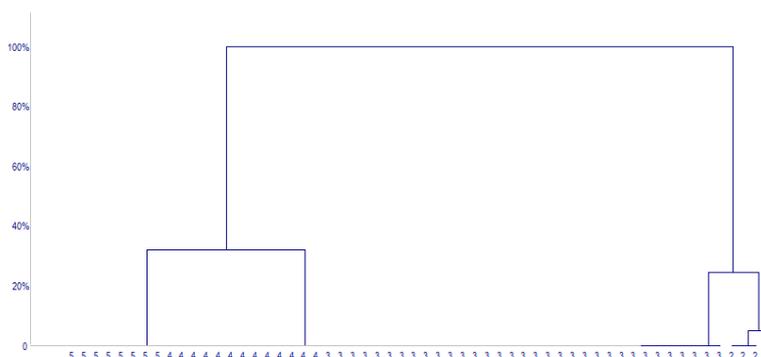
**LUC: and will is used for predictions .*

**LUC: so <you> [/] <you> [/] you use will (.) you are talking about a situation in the future that you (.) are not sure (.) that might happen .*

Candidato LUC. Faixa B, Tarefa 5, EPPLE, Versão 3.0, 2015.

A análise dos dados de fala revelou a existência de dois grupos de emprego de recursos metalinguísticos: o primeiro, com emprego de cinco a quatro funções; e o segundo, com emprego de três ou menos, conforme exibido no Gráfico 10.

Gráfico 10. Dendrograma do emprego de recursos metalinguísticos



Nota: O eixo das abcissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

A Tabela 8 apresenta o emprego mediano de recursos para explicação metalinguística nos dados analisados, apontando para a possibilidade de diferenciação entre níveis.

Tabela 8. Resumo descritivo do emprego mediano de explicação metalinguística com acurácia

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	5.0	0.0000	5.0000	5.0000
B	3.5	1.0000	3.0000	4.0000
C	3.0	0.0000	3.0000	3.0000

D	3.0	0.0000	3.0000	3.0000
E	3.5	0.5000	3.2500	3.7500

Fonte: Elaborado pela autora.

Na faixa A, ambos os candidatos analisados cumpriram as cinco funções comunicativas em suas falas metalinguísticas. Na faixa B, por sua vez, o emprego de funções comunicativas variou de três a cinco funções. Nas faixas C e D restringiu-se a três e, na faixa E, um candidato utilizou-se de quatro e o outro de três funções. Apesar de os dados indicarem um caminho para diferenciação de níveis de proficiência, o emprego de funções comunicativas de explicação metalinguística por si não garante a precisão da explicação fornecida, como exibem os excertos 30 e 31.

Excerto 30:

HEL: I chose [: choose] [] the second (.) situation .
 *HEL: &-um: (.) &u <"will" talk &-um: wi(th)> [//] &u use "will" when you (.) are (.) talking about a prediction .
 *HEL: we (a)re not certain <of> [//] about the future .
 *HEL: I don't know if <the> [//] this thing <will> [//] will happen (.) in a certain way .
 *HEL: and you use (.) "going to" <when you have> [//] (.) &-um: (.) when <the> [//] <the> [//] (.) the thing 0that will happen is closer .
 *HEL: and you <have the certainty (.) more> [//] &-um you are <more &c> [//] more certain that 0it will happen .

Candidato HEL. Faixa C, Tarefa 5, EPPLÉ, Versão 3.0, 2015.

Excerto 31:

LID: so: (.) I chose [: choose] [] situation B .
 *LID: &-um what do you have to do here ?
 *LID: read letter A and then letter B .
 *LID: "she might be away (.) you never know with her" .
 *LID: in this situation you are using the wrong preposition .
 LID: so "with" and "her" is [: are] [] not in the context .
 *LID: you never know (.) about her .
 *LID: you would never know about her .
 *LID: you should use (.) another expression (.) to: (.) mean what you (.) want to say .
 *LID: letter B is better .
 *LID: "she must be in Salvador (.) you never know" .
 *LID: "you never know" because you do not know where (.) this person is where Mary is .

*LID: so: (.) in letter A (.) “you never know with her” (.)
the preposition (.) is being used wrongly .

*LID: while in (.) letter B (.) “you never know” (.) is better .

*LID: because you do not have a preposition .

*LID: and you do not have (.) an undirect [: indirect]
[*] object in your way .

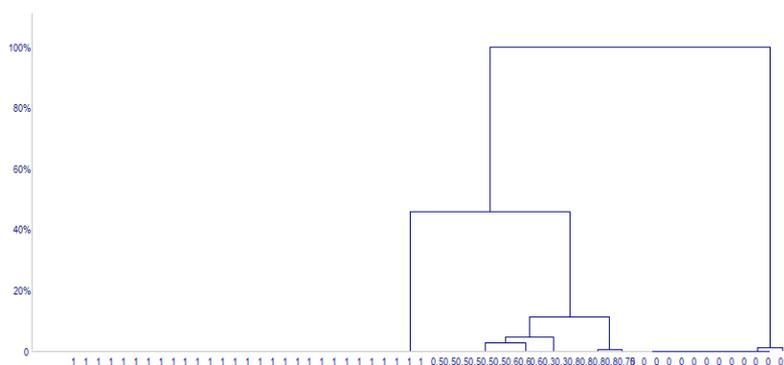
Candidato LID. Faixa B, Tarefa 5, EPPLE, Versão 3.1, 2015.

No excerto 30, a explicação fornecida pelo candidato é precisa; no entanto, seu desempenho no teste oral foi classificado como indicativo da faixa C. Em 31, por outro lado, apesar de a explicação estar incorreta, o candidato foi classificado na faixa B.

A mensuração da acurácia, neste trabalho, dá-se, portanto, a partir do número de explicações a respeito do tópico solicitado fornecidas pelo candidato com acurácia, ou seja, considera, como sugerem Farrell e Richards (2007), o conhecimento de tópico a respeito do assunto gramatical abordado. Para tanto, quantifica-se os acertos nas funções de identificação do problema, descrição do uso, na exemplificação – quando presente e relevante para a tarefa – e/ou no fornecimento da resposta correta. Optou-se por desconsiderar a conclusão graças a sua pouca relevância para a situação do teste em si.

A análise dos dados revelou a existência de quatro grupos: o primeiro, em que as explicações são 100% precisas; o segundo, em que a acurácia varia de 80 a 70%; o terceiro, com 60 a 50% de precisão; e o quarto, com 30% de precisão ou menos. O Gráfico 11 ilustra os resultados da análise acurácia da explicação.

Gráfico 11. Dendrograma da porcentagem de acurácia da explicação metalinguística.



Nota: O eixo das abcissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

A análise dos dados não apontou, contudo, diferenças medianas no emprego acurado de explicação metalinguística e nas classificações obtidas pelos candidatos no exame para as faixas A, B e C, conforme exibido na Tabela 9.

Tabela 9. Resumo descritivo do emprego mediano de explicação metalinguística com acurácia

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	5.0	0.0000	5.0000	5.0000
B	5.0	3.0000	2.0000	5.0000
C	5.0	3.0000	2.0000	5.0000
D	3.0	2.0000	1.0000	3.0000
E	0.0	0.0000	0.0000	0.0000

Fonte: Elaborado pela autora.

Tem-se, com isso, que a avaliação da metalinguagem dos candidatos não levou adequadamente em consideração a acurácia da explicação por eles fornecida em relação aos descritores da escala ou foi influenciada por outros critérios. De maneira geral, dos 59 dados analisados na tarefa metalinguística, 71,18% (42) apresentaram inconsistências na avaliação. Destes, 52,54% (31) das ocorrências apresentaram características de fala compatíveis com ao menos um nível superior na escala e 18,64% (11) tinham desempenhos compatíveis com ao menos um nível inferior. A consistência da avaliação ficou demonstrada em apenas 28,81% (17) das ocorrências.

Também não foram encontradas diferenças significativas com relação à acurácia terminológica, uma vez que, dos 177 vocábulos terminológicos identificados, houve inconsistência no emprego de apenas cinco, divididos entre as faixas B, C e D. Desse modo, tem-se que 97,17% (172) dos vocábulos empregados eram acurados. A Tabela 10 ilustra os resultados da análise da acurácia terminológica.

Tabela 10. Resumo descritivo do emprego mediano de explicação metalinguística com acurácia

	Candidatos	Terminologia	Erros	% Acertos
A	2	8	0	100
B	30	99	2	97,97
C	17	41	2	95,12
D	8	22	1	95,45
E	2	7	0	100
Total	59	177	5	97,17

Fonte: Elaborado pela autora.

Ainda com relação ao emprego de terminologia, a análise dos dados revelou que algumas das situações de dúvidas linguísticas propostas pelas tarefas eram mais propícias ao emprego de recursos terminológicos do que outras. A situação 1 da Tarefa 3 da versão 3.0 e as situações A e B da Tarefa 5 da versão 3.1 foram aquelas que permitiram um emprego mais diversificado e sofisticado de linguagem terminológica, conforme exibido nos quadros 28, 29, 30 e 31. A situação 2 da Tarefa 3 da versão 3.0, contudo, limitou o emprego por permitir respostas mais objetivas.

Quadro 28. Terminologia empregada na Situação 1 da Tarefa 3 do teste oral do EPPL

Uso de metalinguagem: Tarefa 3 Situação 1 (v. 3.0)		
Verbos	Nomes/Expressões	Recursos linguísticos
Express	(In)Correct(ly)	Grammar (rules)
Emphasize	For example / like	Auxiliary
Mean	Appropriate	Verb (tense)
-	Idea / concept	Progressive action
-	Meaning	Present (tense/time)
-	-	(Simple) Present
-	-	Past (Tense)
-	-	Perfective aspect
-	-	Present Perfect (Continuous)
-	-	Subject
-	-	(Main) Verb
-	-	Past participle / Third form

Fonte: Elaborado pela autora.

Quadro 29. Terminologia empregada na Situação 2 da Tarefa 3 do teste oral do EPPL

Uso de metalinguagem: Tarefa 3 Situação 2 (v. 3.0)		
Verbos	Nomes/Expressões	Recursos linguísticos
Explain	(Grammatical) Difference	Future (time)
Express	For example	Sentence
-	-	Present

Fonte: Elaborado pela autora.

Quadro 30. Terminologia empregada na Situação A da Tarefa 5 do teste oral do EPPL

Uso de metalinguagem: Tarefa 5 Situação A (V. 3.1)		
Verbos	Nomes/Expressões	Recursos linguísticos
Indicate	(Not) correct / Error	Object
Complete	Difference	Conditional sentences

-	Like	If-clauses
-	-	Present (form)
-	-	Future (tense/time/auxiliary)
-	-	Past
-	-	Rules
-	-	Verb (form)
-	-	Sentence
-	-	Modals
-	-	First conditional
-	-	Preposition
-	-	Redundancy
-	-	Pleonasm

Fonte: Elaborado pela autora.

Quadro 31. Terminologia empregada na Situação B da Tarefa 5 do teste oral do EPPLE

Uso de metalinguagem: Tarefa 5 Situação B (V. 3.1)		
Verbos	Nomes/Expressões	Recursos linguísticos
Complete	Appropriate	Modals
-	For example	Structure (of the language)
-	Hypothesis	Caluse
-	Difference	Sentence
-	Context	Verb tenses
-	(In)Correct(ly) / Wrong	Indirect object
-	Alternative	Expression
-	Translation	Future
-	-	(Simple) Present
-	-	Conditionals
-	-	(Simple) Past (time)
-	-	Present Perfect

Fonte: Elaborado pela autora.

Com relação à lista, é válido, destacar, ainda, a presença de alguns vícios de linguagem pedagógica, como o emprego de indeterminado de *time* por *tense*, bem como alguns neologismos, como *future verb* por *future auxiliary*, *perfect aspect* por *perfective aspect* e *redundance* por *redundancy*. Tanto os vícios quanto os neologismos estiveram presentes na fala de candidatos das faixas B, C e E.

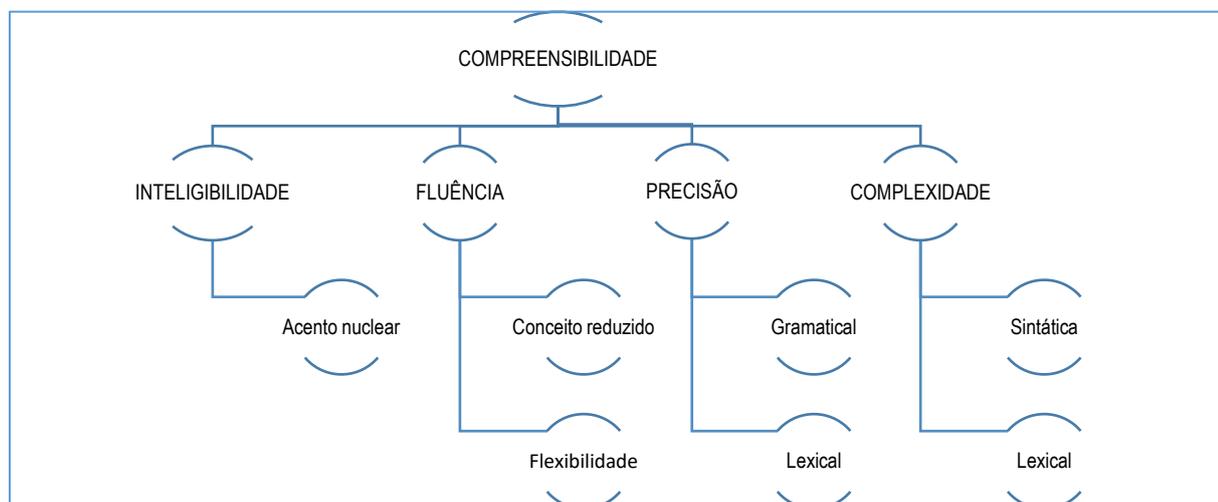
Percebe-se, portanto, que os descritores atuais da escala de proficiência do EPPLE associados à abordagem avaliativa do exame não representam com fidelidade os dados de fala da linguagem metalinguística dos candidatos, uma vez que, em sua maioria, não há concordância entre as afirmações tecidas pelo descritor da faixa e a qualidade real do desempenho dos candidatos.

A proximidade da mediana do emprego de explicações acuradas e a porcentagem de emprego de terminologia acurada explicam a baixa concordância da avaliação do desempenho dos candidatos nesta tarefa, ressaltando a necessidade de elementos distintivos menos fluidos. A tendência demonstrada pela frequência e os valores do primeiro e terceiro quartil, contudo, podem ser úteis nessa delimitação. Além disso, a diferença na quantidade e no grau de sofisticação da terminologia elicitada pelas tarefas impõe uma dificuldade de mensuração dessa competência de uso e, portanto, a necessidade de se modalizar o discurso da escala. Além disso, diante do alto grau de concordância exibido pelos quadros de referência e pelos usuários do exame acerca da importância da terminologia para a fala pedagógica, mostra-se necessária a inclusão de orientações que explicitem a necessidade de emprego desse recurso nas instruções das tarefas de metalinguagem do exame.

4.2.3.3 Avaliação das competências linguístico-comunicativas no exame EPPL

A avaliação das competências de ordem linguístico-comunicativa no EPPL orientou-se pela análise da compreensibilidade dos dados de fala por entender esta dimensão da fala como crucial para o atendimento dos objetivos comunicacionais da fala. Deste modo, a compreensibilidade, neste estudo, é entendida como o grau de dificuldade ou facilidade de compreensão da fala na perspectiva do interlocutor (THOMSON, 2018) e, portanto, esta pesquisa entende que a compreensibilidade consiste no aspecto crucial da avaliação de proficiência, sofrendo influências de outros elementos de ordem linguística, como os expostos na Figura 22.

Figura 22. Relação entre os componentes da competência linguístico-comunicativa avaliados.

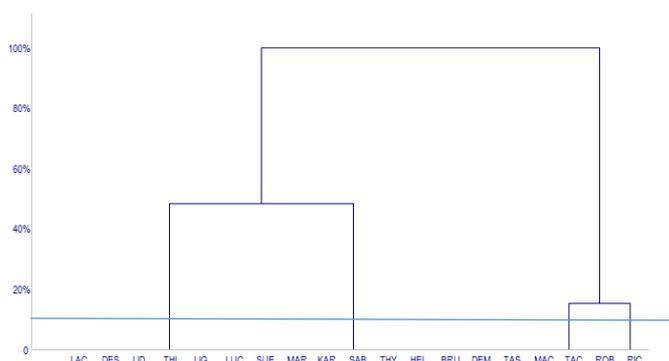


Fonte: Elaborado pela autora.

A inteligibilidade é entendida neste estudo como a efetividade da mensagem produzida pelo falante, levando-se em consideração suas intenções comunicativas (THOMSON, 2018) e, como tal, encontra-se intimamente ligada a elementos de ordem fonético-fonológica, ou seja, a componentes da pronúncia. Diante das necessidades apontadas pelos usuários do exame de se considerar padrões de inglês como língua franca na avaliação da proficiência do professor de línguas, delimitou-se para a investigação dos padrões de pronúncia, os componentes fonológicos essenciais que compõem o *Lingua Franca Core* proposto por Jenkins (2000). De acordo com o quadro, a (i.) divergência fonêmica, caracterizada, em termos vocálicos, pela qualidade de produção do fonema /ɜ:/ e pela consistência na preservação de contrastes e ditongos e, em termos consonantais, pela qualidade da preservação de fonemas na posição inicial das palavras e de sons de articulação distantes, juntamente com (ii.) o acento nuclear são características que devem ser preservadas. No entanto, diante da inexistência de consenso com relação à proposta de Jenkins (2000), Walker (2010) ressalta a importância do acento nuclear em especial para a efetividade comunicacional.

A esse respeito, o estudo de Manosso (2013) identificou a existência de quatro perfis de posicionamento de acento nuclear em falantes de inglês como língua estrangeira brasileiros. O primeiro grupo, com nível de proficiência C1, era capaz de posicionar o acento nos verbos e nos substantivos nucleares da sentença; no grupo de proficiência B2, os participantes foram capazes de posicionar o acento nos verbos nucleares, mas também em final de sentença; em B1, a posição ocorreu apenas no início e no final da sentença; nos níveis A2 e A1, o posicionamento de acento ocorreu majoritariamente no final da sentença.

O desempenho dos candidatos com relação à inteligibilidade neste estudo levou em consideração o acento nuclear em enunciados nos quais os participantes forneciam informações sobre suas experiências com o aprendizado de língua inglesa – presente em 96,55% dos dados de fala. Os dados foram analisados por meio do espectrograma de fala produzido pelo *software* Praat e resultou nos quatro padrões de análise identificados por Manosso (2013), conforme exibido no Gráfico 12.

Gráfico 12. Dendrograma do desempenho de fala com relação ao acento nuclear.

Nota: O eixo das abscissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

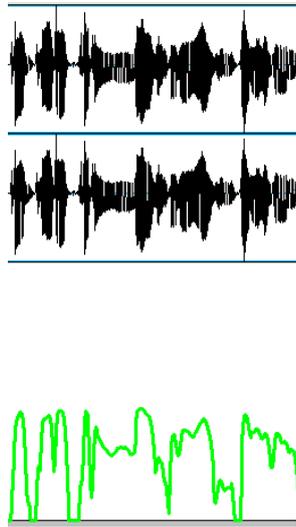
Com relação à avaliação do desempenho no exame EPPLE, evidencia-se a existência de uma relação entre os grupos encontrados na análise e as faixas atribuídas aos candidatos, conforme evidenciado pela Tabela 11.

Tabela 11. Frequência do emprego de acento nuclear.

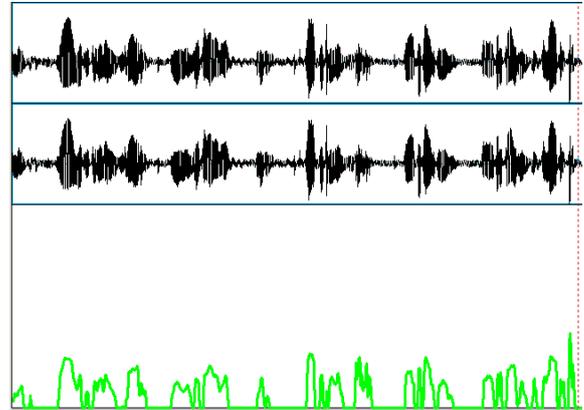
	Verbos e substantivos	Verbos e posição final	Posição final e inicial	Posição final
A	2	0	0	0
B	3	27	0	0
C	1	15	1	0
D	1	0	2	5
E	0	0	1	1

Fonte: Elaborado pela autora.

A Figura 23 ilustra o posicionamento dos acentos nucleares na frase “*I started studying English when I was: eleven*” produzida pelo candidato DES e exemplificativa de desempenhos da faixa B. A Figura 24, por sua vez, exibe o posicionamento de acentos nucleares em “*my: experience learning (.) <is: &-um:> [/] (.) <is> [//] <&-um begins> [//] &-um begun in: (.) &-um two thousand (.) eleven*” produzida pelo candidato ROB e típica de faixas de proficiência inferiores.

Figura 23. Acento nuclear de DES.

Fonte: Elaborado pela autora.

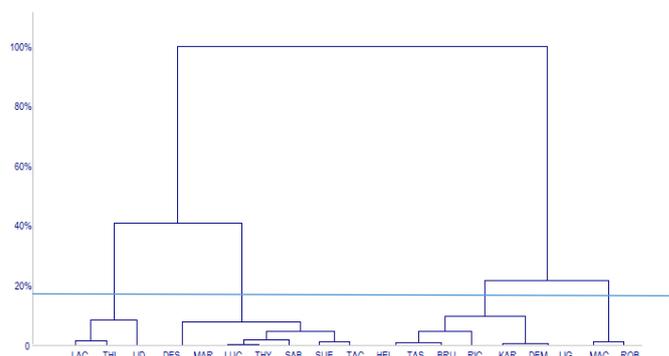
Figura 24. Acento nuclear de ROB.

Fonte: Elaborado pela autora.

Com relação à fluência, dentre as medidas mais empregadas encontram-se a velocidade de fala, a quantidade de pausas silenciosas, a quantidade de pausas preenchidas, a quantidade de repetições, a quantidade de reparos e a extensão média dos enunciados (BROWN; IWASHITA; MCNAMARA, 2005; DE JONG, 2018). De Jong (2018) acredita que, dos critérios comumente empregados para mensuração da fluência de fala dos candidatos, medidas correlatas à velocidade de fala mostram-se altamente eficientes para falantes de língua estrangeira.

Neste estudo, dentre os critérios apontados pelos usuários do exame, levou-se em consideração o alongamento de sons – em virtude de ser essa uma característica típica do português brasileiro, não comum na língua inglesa –, a velocidade de fala e a frequência de pausas e de pausas preenchidas. Os cálculos foram realizados automaticamente pela função Flucalc do programa CLAN.

Com relação à velocidade de fala, também foi possível a identificação de quatro grupos. O primeiro, composto por desempenhos iguais ou superiores a 121 palavras/minuto; o segundo, entre 95 e 120 palavras/minuto; o terceiro, entre 95 e 60 palavras/minuto; e o quarto, composto por desempenhos iguais ou inferiores a 59 palavras/minuto, conforme exibido no Gráfico 13.

Gráfico 13. Dendrograma de desempenhos de fala com relação à velocidade de fala

Nota: O eixo das abcissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

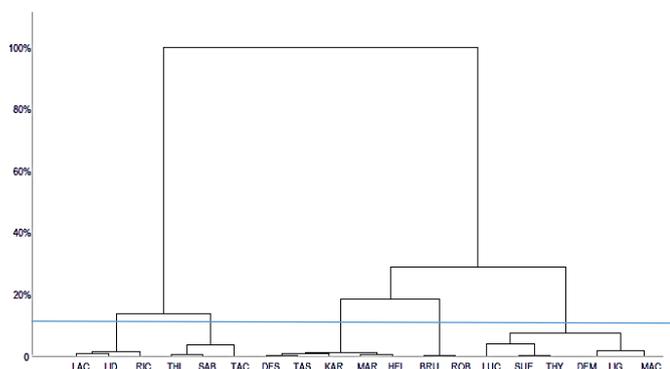
A análise do desempenho de fala confirma a existência de uma tendência de caráter discriminatório por parte do critério, conforme exibido pela Tabela 12.

Tabela 12. Resumo descritivo do emprego mediano de velocidade de fala

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	138.4800	0.0000	138.4800	138.4800
B	104.8890	19.1130	98.6535	117.7665
C	80.7905	25.9450	75.0000	100.9450
D	101.6160	24.1475	81.6190	105.7665
E	73.6925	14.7495	66.3178	81.0673

Fonte: Elaborado pela autora.

No que se refere ao alongamento de sons, a análise de agrupamentos possibilitou a identificação de quatro grupos: o primeiro, composto por uma frequência de alongamentos inferior a 2%; o segundo, composto por uma frequência entre 3 e 4%; o terceiro, por uma frequência de 5 a 6%; e o último com uma frequência superior a 7%, conforme exibido pelo Gráfico 14.

Gráfico 14. Dendrograma dos desempenhos de fala com relação ao alongamento de sons

Nota: O eixo das abscissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

A análise dos desempenhos de fala informou que a avaliação considerou os grupos de fala encontrados na análise de agrupamento, conforme exibido na Tabela 13.

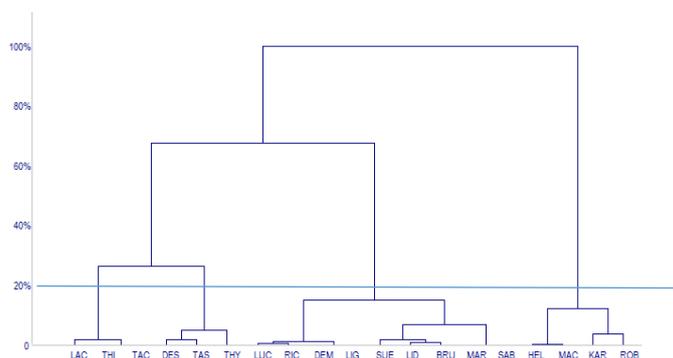
Tabela 13. Resumo descritivo do emprego mediano de alongamento de sons

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	1.9000	0.0000	1.9000	1.9000
B	5.3920	3.4045	3.5840	6.9885
C	7.2200	1.0455	6.3650	7.4105
D	5.8540	1.3905	4.8940	6.2845
E	5.6735	3.5135	3.9168	7.4303

Fonte: Elaborado pela autora.

Com relação às pausas, De Jong (2018) ressalta que pausas que ocorrem entre um enunciado e outro e que precedem palavras de baixa frequência devam ser tratadas como pausas esperadas, e não como rupturas. No entanto, durante a transcrição dos dados de fala para análise, percebeu-se que esse padrão não era corroborado pelas amostras do exame EPPLE, pois o emprego de pausas tanto silenciosas quanto preenchidas era extremamente frequente e diversificado. Desse modo, a análise focou-se apenas na contagem e categorização da quantidade de pausas silenciosas e preenchidas das amostras.

Com relação às pausas silenciosas, o Gráfico 15 exibe que foi possível determinar a existência de quatro grupos. No primeiro, as pausas eram inferiores a 4 pausas por minuto; no segundo, elas variavam entre 5 e 7 por minuto; no terceiro, a variação se dava entre 8 e 10 pausas por minuto; e no quarto, era superior a 11 pausas por minuto.

Gráfico 15. Dendrograma dos desempenhos de fala com relação às pausas silenciosas

Nota: O eixo das abscissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

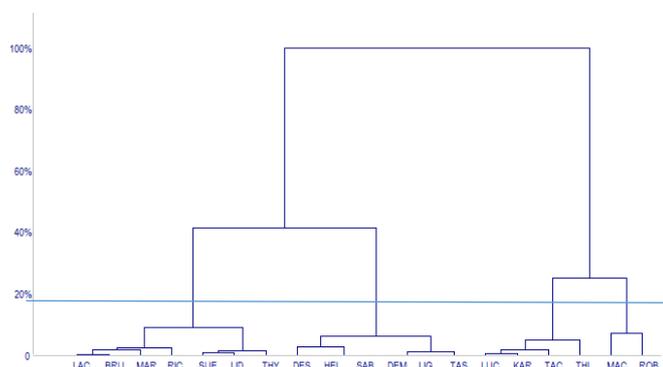
A Tabela 14 exibe que as pausas silenciosas aparentemente foram consideradas na análise do desempenho de fala dos candidatos.

Tabela 14. Resumo descritivo do emprego mediano de pausas silenciosas.

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	3.8630	0.0000	3.8630	3.8630
B	8.6520	1.5575	7.8260	9.3835
C	9.6265	2.7743	8.0335	10.8078
D	6.2200	4.4725	5.2320	9.7045
E	12.9240	2.7390	11.5545	14.2935

Fonte: Elaborado pela autora.

As pausas preenchidas, por sua vez, apresentaram um resultado na contramão da literatura. De acordo com De Jong *et al.* (2013), pausas preenchidas são mais frequentes em falas em língua estrangeira com níveis maiores de proficiência. No caso dos agrupamentos para o teste oral do EPPLE, contudo, identificou-se um padrão no qual elas foram empregadas com maior frequência por falante menos proficientes. No Gráfico 16, é possível visualizar a existência de quatro grupos. No primeiro, as pausas são empregadas com uma frequência igual ou inferior a uma por minuto; no segundo, o emprego varia de 2 a 3 pausas por minuto; no terceiro, de 4 a 5; e no quarto, são iguais ou superiores a 6 pausas por minuto.

Gráfico 16. Dendrograma dos desempenhos de fala com relação às pausas preenchidas

Nota: O eixo das abcissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

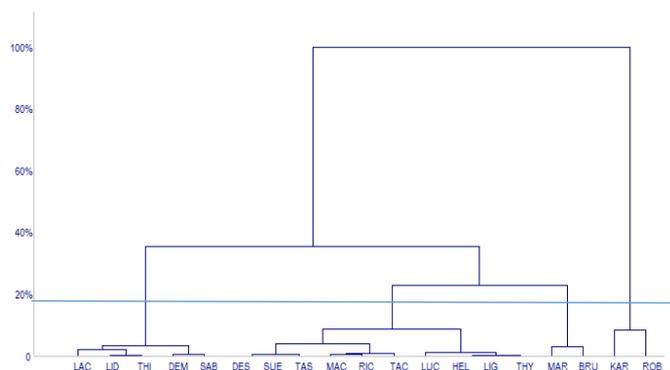
A tabela 15 ressalta a criterialidade das pausas preenchidas na contramão da literatura internacional.

Tabela 15. Resumo descritivo do emprego mediano de pausas preenchidas

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	0.6330	0.0000	0.6330	0.6330
B	1.7840	2.5073	1.2335	3.7408
C	2.9450	0.4603	2.6518	3.1120
D	5.2800	2.5650	3.4325	5.9975
E	4.6045	3.6785	2.7653	6.4438

Fonte: Elaborado pela autora.

Os desempenhos relativos ao emprego de repetições e de reparos nos enunciados foram analisados como indicativos de flexibilidade e seus cálculos também foram fornecidos pela função Flucal do programa CLAN. Com relação às repetições, foi possível identificar a presença de quatro grupos, conforme exibido no Gráfico 17. O primeiro grupo apresentou um desempenho em que a taxa das repetições era inferior a 0,4; o segundo apresentou uma variação entre 0,5 e 1,5; o terceiro, entre 1,5 e 2,5; e o quarto, acima de 2,5.

Gráfico 17. Dendrograma dos desempenhos de fala com relação às repetições

Nota: O eixo das abscissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

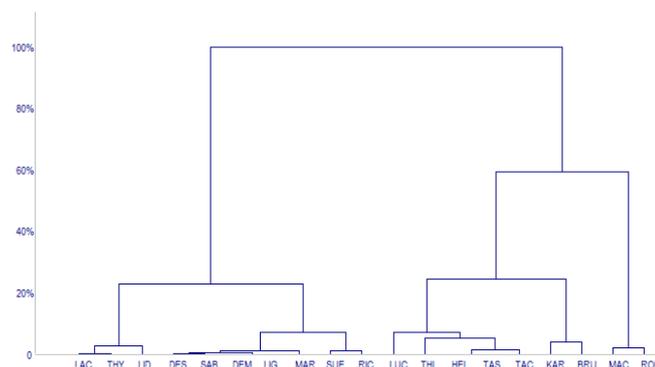
A Tabela 16 aponta para uma tendência na consideração do critério na avaliação dos dados de fala do teste oral do exame.

Tabela 16. Resumo descritivo do emprego mediano de repetições

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	0.0000	0.0000	0.0000	0.0000
B	0.5830	0.8715	0.1998	1.0713
C	1.0190	0.9535	0.4240	1.3775
D	0.8280	0.2505	0.7735	1.0240
E	2.6450	1.8730	1.7085	3.5815

Fonte: Elaborado pela autora.

Com relação às reformulações, tem-se que foi possível encontrar dados significativos para apenas três grupos, conforme exibido pelo Gráfico 18. No primeiro grupo, elas eram superiores a 0,5; no segundo, variavam entre 0,6 e 2,9; e, no terceiro, eram superiores a 3.

Gráfico 18. Dendrograma dos desempenhos de fala com relação às reformulações

Nota: O eixo das abcissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

A Tabela 17 aponta para a consideração deste critério na avaliação dos dados de fala do exame.

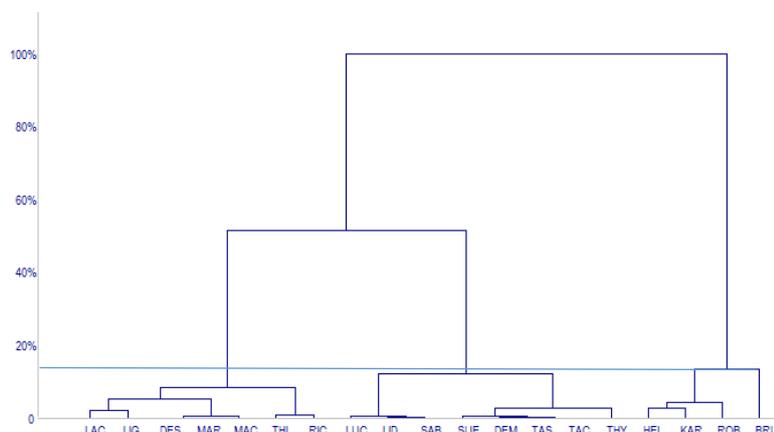
Tabela 17. Resumo descritivo do emprego mediano de reformulações

Faixa	Mediana	Desvio Interquartil	Primeiro Quartil	Terceiro Quartil
A	0.2530	0.0000	0.2530	0.2530
B	0.7200	0.3053	0.5850	0.8903
C	1.2280	0.6788	0.7070	1.3858
D	1.1390	1.3170	0.6915	2.0085
E	1.7375	1.2745	1.1003	2.3748

Fonte: Elaborado pela autora.

Com relação ao critério de precisão linguística, neste trabalho, empregou-se a medida tradicional: a contagem dos erros em relação ao total de palavras produzidas. Em termos lexicais, os erros considerados foram resumidos a erros idiossincráticos, ou seja, à produção de não-palavras, a erros gerados por transferência linguística, a erros de escolhas de vocábulos de conteúdo e a erros na escolha de palavras funcionais. Com relação aos erros de ordem gramatical, considerou-se o cometimento de erros de concordância, de erro de ordem de palavras na sentença, omissões de constituintes funcionais e erros de transferência linguística.

A análise dos desempenhos possibilitou a categorização dos perfis em quatro grupos. O primeiro, composto pela ocorrência de um índice de 0,03 ou inferior; o segundo, de 0,04 a 0,06; o terceiro, de 0,07 a 0,09; e o quarto, igual ou superior a 0,1, conforme exibido no Gráfico 19.

Gráfico 19. Dendrograma dos desempenhos de fala com relação à precisão linguística

Nota: O eixo das abscissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

A análise das avaliações revelou a existência de uma relação direta entre as faixas e as categorias encontradas, como exibe a Tabela 18.

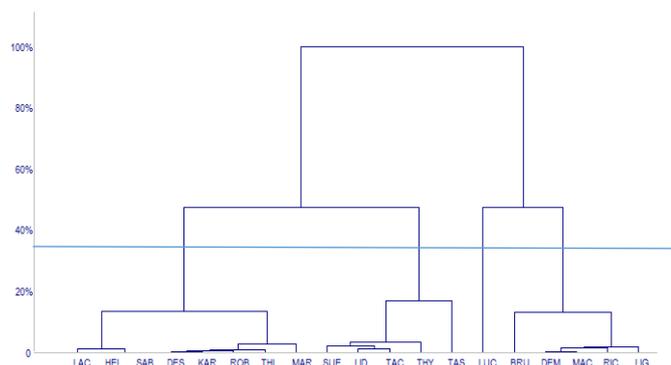
Tabela 18. Resumo descritivo do emprego mediano de precisão linguística

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	0.0151	0.0001	0.0150	0.0151
B	0.0416	0.0111	0.0315	0.0426
C	0.0548	0.0467	0.0484	0.0950
D	0.0493	0.0168	0.0353	0.0521
E	0.0667	0.0360	0.0486	0.0847

Fonte: Elaborado pela autora.

Com relação à complexidade sintática, por sua vez, tem-se que, tradicionalmente, ela é mensurada por meio de medidas como a medida global, que considera o número de palavras produzidas por sentença e o índice de coordenação que, no caso deste trabalho, levou em consideração a proporção de orações coordenadas em relação às subordinadas. Os cálculos também foram obtidos por meio do programa CLAN.

Com relação à medida global, o Gráfico 20 exibe o encontro de quatro grupos de análise. O primeiro, composto por 16 ou mais palavras por sentença; o segundo, de 13 a 15; o terceiro, de 12 a 8; e o quarto, inferior a 11 palavras por sentença.

Gráfico 20. Dendrograma dos desempenhos de fala com relação à medida global.

Nota: O eixo das abcissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

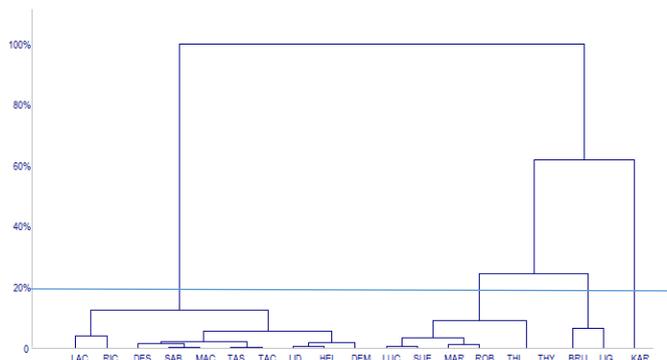
No entanto, conforme exibido na Tabela 19, a avaliação dos desempenhos não parece relacionar-se de maneira direta à medida global.

Tabela 19. Resumo descritivo do emprego mediano de medida global

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	10.9417	0.0000	10.9417	10.9417
B	11.6529	1.3377	10.6045	11.9422
C	11.7059	2.0031	11.0915	13.0946
D	9.8806	1.8071	9.7419	11.5490
E	12.5758	0.8066	12.1725	12.9791

Fonte: Elaborado pela autora.

Já com relação ao índice de coordenação, o Gráfico 21 exibe a existência de quatro grupos. O primeiro, inferior a uma proporção de 1,2; o segundo, entre 1,3 e 1,9; o terceiro, de 2 a 2,5; e o quarto, superior a 2,5.

Gráfico 21. Dendrograma dos desempenhos de fala com relação ao índice de coordenação

Nota: O eixo das abcissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

Os dados da Tabela 20, por sua vez, não demonstram a existência de relação entre a proporção de orações coordenadas em relação às subordinadas e à classificação dos candidatos.

Tabela 20. Resumo descritivo do emprego mediano do índice de coordenação

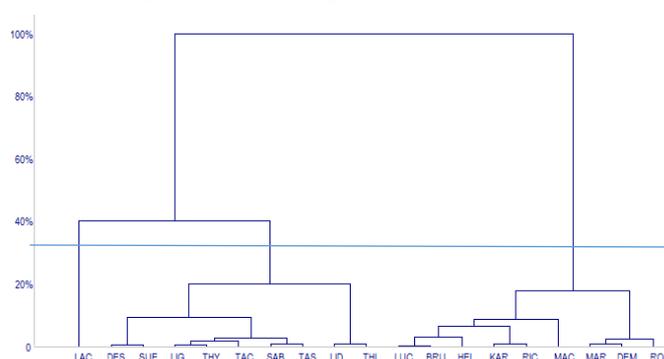
Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	0.9761	0.0000	0.9761	0.9761
B	1.6060	0.4539	1.3128	1.7667
C	1.2939	1.0768	1.1832	2.2599
D	1.1538	0.2102	1.1394	1.3496
E	1.2513	0.4819	1.0103	1.4922

Fonte: Elaborado pela autora.

Dentre as medias de complexidade lexical mais empregadas na literatura, encontra-se a avaliação da relação entre *types* e *tokens* na fala dos candidatos. *Tokens* consistem em efetivamente todas as palavras produzidas em um texto; ao passo que *types* se referem apenas às palavras produzidas apenas uma vez, não contabilizando as repetições. A relação entre *types* e *tokens* pode ser calculada de diversas maneiras. Uma delas consiste no Índice de Guiraud, que é calculado dividindo-se o número total de *types* pela raiz quadrada do número total de *tokens* da fala. Hout e Vermeer (2007) afirmam que os resultados fornecidos pelo índice apresentam menos variações em relação à *type-token ratio* (TTR), geralmente empregada em estudos da Linguística Aplicada, especialmente se a relação entre *types* e *tokens* não é linear, sendo, portanto, uma medida mais válida.

Nos dados de fala investigados neste estudo, identificou-se a presença de três grupos. No primeiro, o índice tem valor superior a 52; no segundo, varia de 39 a 51; e, no terceiro, é inferior a 38, conforme exibido no Gráfico 22.

Gráfico 22. Dendrograma dos desempenhos de fala com relação ao Índice de Guiraud.



Nota: O eixo das abcissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

A Tabela 21, exibe uma provável consideração da complexidade lexical na avaliação dos desempenhos.

Tabela 21. Resumo descritivo do emprego mediano de Índice de Guiraud

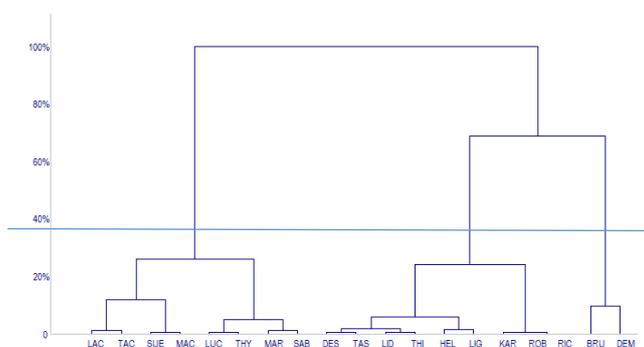
Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	67.4717	0.0000	67.4717	67.4717
B	42.3756	9.4944	39.5716	49.0660
C	38.1064	12.3616	33.5385	45.9000
D	44.4305	9.1710	35.9767	45.1477
E	35.8681	0.7641	35.4861	36.2501

Fonte: Elaborado pela autora.

Outra maneira de se mensurar a complexidade lexical, agora por meio de sofisticação, consiste na avaliação do emprego de palavras de baixa frequência. Pela literatura o emprego deste tipo de palavra relaciona-se a níveis maiores de proficiência. No entanto, o estudo de Brown, Iwashita e McNamara (2005) não encontrou uma relação significativa entre ambas as variáveis. No caso de falantes brasileiros de inglês como língua estrangeira, é válido ressaltar, ainda, que, pela origem latina do português, o emprego de palavras de menor frequência acaba sendo mais confortável, conforme exibido em Colombo (2014).

A contagem dos tipos de palavras neste trabalho foi realizada por meio do programa RANGE. O resultado da análise de conglomerados nos dados de fala do EPPLÉ informa a existência de três grupos com relação à proporção de emprego de palavras de alta e baixa frequência que vão na contramão da literatura. O primeiro apresenta uma proporção inferior a 5; o segundo, uma proporção que varia de 10 a 5; e o terceiro, superior a 10.

Gráfico 23. Dendrograma dos desempenhos de fala com relação à frequência de palavras.



Nota: O eixo das abscissas (x) exibe a quantidade de funções empregada por candidato. O eixo das ordenadas (y) exibe a distância dos grupos de conglomerados.

Fonte: Elaborado pela autora.

A Tabela 22 exibe que a avaliação dos desempenhos de fala parece ter considerado o emprego de palavras de baixa frequência de acordo com o exibido pelo Gráfico 23.

Tabela 22. Resumo descritivo do emprego mediano de frequência de palavras

Faixa	Mediana	Desvio Interquartilico	Primeiro Quartil	Terceiro Quartil
A	4.7619	0.0000	4.7619	4.7619
B	8.0313	1.7826	7.0248	8.8074
C	8.9552	3.5146	8.3039	11.8185
D	6.0150	1.1670	5.2616	6.4286
E	10.1157	0.0033	10.1140	10.1174

Fonte: Elaborado pela autora.

Ressalta-se que, para os dendrogramas apresentados nesta subseção, empregou-se apenas alguns grupos de candidatos representativos da amostra geral, em virtude da dificuldade de visualização dos valores geradas pela exibição gráfica com todos os participantes. Os resultados obtidos nesta etapa de análise consistem em informações cruciais para a elaboração da escala de mensuração para o teste oral do exame.

4.3 Proposição de uma abordagem avaliativa da proficiência oral do professor de línguas para o EPPLE

A abordagem avaliativa aqui proposta consiste na reestruturação de parâmetros de mensuração para o exame EPPLE. Assim, propõe-se a geração de escalas que congreguem perfis de proficiência internacionalmente aceitos e perfis de proficiência elicitados pelo exame. Além disso, sugere-se o tratamento da dinamicidade interacional dos critérios de avaliação que integram essas escalas por meio do método *Fuzzy TOPSIS* em uma tentativa de agregar valores dos usuários (*stakeholders*) ao exame e de oferecer um caminho prático e válido de avaliação.

4.3.1 A proposição de escalas para o teste oral do exame EPPLE

A análise da relação entre a escala de proficiência empregada pelo exame e os critérios elicitados pelas tarefas revela a necessidade de criação de duas escalas com critérios equivalentes, mas finalidades distintas para o EPPLE. Uma delas consiste na escala de proficiência propriamente dita, que tem a função de informar os candidatos a respeito da caracterização de sua proficiência em relação a quadros de referência internacionalmente aceitos, como o CEFR e o CETF. Semelhantemente à escala atual do exame, propõe-se que ela seja de caráter holístico, ou seja, preocupe-se majoritariamente com o atingimento da

comunicação verbal, mas inclua descritores analíticos, ou seja, mostre como os componentes da competência linguística interagem, garantindo eficácia ou prejudicando a comunicação.

A outra escala consiste em uma escala de mensuração, ou seja, uma escala que esteja diretamente vinculada às tarefas do exame e que informe as variações dos perfis linguísticos previstos pela escala de proficiência. Tal escala não tem finalidade informativa, ou seja, não é disponibilizada aos candidatos, ficando restrita aos examinadores e guiando a análise dos desempenhos de fala. É imprescindível, contudo, que ambas as escalas estejam fortemente vinculadas, pois, como afirma McNamara (2006), em exames de proficiência, as inferências – presentes na escala de caráter informativo – devem ser possibilitadas adequadamente pela observação de desempenho em critérios confiáveis – os quais se encontram dispostos na escala de mensuração.

4.3.1.1 A escala de mensuração analítica para o teste oral do exame EPPLE

A escala de mensuração aqui proposta leva em consideração as respostas dos usuários (*stakeholders*) aos questionários, as funções comunicativas elicitadas pelas tarefas do teste oral, os critérios do CETF e do CEFRL, e a criterialidade informada por meio da análise de agrupamentos dos desempenhos de fala dos candidatos ao exame. Diante do construto comunicativo do EPPLE, considera-se necessário que todos os dados de fala, para serem analisados e incluídos em uma faixa de proficiência de maneira adequada, devem ser previamente analisados com relação ao atendimento ao tópico/tema da tarefa, ou seja, não se recomenda, pelos padrões aqui estabelecidos, a avaliação de desempenhos de fala que não tratem especificamente dos assuntos elicitados.

A escala de mensuração de orientação analítica foi elaborada considerando-se as variações de desempenho informadas pela análise de agrupamento de desempenho dos candidatos nos critérios possibilitados pelas tarefas. A escala tem a função de minimizar possíveis interferências de preferências particulares por parte dos avaliadores e de oferecer padrões de mensuração reais e padrões de inferência úteis para o teste oral do exame EPPLE, em virtude de sua elaboração com base em desempenhos de candidatos do exame e, paralelamente, nos quadros de referência internacionais.

Desse modo, tem-se que a escala se subdivide em elementos comunicativos, elementos linguístico-pedagógicos e elementos linguísticos. No primeiro grupo, são apresentadas as variações de desempenho para as três primeiras tarefas do exame, que requerem o emprego da língua para finalidades gerais. Os critérios empregados, a saber a organização textual e a extensão do tópico, são passíveis de variações a depender das instruções fornecidas e, portanto,

devem ser analisados de maneira independente para cada uma das tarefas. No segundo grupo, são condensados critérios relativos, especificamente, às tarefas de uso de linguagem para finalidades pedagógicas, os quais são, novamente, limitados ao contexto da tarefa. Compõem este grupo os critérios de mobilização de funções e de precisão de informações. A terminologia não foi empregada em virtude de seu uso não ser explicitamente solicitado em nenhuma das tarefas do exame. O terceiro grupo é composto, por fim, por critérios de ordem linguística e, diferentemente dos demais, são avaliados em relação ao desempenho global apresentado pelo candidato no teste. O Quadro 32 condensa a escala de orientação analítica gerada.

Quadro 32. Escala de mensuração para o teste oral do EPPL.

Critério	Relação com a tarefa	Unidade de mensuração	Tipo	Variação					
				A	B	C	D		
				C1+	B2	B1	A2-		
				<i>Expert</i>	<i>Proficiency</i>	<i>Developing</i>	<i>Foundation</i>		
Elementos comunicativos	Dependente	T1	Expansão do tópico	# tópicos expandidos	Max.			3 ou mais tópicos expandidos	2 ou menos tópicos expandidos
		T2	Expansão do tópico	Presença/Ausência de detalhes Relevância de detalhes	Bin.		Há detalhes relevantes	Há detalhes não relevantes/ Não há detalhes	
			Organização textual	Tipo de organização	Cat.		Sequência linear		Lista
		T3	Expansão do tópico	Presença/Ausência de detalhes Relevância de detalhes	Bin.		Há detalhes relevantes	Há detalhes não relevantes/ Não há detalhes	
			Organização textual	Tipo de organização			Sequência linear		Lista
Elementos linguístico-pedagógicos	T4	Funções mobilizadas	# funções pedagógicas Variedade de funções	Max. Bin.	6 ou mais funções Presença de município linguístico	5 funções	4 funções	3 ou menos funções Ausência de município linguístico	
		T5	Precisão de informações	(# tópicos corretos / # tópicos) x 100	Max.	100% são precisas	90% ou menos são precisas	60% ou menos são precisas	40% ou menos são precisas
			Elementos linguísticos	Independente	Precisão linguística	Índice de erros = (# erros gramaticais + # erros lexicais) / # tokens	Min.	0,03 ou menos	entre 0,04 e 0,05
Elementos linguísticos	Independente	Complexidade linguística	Índice de Guiraud = # tokens / $\sqrt{\# \text{types}}$	Max.	52 ou mais	entre 39 e 51	38 ou menos		
			Sofisticação lexical = $[(\# \text{token} - (\# \text{token L1} + \text{L2} + \text{L3})) \times 100 / \# \text{token L1} + \text{L2} + \text{L3}]$	Min.	5% ou menos	entre 9 e 4%		10% ou mais	
			Índice de extensão média do enunciado	Max.	18 ou mais	entre 14 e 17	entre 13 e 9	8 ou menos	
			Índice de coordenação = # conjunções coordenativas / # conjunções subordinativas	Min.	1,2 ou menos	entre 1,3 e 1,9	entre 2 e 2,5	2,5 ou mais	
			Medida global = # tokens / oração	Max.	16 ou mais	entre 13 e 15	entre 11 e 12	10 ou menos	
	Independente	Inteligibilidade da pronúncia	Colocação do acento nuclear	Cat.	Substantivos e verbos nucleares	Verbos nucleares e final da sentença	Posição inicial e final da setença	Posição final da sentença	
	Independente	Fluência	Frequência de alongamento	Min.	2% ou menos	entre 3 e 4%	entre 5 e 7%	8% ou mais	
			Velocidade de fala	Max.	121 palavras/min ou mais	entre 95 e 120	entre 60 e 95	59 palavras/minuto ou menos	
			Frequência de reparo	Min.	0,5 ou menos	entre 0,6 e 2,9		3 ou mais	
			Frequência de repetições	Min.	0,4 ou menos	entre 0,5 e 1,5	entre 1,5 e 2,5	acima de 2,5	
Frequência de pausas silenciosas			Min.	4 ou menos	entre 5 e 7	entre 8 e 10	11 ou mais		
Frequência de pausas preenchidas	Min.	2 ou menos	entre 2 e 3	entre 4 e 5	6 ou mais				

Fonte: Elaborada pela autora.

*Nota: Tipo de avaliação do critério: quantitativa (máx. = maximização, min. = minimização), categórico (cat.) ou binário (bin.).

4.3.1.2 A escala de proficiência oral para o exame EPPLE

A escala de proficiência proposta ao exame EPPLE orienta-se por uma abordagem comunicativa de língua (MORROW, 2018) e estrutura-se pela descrição de perfis de desempenho linguístico em situações de uso de língua para finalidades gerais definidos pelo CEFR (COE, 2018) e de desempenho de uso para finalidades pedagógicas definidos por alguns critérios do CETF (CAE, 2018).

A escala encontra-se dividida em quatro faixas principais – a saber, faixa A, faixa B, faixa C e faixa D – e em um nível complementar, que tem por finalidade, unicamente informar que os dados de determinado desempenho de fala foram recebidos pela equipe de correção, mas que tais dados não são adequados para avaliação em virtude de características de fala do próprio candidato – como fuga ao tema ou ininteligibilidade da fala – ou por questões de ordem técnica – como chiados ou cortes de gravação. A escala tem seu limiar inferior, representado pela faixa D, definido pelo nível A2 do CEFR em virtude de ser esse o nível mínimo para se atuar com o ensino de línguas estabelecido pelo CETF (CAE, 2018). A determinação do nível C1 como limiar superior, por sua vez, representado pela faixa A, atrela-se aos níveis de desempenho possibilitados pelas tarefas que compõem os dados do teste oral empregados neste estudo.

A segmentação da escala acabou por contemplar, assim, além dos perfis estabelecidos pelo CETF, os quatro níveis de fala desejados pelo coordenador da equipe responsável pelo exame²⁵, que visava à eliminação da faixa de desempenho intermediária – até então representada pela faixa C – pela possibilidade de preferência de classificação de candidatos nesta faixa diante da presença de dúvidas de desempenho por parte dos examinadores.

Os critérios que possibilitam as inferências expressas pelos descritores que compõem as faixas encontram-se ancorados no modelo de competência linguística de Bachman (1990) e no modelo de competência linguístico-comunicativa de Almeida Filho (1993). Para uma melhor informatividade, optou-se pela elaboração de uma escala holística com descritores analíticos, subdivididos em três grandes dimensões. A primeira dimensão mobiliza as subcompetências organizacional e pragmática do modelo de Bachman (1990) ao focar-se nas competências funcional, textual e sociolinguística, as quais são representadas pelos critérios de desenvolvimento de tópico, precisão proposicional, coerência e coesão, flexibilidade e registro. A segunda dimensão volta-se, exclusivamente, à efetividade do uso da competência gramatical para atingir os objetivos comunicacionais e aborda aspectos relativos ao léxico, em termos de

²⁵ CONSOLO, D. A. Comunicação pessoal.

precisão e variedade; à sintaxe, em termos de complexidade; à fonologia, em termos de clareza, precisão e monitoramento; e à morfologia, em termos de precisão. A terceira e última dimensão consiste, por fim, em uma mescla das competências abordadas anteriormente, mas, agora, orientadas exclusivamente pela função heurística, mais especificamente, de ensino, agregando critérios relativos à linguagem docente, especialmente em termos de precisão, variedade e sofisticação terminológica. No entanto, diferentemente das demais, a terceira dimensão requer a presença do critério de conhecimento sobre a língua, cuja descrição é orientada pela capacidade de reconhecimento de erros autênticos cometidos por alunos imaginários e pela capacidade de atuar como modelo linguístico a esses alunos em diferentes níveis de proficiência.

De maneira geral, as funções comunicativas – descrição, apresentação, fornecimento de opinião, sustentação de opinião e fornecimento de exemplos – empregadas na escala encontram-se intimamente relacionadas às funções mobilizadas pelas tarefas do exame.

Padrão fonológico empregado passa a ser o do inglês como língua franca que, apesar de controverso, consegue funcionar como um padrão um pouco mais distante da figura do falante nativo do que a noção de falante competente, além de ser menos subjetivo do que este pela existência de padrões empíricos.

Interrupções, autocorreção e imprecisões linguísticas são incluídas inclusive na faixa A e tratadas como normais na fala, desde que não afetem a compreensibilidade geral.

A definição das características distintivas dos componentes de cada faixa foi determinada com base nos perfis linguísticos apresentados nos quadros de referência internacionais mencionados. Deste modo, entre os níveis B, C e D, a distintividade dos critérios é definida por meio da qualidade da precisão e da complexidade da linguagem utilizada, ao passo que a faixa A se diferencia das demais por conter informações voltadas à variedade de uso da língua.

O Quadro 33 apresenta a escala de proficiência proposta.

Quadro 33. Escala de proficiência oral sugerida ao exame EPPLE com base em quadros de referência internacionais.

ESCALA DE PROFICIÊNCIA ORAL DO EXAME EPPLE			
CEFR/CEFT	FAIXAS	Avaliação holística	Crítérios analíticos
C1+	A	<p>A1. Fornece descrições, apresentações e opiniões detalhadas sobre assuntos complexos, sustentando suas ideias com argumentos elaborados, precisos e relevantes.</p> <p>Sua fala é bem estruturada, utilizando-se de uma variedade de recursos de organização, articulação e coesão textual.</p> <p>Consegue modificar sua fala para expressar graus de compromisso e confiança. É capaz de ajustar o nível de formalidade de sua fala para adequar-se ao contexto social.</p>	<i>Desenvolvimento de tópico:</i>
			<i>Precisão proposicional:</i>
			<i>Coerência e coesão:</i>
			<i>Flexibilidade:</i>
		<p>A2. A fala é inteligível. Interrupções no fluxo da fala ocorrem apenas para selecionar palavras e exemplos precisos para expressão do pensamento.</p> <p>Demonstra controle sobre a articulação clara e precisa dos sons do inglês como língua franca. É capaz de notar deslizes de pronúncia e se autocorriger. Emprega acento, ritmo e entonação de maneira apropriada e efetiva para comunicar pequenas diferenças de sentido.</p> <p>Demonstra controle sobre um amplo repertório lexical, sendo capaz de contornar falhas, quando estas ocorrem, por meio do emprego de perífrases.</p> <p>Demonstra um alto grau de precisão gramatical para estruturas simples e complexas. A presença de erros é escassa e não interfere na comunicação.</p>	<i>Registro:</i>
			<i>Precisão lexical:</i>
			<i>Variedade lexical:</i>
			<i>Precisão gramatical:</i>
		<p>A3. Utiliza ampla variedade de linguagem de sala de aula com consistência. Consegue empregar apropriadamente vocabulário técnico e expressões idiomáticas sofisticadas para descrição de língua.</p> <p>É capaz de reconhecer todos os erros cometidos por alunos e de fornecer exemplos adequados de assuntos linguísticos dos níveis A1, A2, B1, B2, C1 e C2 do CEFR.</p>	<i>Complexidade sintática:</i>
			<i>Inteligibilidade:</i>
			<i>Fluência:</i>
			<i>Linguagem docente:</i>
			<i>Modelo de língua:</i>
			<i>Reconhecimento de erros:</i>
			<i>Terminologia:</i>

B2	B	<p>B1. Fornece descrições, apresentações e opiniões detalhadas sobre uma variedade de assuntos relacionados a sua área de interesse e sobre o significado pessoal de eventos e experiências, sustentando suas ideias com argumentos e exemplificações simples, mas precisas e relevantes.</p> <p>Sua fala é bem organizada, utilizando-se de uma variedade de recursos coesivos para encadeamento de ideias. Pode haver alguns problemas no encadeamento de contribuições longas.</p> <p>Consegue reformular sua fala para enfatizar ou explicar um assunto. É capaz de se expressar com confiança, clareza e polidez em registros formais e informais, apropriados ao contexto social.</p>	<i>Desenvolvimento de tópico:</i>
			<i>Precisão proposicional:</i>
			<i>Coerência e coesão:</i>
			<i>Flexibilidade:</i>
			<i>Registro:</i>
		<p>B2. A fala é inteligível, apesar de haver alguns deslizes de pronúncia. Interrupções no fluxo da fala ocorrem na busca por padrões e expressões para comunicação.</p> <p>Demonstra controle sobre a articulação clara e precisa de uma variedade de sons do inglês como língua franca. É capaz de empregar acento, ritmo e entonação de maneira apropriada e efetiva.</p> <p>Demonstra controle sobre um repertório lexical para tratar de tópicos genéricos e de assuntos relacionados a sua área. Consegue variar sua formulação para evitar repetições frequentes. Lacunas lexicais geram hesitações e circunlóquios, contudo, escolhas lexicais inapropriadas não interferem nos objetivos da comunicação.</p> <p>Demonstra bom controle gramatical sobre estruturas simples. Desvios, erros não sistemáticos e pequenas falhas na estrutura da sentença podem ocorrer, mas são escassos e podem ser corrigidos. O uso de estruturas complexas pode ocorrer com algumas inconsistências.</p>	<i>Precisão lexical:</i>
			<i>Variedade lexical:</i>
			<i>Precisão gramatical:</i>
			<i>Complexidade sintática:</i>
			<i>Inteligibilidade:</i>
		<p>B3. Utiliza linguagem de sala de aula consistentemente apropriada. Consegue empregar adequadamente uma variedade de terminologia para descrição de língua.</p> <p>É capaz de reconhecer quase todos os erros cometidos por alunos e de fornecer exemplos adequados de assuntos linguísticos dos níveis A1, A2, B1 e B2 do CEFR.</p>	<i>Fluência:</i>
			<i>Linguagem docente:</i>
			<i>Modelo de língua:</i>
B1	C	<p>C1. Fornece descrições, apresentações e opiniões diretas sobre uma variedade de assuntos familiares, reais ou imaginários, relacionados a sua área de interesse e a experiências, sentimentos e reações. Consegue explicar os pontos principais de suas ideias em detalhes simples e curtos, mas precisos.</p>	<i>Reconhecimento de erros:</i>
			<i>Terminologia:</i>
			<i>Desenvolvimento de tópico:</i>
			<i>Precisão proposicional:</i>
			<i>Coerência e coesão:</i>

		<p>Sua fala é estruturada em uma sequência linear de frases longas, utilizando-se de pouca variedade de recursos coesivos para encadeamento. Consegue apresentar contra-argumentos em um texto discursivo simples.</p> <p>Consegue explorar uma ampla gama de linguagem simples para expressar o que quer dizer e reformular a fala para tratar de situações menos familiares. É capaz de agir de acordo com convenções de polidez importantes.</p>	<i>Flexibilidade:</i>		
			<i>Registro:</i>		
		<p>C2. A fala é geralmente inteligível, apesar de deslizes de pronúncia regulares. Interrupções no fluxo da fala são frequentes e ocorrem para fins de planejamento gramatical e lexical e para realizar reparos na formulação, especialmente em trechos de fala longos.</p> <p>Apresenta erros regulares de pronúncia de sons individuais e de palavras que interferem na inteligibilidade. Acento, ritmo e entonação não são empregados de maneira efetiva, mas não interferem na inteligibilidade da fala.</p> <p>Demonstra controle suficiente sobre um repertório lexical simples para tratar de assuntos relacionados a tópicos familiares e situações cotidianas. Sua fala é marcada pela presença de circunlóquios. Comete erros graves quando fala sobre tópicos complexos ou assuntos não familiares.</p> <p>Demonstra precisão gramatical razoável no uso de estruturas simples e familiares. Erros ocorrem, mas não interferem no conteúdo da mensagem. Utiliza-se de um repertório de padrões frequentemente empregados em situações mais previsíveis.</p>	<i>Precisão lexical:</i>		
			<i>Variedade lexical:</i>		
			<i>Precisão gramatical:</i>		
			<i>Complexidade sintática:</i>		
			<i>Inteligibilidade:</i>		
			<i>Fluência:</i>		
		<p>C3. Utiliza linguagem de sala de aula com algumas inadequações. Consegue empregar pouca terminologia adequada para descrição de língua.</p> <p>É capaz de reconhecer a maioria dos erros cometidos por alunos e de fornecer exemplos adequados de assuntos linguísticos dos níveis A1, A2 e B1 do CEFR.</p>	<i>Linguagem docente:</i>		
			<i>Modelo de língua:</i>		
			<i>Reconhecimento de erros:</i>		
			<i>Terminologia:</i>		
		A2-	D	<p>D1. Fornece descrições e apresentações simples e curtas sobre posses, lugares, rotinas, preferências, habilidades, planos e experiências. Consegue fornecer exemplos e comparar elementos de maneira muito simples e direta.</p> <p>Sua fala é estruturada em frases simples e utiliza-se de recursos coesivos básicos para encadeamento de ideias em listas.</p> <p>Consegue adaptar frases memorizadas e simples por meio de substituição lexical limitada ou expandir frases aprendidas por meio da recombinação simples de seus elementos. É capaz de socializar de maneira simples, mas efetiva, utilizando-se de expressões comuns e seguindo rotinas básicas.</p>	<i>Desenvolvimento de tópico:</i>
					<i>Precisão proposicional:</i>
<i>Coerência e coesão:</i>					
<i>Flexibilidade:</i>					
<i>Registro:</i>					

		<p>D2. A fala é geralmente inteligível para assuntos restritos a situações cotidianas, apesar de deslizes frequentes de pronúncia. Interrupções no fluxo da fala, como pausas, falsos inícios e reformulações são frequentes inclusive em trechos de fala curtos.</p> <p>Apresenta erros sistemáticos de pronúncia de fonemas do inglês como língua franca. Acento, ritmo e entonação são empregados de maneira adequada apenas para palavras e frases simples e familiares, podendo interferir na inteligibilidade da fala.</p> <p>Demonstra controle sobre um repertório lexical limitado, mas suficiente para tratar de necessidades cotidianas básicas e concretas.</p> <p>Demonstra controle sobre o uso de estruturas gramaticais simples. Erros básicos são cometidos sistematicamente, mas geralmente não interferem na inteligibilidade.</p>	<i>Precisão lexical:</i>
			<i>Variedade lexical:</i>
		<p>D3. Utiliza linguagem de sala de aula básica com inadequações. Não consegue empregar terminologia adequada para descrição de língua.</p> <p>É capaz e reconhecer erros básicos cometidos por alunos e de fornecer exemplos adequados de assuntos linguísticos dos níveis A1 e A2 do CEFR.</p>	<i>Precisão gramatical:</i>
			<i>Complexidade sintática:</i>
			<i>Inteligibilidade:</i>
			<i>Fluência:</i>
			<i>Linguagem docente:</i>
			<i>Modelo de língua:</i>
			<i>Reconhecimento de erros:</i>
			<i>Terminologia:</i>
-	E	Os dados de fala não são apropriados para avaliação.	<i>Motivo:</i>

Fonte: Elaborado pela autora.

Em relação ao CEFR e aos exames de proficiência apresentados no Quadro 6, a escala proposta para o EPPLÉ tem a pretensão exibida no Quadro 34. Ressalta-se que o quadro não tem por finalidade comparar ou equiparar as faixas de proficiência dos diferentes exames, mas, sim, apresenta-lás em relação ao CEFR.

Quadro 34. CEFR em relação a exames consagrados e ao EPPLÉ

CEFR	TOEFL iBT	IELTS	B2 First	C1 Advanced	C2 Proficiency	EPPLÉ
C2	95	9			220-230 Grade A	A
					213-219 Grade B	
C1		7-8		200-2010 Grade A	200-212 Grade C	
				193-199 Grade B	180-199	
B2	72	5,5-6,5	180-190 Grade A	180-192 Grade C		B
			173-179 Grade B	160-179		
B1	42	4,5-5	140-159			C
A2		4				D
A1						

Fonte: Elaborado pela autora.

4.3.2 A avaliação da proficiência oral com auxílio do método multicritério de análise da decisão *Fuzzy* TOPSIS

Para a reavaliação dos dados de fala por meio do método Fuzzy TOPSIS, foi necessário, primeiramente, a definição dos pesos a serem empregados na análise e a preparação de matrizes de comparação dos critérios em relação às alternativas.

Os critérios empregados para avaliação dos dados de fala encontram-se discriminados na escala de mensuração. O peso dos critérios foi atribuído levando-se em consideração (i.) a hierarquização de importância para os avaliadores identificada em Brown, Iwashita e McNamara (2005) – considerando-se como critérios conteúdo, precisão linguística, complexidade linguística, precisão fonológica e fluência – e para os especialistas em ensino de língua estrangeira identificados em Consolo (2017) – considerando-se flexibilidade, precisão linguística, conteúdo, inteligibilidade fonológica e fluência –, (ii.) a opinião discriminada na análise das respostas fornecidas aos questionários neste estudo, bem como (iii) a criterialidade indentificada na preparação dos dados para a escala de mensuração. As variáveis linguísticas e seus respectivos valores difusos foram definidas por Chen (2000) e consistem em:

- VH: extremamente importante (0,9 – 1,0 – 1,0),

- H: muito importante (0,7 – 0,9 – 1,0),
- MH: importante (0,5 – 0,7 – 0,9),
- M: medianamente importante (0,3 – 0,5 – 0,7),
- ML: medianamente pouco importante (0,1 – 0,3 – 0,5),
- L: pouco importante (0 – 0,1 – 0,3),
- VL: muito pouco importante (0 – 0 – 0,1).

A variação no emprego das variáveis foi realizada considerando-se a frequência de menção aos critérios nos levantamentos realizados – considerando-se uma variação de 15% entre eles – atrelada à criterialidade destes para o teste oral do EPPL. Os resultados obtidos encontram-se resumidos no Quadro 35.

Quadro 35. Definição dos pesos dos critérios de avaliação do teste oral do EPPL.

	Crítérios	Literatura	Usuários	Criterialidade	Peso final	
1	Uso de linguagem geral	H	H	VH	H	0.7 – 0.9 – 1.0
2	Uso de linguagem de sala de aula	VH	VH	VH	VH	0.9 – 1.0 – 1.0
3	Precisão linguística	VH	L	VH	MH	0.5 – 0.7 – 0.9
4	Complexidade linguística	VL	VL	VH	M	0.3 – 0.5 – 0.7
5	Fluência	VH	VH	VH	VH	0.9 – 1.0 – 1.0
6	Inteligibilidade	VH	VH	VH	VH	0.9 – 1.0 – 1.0
7	Flexibilidade	VH	H	H	H	0.7 – 0.9 – 1.0

Fonte: Elaborado pela autora.

Por trabalhar com uma escala de mensuração de orientação analítica e objetivar a definição de padrões de desempenho de uso da língua sobre os quais seja possível a realização de inferências, a abordagem avaliativa proposta requer que, inicialmente, os dados de fala coletados por meio das aplicações do teste oral do exame passem por uma fase de tratamento por especialistas para definição dos atributos da competência linguístico-comunicativo-pedagógica dos candidatos. Assim, requer-se a transcrição desses dados e a realização de cálculos e categorizações dos desempenhos.

De posse dessas informações, é necessário que o especialista prossiga para a atribuição de variáveis linguísticas, nos moldes definidos por Chen (2000), ao desempenho analisado em cada critério diante dos valores informados na escala de mensuração. Tais variáveis consistem em:

- VG: muito bom (9 – 10 – 10),
- G: bom (7 – 9 – 10),
- MG: medianamente bom (5 – 7 – 9),
- F: razoável (3 – 5 – 7),
- MP: medianamente insuficiente (1 – 3 – 5),

- P: insuficiente (0 – 1 – 3),
- VP: muito insuficiente (0 – 0 – 1).

Diante da existência de limites superiores e inferiores de variação de desempenho dentro das faixas de mensuração propostas e da necessidade de padronização da avaliação, com minimização de interferências geradas por enviesamentos individuais dos examinadores, propõe-se que as variáveis linguísticas representem as possíveis variações presentes em cada faixa com relação à mediana dos valores que as compõem. Deste modo, sugere-se que seja considerado VG desempenhos que representem valores que estejam dentro ou acima da mediana da escala de mensuração para o limiar máximo de variação do critério na faixa A; G, desempenhos dentro da mediana do limiar máximo de variação na faixa B; MG, desempenhos um pouco abaixo da mediana do limiar de variação da faixa B, mas acima do limiar da faixa C; F, desempenhos dentro da mediana do limiar de variação da faixa C; MP, desempenhos levemente abaixo da mediana do limiar mínimo da faixa C, mas acima da faixa D; P, desempenhos dentro ou abaixo da mediana do limiar mínimo de variação da faixa D; e VP, desempenhos que não sejam passíveis de avaliação. O Quadro 36 resume essas sugestões.

Quadro 36. Variáveis linguísticas associadas às faixas de proficiência.

Níveis	Limiares	Variável linguística	
A	limiar superior	1.0	VG
	mediana	0.9:1.0	VG
	limiar inferior	0.8:0.9	G
B	limiar superior	0.7:0.8	G
	mediana	0.6:0.7	MG
	limiar inferior	0.5:0.6	MG
C	limiar superior	0.4:0.5	F
	mediana	0.3:0.4	F
	limiar inferior	0.2:0.3	MP
D	limiar superior	0.1:0.2	MP
	mediana	0.1	P
	limiar inferior	0.0	P
E	Desempenho insuficiente	0.0	VP

Fonte: Elaborado pela autora.

O desenho padronizado e possibilitado pelos valores do Quadro 39 das matrizes de insumo de comparação de cada critério em relação às alternativas que deve ser empregado para o cálculo da classificação do candidato encontra-se expresso nas tabelas 23, 24, 25, 26, 27, 28, 29, 30, 31, 32 e 33.

Tabela 23. Mediana ou acima do limiar de A

CRITÉRIO X							
	VP	P	MP	F	MG	G	VG
Alternativa 1	0	0	0	0	0	0	1
Alternativa 2	1	0	0	0	0	0	0
Alternativa 3	1	0	0	0	0	0	0
Alternativa 4	1	0	0	0	0	0	0
Alternativa 5	1	0	0	0	0	0	0

Fonte: Elaborado pela autora.

Tabela 24. Limiar inferior de A

CRITÉRIO X							
	VP	P	MP	F	MG	G	VG
Alternativa 1	0	0	0	0	0	1	0
Alternativa 2	0	0	0	0	0	0	1
Alternativa 3	1	0	0	0	0	0	0
Alternativa 4	1	0	0	0	0	0	0
Alternativa 5	1	0	0	0	0	0	0

Fonte: Elaborado pela autora.

Tabela 25. Limiar superior de B

CRITÉRIO X							
	VP	P	MP	F	MG	G	VG
Alternativa 1	0	0	0	0	1	0	0
Alternativa 2	0	0	0	0	0	0	1
Alternativa 3	1	0	0	0	0	0	0
Alternativa 4	1	0	0	0	0	0	0
Alternativa 5	1	0	0	0	0	0	0

Fonte: Elaborado pela autora.

Tabela 26. Mediana de B

CRITÉRIO X							
	VP	P	MP	F	MG	G	VG
Alternativa 1	1	0	0	0	0	0	0
Alternativa 2	0	0	0	0	0	0	1
Alternativa 3	1	0	0	0	0	0	0
Alternativa 4	1	0	0	0	0	0	0
Alternativa 5	1	0	0	0	0	0	0

Fonte: Elaborado pela autora.

Tabela 27. Limiar inferior de B

CRITÉRIO X							
	VP	P	MP	F	MG	G	VG
Alternativa 1	1	0	0	0	0	0	0
Alternativa 2	0	0	0	0	0	0	1
Alternativa 3	0	0	0	0	0	1	0
Alternativa 4	1	0	0	0	0	0	0
Alternativa 5	1	0	0	0	0	0	0

Fonte: Elaborado pela autora.

Tabela 28. Limiar superior de C

CRITÉRIO X							
	VP	P	MP	F	MG	G	VG
Alternativa 1	1	0	0	0	0	0	0
Alternativa 2	0	0	0	0	0	1	0
Alternativa 3	0	0	0	0	0	0	1
Alternativa 4	1	0	0	0	0	0	0
Alternativa 5	1	0	0	0	0	0	0

Fonte: Elaborado pela autora.

Tabela 29. Mediana de C

CRITÉRIO X							
	VP	P	MP	F	MG	G	VG
Alternativa 1	1	0	0	0	0	0	0
Alternativa 2	1	0	0	0	0	0	0
Alternativa 3	0	0	0	0	0	0	1
Alternativa 4	1	0	0	0	0	0	0
Alternativa 5	1	0	0	0	0	0	0

Fonte: Elaborado pela autora.

Tabela 30. Limiar inferior de C

CRITÉRIO X							
	VP	P	MP	F	MG	G	VG
Alternativa 1	1	0	0	0	0	0	0
Alternativa 2	1	0	0	0	0	0	0
Alternativa 3	0	0	0	0	0	0	1
Alternativa 4	0	0	0	0	0	1	0
Alternativa 5	1	0	0	0	0	0	0

Fonte: Elaborado pela autora.

Tabela 31. Limiar superior de D

CRITÉRIO X							
	VP	P	MP	F	MG	G	VG
Alternativa 1	1	0	0	0	0	0	0
Alternativa 2	1	0	0	0	0	0	0
Alternativa 3	0	0	0	0	0	1	0
Alternativa 4	0	0	0	0	0	0	1
Alternativa 5	1	0	0	0	0	0	0

Fonte: Elaborado pela autora.

Tabela 32. Mediana ou inferior ao limiar de D

CRITÉRIO X							
	VP	P	MP	F	MG	G	VG
Alternativa 1	1	0	0	0	0	0	0
Alternativa 2	1	0	0	0	0	0	0
Alternativa 3	1	0	0	0	0	0	0
Alternativa 4	0	0	0	0	0	0	1
Alternativa 5	1	0	0	0	0	0	0

Fonte: Elaborado pela autora.

Tabela 33. Desempenho insuficiente

CRITÉRIO X							
	VP	P	MP	F	MG	G	VG
Alternativa 1	1	0	0	0	0	0	0
Alternativa 2	1	0	0	0	0	0	0
Alternativa 3	1	0	0	0	0	0	0
Alternativa 4	1	0	0	0	0	0	0
Alternativa 5	0	0	0	0	0	0	1

Fonte: Elaborado pela autora.

4.3.3 Os resultados da reavaliação dos desempenhos de fala

A proposta de aplicação do método *Fuzzy* TOPSIS neste trabalho garante que o processo de avaliação dos desempenhos seja realizado de maneira semi-automática. Isso porque, a presença do examinador humano é requerida, na condição de especialista, na etapa de preparação prévia dos dados – transcrição, contagem e preparação dos perfis de desempenho para cada critério em relação à escala de mensuração. No entanto, a decisão de classificação do candidato é informada pelo método, e não pelo examinador em si.

A reavaliação dos desempenhos de fala foi realizada para aproximadamente 10% (7) da amostra. A seleção dos dados ocorreu de forma intencional, visando-se à obtenção de perfis de candidatos em que a amostra apresentava perfis compatíveis com o desempenho de uso de linguagem pedagógica e com inconsistências com relação à avaliação desse uso. O desempenho no uso da linguagem pedagógica foi escolhido como critério de seleção de dados para reavaliação em virtude de ser esse o aspecto que diferencia o exame EPPL de exames com abordagens mais gerais. Desse modo, foram selecionados para reavaliação os candidatos LID, DES, LIG, MAC, THY, ROB e LAC, conforme exibido no Quadro 37.

Quadro 37. Seleção de candidatos para reavaliação.

Candidato	Faixa original	Motivação
LID	B	Inconsistência no desempenho de metalinguagem: o candidato não demonstra conhecimento de conteúdo suficiente
DES	B	Consistência no desempenho de metalinguagem
LIG	C	Consistência no desempenho de metalinguagem
MAC	D	Consistência no desempenho de metalinguagem
THY	D	Inconsistência no desempenho de metalinguagem: o candidato demonstra desempenho superior
ROB	E	Consistência no desempenho de metalinguagem
LAC	A	Consistência no desempenho de metalinguagem

Fonte: Elaborado pela autora.

O candidato LID foi inserido na amostra porque seu perfil de desempenho apresentava inconsistências graves em relação ao uso da linguagem pedagógica. Embora na tarefa de fornecimento de instruções o candidato tenha apresentado um desempenho bastante satisfatório, fazendo uso de cinco funções comunicativas de maneira apropriada, incluindo-se o emprego de municiação linguística; na tarefa de explicação gramatical o candidato não foi capaz de identificar apropriadamente os erros apresentados, incorrendo em um erro de conteúdo grave, conforme exibido pelo excerto 32. O desempenho geral do referido candidato encontra-se resumido no Quadro 38.

Excerto 32:

- *LID: so: (.) I chose [: choose] [*] situation B .
- *LID: &-um what do you have to do here ?
- *LID: read letter A and then letter B .
- *LID: "she might be away (.) you never know with her" .
- *LID: in this situation you are using the wrong preposition .
- *LID: so "with" and "her" is [: are] [*] not in the context .
- *LID: you never know (.) about her .
- *LID: you would never know about her .
- *LID: you should use (.) another expression (.) to: (.) mean what you (.) want to say .
- *LID: letter B is better .
- *LID: "she must be in Salvador (.) you never know" .
- *LID: "you never know" because you do not know where (.) this person is where Mary is .
- *LID: so: (.) in letter A (.) "you never know with her" (.) the preposition (.) is being used wrongly .
- *LID: while in (.) letter B (.) "you never know" (.) is better .
- *LID: because you do not have a preposition .
- *LID: and you do not have (.) an undirect [: indirect] [*] object in your way .

Candidato LID. Faixa B. EPPLÉ, Tarefa 5. Versão 3.1, 2015.

Quadro 38. Perfil de desempenho de LID.

LID			
Crítérios	Variáveis		
Uso de linguagem geral	MG		
	5	7	9
Uso de linguagem de sala de aula	MP		
	1	3	5
Precisão linguística	MG		
	5	7	9
Complexidade linguística	G		
	7	9	9
Inteligibilidade	G		
	7	9	9
Fluência	G		
	7	9	9
Flexibilidade	G		
	7	9	9
Ranking de Alternativas			
A	0.1313	4	
B	0.4384	1	
C	0.1384	3	
D	0.0476	5	
E	0.1706	2	

Fonte: Elaborado pela autora.

O desempenho de fala de LID corrobora a crença segundo a qual o domínio de uso da língua para finalidades gerais não é suficiente para atuação voltada ao ensino de línguas, uma vez que, conforme ressalta Douglas (2000), no desempenho em contextos específicos, o conhecimento de tópico exerce interferências importante. No entanto, conforme ressaltam Farrell e Richards (2007), diante da relação direta existente entre o conhecimento de tópico e a capacidade de uso da língua para finalidade de ensino, é importante que tal aspecto sejam considerados pela avaliação.

O candidato THY, por sua vez, foi selecionado em virtude de haver uma discrepância expressiva entre a faixa a ele atribuída (D) e seu desempenho de uso de língua tanto para finalidades gerais quanto específicas, conforme exibido pelo Quadro 39. Com relação à tarefa de explicação gramatical, mais especificamente, o excerto 33 exibe que o candidato foi capaz de identificar e corrigir o erro de maneira apropriada, apesar de não fazer uso de terminologia específica em sua explicação – o que, conforme já apontado por este estudo, não consistia em uma exigência da tarefa.

Excerto 33:

*THY: &-um situation B .
 *THY: "I (ha)ve written to Maria but she has n(o)t replied".
 *THY: I think that the best (.) alternative is A .
 *THY: because "might" is less strong than "must" .
 *THY: "she must be" (.) you are strongly affirming (.) that she is in Salvador .
 *THY: and (.) you do n(o)t know if she is in Salvador .
 *THY: she probably did n(o)t answer for some other reason that you do n(o)t know .
 *THY: so "might be" (.) I think it (i)s (.) a better option than "must be" .
 THY: unless that you (.) have a previously [: previous] [] information that she has (.) traveled to: a place .
 *THY: but you do n(o)t know (.) which place .
 *THY: an:d "you never know with her" .
 *THY: I guess it (i)s <a &w> [//] a: way to: (.) affirm the (.) previous sentence .

Candidato THY. Faixa D. EPPLÉ, Tarefa 5. Versão 3.1, 2015.

Quadro 39. Perfil de desempenho de THY.

THY			
Critérios	Variáveis		
Uso de linguagem geral	MG		
	5	7	9
Uso de linguagem de sala de aula	MG		
	5	7	9

<i>Precisão linguística</i>	G		
	7	9	9
<i>Complexidade linguística</i>	F		
	3	5	7
<i>Inteligibilidade</i>	G		
	7	9	9
<i>Fluência</i>	MG		
	5	7	9
<i>Flexibilidade</i>	G		
	7	9	9
Ranking de Alternativas			
A	0.2743	3	
B	0.6023	1	
C	0.3616	2	
D	0.0476	4	
E	0.0476	4	

Fonte: Elaborado pela autora.

Os candidatos DES, LIG, MAC, ROB e LAC, por fim, cada um representante de uma faixa diferente da escala original – respectiva e originalmente B, C, D, E e A – foram selecionados em virtude da existência de consistências entre a avaliação realizada pelos examinadores e seus perfis de desempenho. Os quadros 40, 41, 42, 43 e 44 resumem as características de seus desempenhos de fala.

Quadro 40. Perfil de desempenho de DES.

DES			
Critérios	Variáveis		
Uso de linguagem geral	G		
	7	9	9
	VG		
Uso de linguagem de sala de aula	9	10	10
	VG		
	9	10	10
Precisão linguística	VG		
	9	10	10
	MG		
Complexidade linguística	5	7	9
	G		
	7	9	9
Fluência	MG		
	5	7	9
	G		
Flexibilidade	7	9	9
	Ranking de Alternativas		
	A	0.1528	3
B	0.6398	1	
C	0.1724	2	
D	0.0476	4	
E	0.0476	4	

Fonte: Elaborado pela autora.

Quadro 41. Perfil de desempenho de LIG.

LIG			
Critérios	Variáveis		
Uso de linguagem geral	F		
	3	5	7
	F		
Uso de linguagem de sala de aula	3	5	7
	MP		
	1	3	5
Complexidade linguística	MP		
	1	3	5
	F		
Inteligibilidade	3	5	7
	F		
	3	5	7
Fluência	F		
	3	5	7
	G		
Flexibilidade	7	9	9
	Ranking de Alternativas		
	A	0.0476	4
B	0.1141	3	
C	0.5789	1	
D	0.1724	2	
E	0.0476	4	

Fonte: Elaborado pela autora.

Quadro 42. Perfil de desempenho de MAC.

MAC			
Critérios	Variáveis		
Uso de linguagem geral	F		
	3	5	7
	F		
Uso de linguagem de sala de aula	3	5	7
	VG		
	9	10	10
Complexidade linguística	MG		
	5	7	9
	F		
Inteligibilidade	3	5	7
	MP		
	1	3	5
Fluência	G		
	7	9	9
	Ranking de Alternativas		
A	0.1730	2	
B	0.1564	3	
C	0.4847	1	
D	0.1384	4	
E	0.0476	5	

Fonte: Elaborado pela autora.

Quadro 43. Perfil de desempenho de ROB.

ROB			
Critérios	Variáveis		
Uso de linguagem geral	MP		
	1	3	5
	MP		
Uso de linguagem de sala de aula	1	3	5
	P		
	0	1	3
Precisão linguística	F		
	3	5	7
	MP		
Inteligibilidade	1	3	5
	P		
	0	1	3
Fluência	P		
	0	1	3
	P		
Flexibilidade	0	1	3
	Ranking de Alternativas		
	A	0.0456	4
B	0.1364	3	
C	0.3058	2	
D	0.4999	1	
E	0.0456	4	

Fonte: Elaborado pela autora.

Quadro 44. Perfil de desempenho de LAC.

LAC			
Critérios	Variáveis		
Uso de linguagem geral	VG		
	9	10	10
	VG		
Uso de linguagem de sala de aula	9	10	10
	VG		
	9	10	10
Precisão linguística	VG		
	9	10	10
	G		
Complexidade linguística	7	9	9
	VG		
	9	10	10
Fluência	VG		
	9	10	10
	VG		
Flexibilidade	9	10	10
	Ranking de Alternativas		
	A	0.6023	1
B	0.0906	2	
C	0.0476	3	
D	0.0476	3	
E	0.0476	3	

Fonte: Elaborado pela autora.

A abordagem proposta mostrou-se adequada para 85,33% (6) dos dados reavaliados, mantendo uma consistência na avaliação dos desempenhos em relação à escala de proficiência sugerida e sendo capaz de corrigir avaliações equivocadas, como a do candidato THY. No caso deste candidato, tem-se que a avaliação de seu desempenho de fala apontava para usos medianamente satisfatórios da língua em quase todos os critérios considerados na análise. O único critério no qual o desempenho do candidato apresentava declínio, referia-se à complexidade linguística, que, na opinião dos usuários do exame, não se mostrava como um critério relevante para a proficiência do professor de línguas. A investigação analítica de seu desempenho e o tratamento da interatividade dos critérios foi capaz de prover solução ao equívoco.

No que se refere ao desempenho altamente discrepante do candidato LID, correspondente a 14,29% (1) da amostra, tem-se que o método não se mostrou capaz de solucionar a inconsistência de seu desempenho na tarefa de metalinguagem, atribuindo-lhe uma faixa muito superior para o desempenho metalinguístico do que deveria ter sido apontado. O resultado fornecido pelo método atribui à LID uma faixa correspondente a um desempenho bom (G) quando, na verdade, seu desempenho na tarefa foi medianamente ruim (P), pois o candidato não foi capaz de identificar ou explicar o erro requerido na situação da tarefa de explicação gramatical. Percebe-se, com isso, que a gravidade do possível impacto gerado por este tipo de erro de mensuração mostra-se aumentada na presença de desempenhos de uso da língua muito discrepantes nas dimensões de finalidades gerais e pedagógicas. A faixa atribuída ao candidato LID tanto na avaliação original quanto na reavaliação representa com fidelidade apenas seu desempenho no uso da língua para finalidades gerais no teste.

Tal ocorrência aponta para duas possibilidades de gerenciamento deste tipo de inconsistência com relação à avaliação da linguagem com finalidades pedagógicas no exame. A primeira, e mais simples, seria a necessidade de se separar o aspecto geral do aspecto específico na escala de mensuração, realizando-se duas avaliações individuais e reportando-se as capacidades separadamente. Outra opção, de caráter mais integrado, seria voltar as funções comunicativas do teste primordial e exclusivamente para o uso da linguagem em situações de sala de aula. Apesar de todas as tarefas analisadas relacionarem-se a tópicos referentes ao ambiente escolar, a linguagem esperada em mais de 60% não é específica desse ambiente. Aventa-se, assim, a possibilidade de remodelar as tarefas do exame, de modo que a capacidade de uso da língua para finalidades gerais seja avaliada em um número de questões inferior às voltadas ao ensino-aprendizagem de línguas ou por meio de tarefas que requeiram simulações de uso da linguagem genérica durante uma atividade pedagógica.

O Quadro 45 resume os resultados encontrados na reavaliação.

Quadro 45. Quadro-resumo dos resultados da avaliação original e da reavaliação dos desempenhos de fala.

Candidato	Faixa original	Nova classificação	Análise
LID	B	B	<p>Não corrige a inconsistência na avaliação do desempenho de metalinguagem: o candidato não demonstra conhecimento de conteúdo suficiente.</p> <p><i>B3. Utiliza linguagem de sala de aula consistentemente apropriada. Consegue empregar adequadamente uma variedade de terminologia para descrição de língua.</i></p> <p><i>É capaz de reconhecer quase todos os erros cometidos por alunos e de fornecer exemplos adequados de assuntos linguísticos dos níveis A1, A2, B1 e B2 do CEFR.</i></p>
DES	B	B	<p>Mantém consistência na avaliação do desempenho de metalinguagem.</p> <p><i>B3. Utiliza linguagem de sala de aula consistentemente apropriada. Consegue empregar adequadamente uma variedade de terminologia para descrição de língua.</i></p> <p><i>É capaz de reconhecer quase todos os erros cometidos por alunos e de fornecer exemplos adequados de assuntos linguísticos dos níveis A1, A2, B1 e B2 do CEFR.</i></p>
LIG	C	C	<p>Mantém consistência na avaliação do desempenho de metalinguagem.</p> <p><i>C3. Utiliza linguagem de sala de aula com algumas inadequações. Consegue empregar pouca terminologia adequada para descrição de língua.</i></p> <p><i>É capaz de reconhecer a maioria dos erros cometidos por alunos e de fornecer exemplos adequados de assuntos linguísticos dos níveis A1, A2 e B1 do CEFR.</i></p>
MAC	D	C	<p>Mantém consistência na avaliação do desempenho de metalinguagem.</p> <p><i>C3. Utiliza linguagem de sala de aula com algumas inadequações. Consegue empregar pouca terminologia adequada para descrição de língua.</i></p> <p><i>É capaz de reconhecer a maioria dos erros cometidos por alunos e de fornecer exemplos adequados de assuntos linguísticos dos níveis A1, A2 e B1 do CEFR.</i></p>
THY	D	B	<p>Corrige a inconsistência da avaliação do desempenho de metalinguagem: o candidato demonstrava desempenho superior</p> <p><i>B3. Utiliza linguagem de sala de aula consistentemente apropriada. Consegue empregar adequadamente uma variedade de terminologia para descrição de língua.</i></p> <p><i>É capaz de reconhecer quase todos os erros cometidos por alunos e de fornecer exemplos adequados de assuntos linguísticos dos níveis A1, A2, B1 e B2 do CEFR.</i></p>
ROB	E	D	<p>Mantém consistência na avaliação do desempenho de metalinguagem.</p> <p><i>D3. Utiliza linguagem de sala de aula básica com inadequações. Não consegue empregar terminologia adequada para descrição de língua.</i></p> <p><i>É capaz e reconhecer erros básicos cometidos por alunos e de fornecer exemplos adequados de assuntos linguísticos dos níveis A1 e A2 do CEFR.</i></p>
LAC	A	A	<p>Mantém consistência na avaliação do desempenho de metalinguagem.</p>

			<p>A3. <i>Utiliza ampla variedade de linguagem de sala de aula com consistência. Consegue empregar apropriadamente vocabulário técnico e expressões idiomáticas sofisticadas para descrição de língua.</i></p> <p><i>É capaz de reconhecer todos os erros cometidos por alunos e de fornecer exemplos adequados de assuntos linguísticos dos níveis A1, A2, B1, B2, C1 e C2 do CEFR.</i></p>
--	--	--	--

Fonte: Elaborado pela autora.

Desse modo, entende-se que os dados apontam para a existência de uma tendência de confiabilidade (i.) no emprego de uma escala de mensuração analítica, (ii.) no emprego do método *Fuzzy* TOPSIS para tratamento da interatividade dos critérios em relação à escala e (iii.) no uso de uma estrutura semi-automática de avaliação, na qual a presença do examinador é requisitada na condição de especialista para análise dos dados de fala, mas a classificação do desempenho do candidato é realizada modelando-se o pensamento humano em relação a uma escala padronizada de desempenhos, evitando-se, assim, discrepâncias e enviesamentos individuais por parte dos examinadores.

4.4 O capítulo em resumo

Neste capítulo foram apresentados os resultados da análise dos dados e as discussões. Inicialmente, apresentou-se uma descrição das respostas fornecidas pelos usuários do exame visando-se à busca por uma delimitação dos critérios de relevância para avaliação da proficiência do professor de línguas. Na sequência, procedeu-se à análise da abordagem de avaliação da proficiência realizada e do possibilitada pelo exame EPPL, levantando-se informações sobre seu construto, seus objetivos, seus critérios, suas potencialidades e sobre as características do desempenho de fala de seus candidatos. A partir destes dados, foi possível a geração de uma escala de mensuração e de uma escala de proficiência para o exame. O capítulo foi finalizado com a realização dos procedimentos de reavaliação semi-automática de uma amostra dos dados de fala por meio de um método multicritério de análise da decisão, o *Fuzzy* TOPSIS, cujos resultados apontaram para um emprego promissor deste tipo de metodologia na avaliação de proficiência.

5 CONSIDERAÇÕES FINAIS

Este capítulo tem por finalidade sintetizar os resultados alcançados pela investigação, respondendo, pontualmente, às perguntas de pesquisa apresentadas no capítulo introdutório. A pesquisa dedicou-se ao estudo da abordagem avaliativa da proficiência oral de professores de inglês como língua estrangeira empregado por um exame de finalidades específicas, o EPPL e teve por objetivo geral reestruturar a abordagem de avaliação do candidato em faixas de proficiência oral. Por objetivos específicos, o estudo determinou a proposição de uma escala de proficiência oral que incorporasse os interesses dos usuários (*stakeholders*), bem como as características de fala dos avaliados do exame e que estivesse alinhada a indicadores de proficiência linguística internacionais; e a proposição de uma estrutura avaliativa que potencializasse a precisão, a transparência, a praticidade e a confiabilidade dos resultados.

Para atingir os objetivos propostos, o estudo orientou a investigação por meio de três perguntas de pesquisa. A primeira pergunta tinha por objetivo realizar um levantamento dos critérios que deveriam integrar a avaliação da proficiência do professor de línguas, ou seja, preocupava-se com a definição de critérios que tivessem relevância avaliativa para determinar níveis dessa proficiência. Orientando-se pelos modelos de competência linguística de Bachman (1990) e do CEFR (COE, 2018) e pelos modelos de competência linguístico-comunicativa de Almeida Filho (1993) e do CETF (CAE, 2018), foi realizada a análise da opinião de usuários do exame EPPL, obtida por meio da aplicação de questionários, a esse respeito. Princípios extraídos do método de estruturação de problemas VFT – *Value-Focused Thinking* (KEENEY, 1992, 1994, 1996) orientaram a elaboração dos questionários e a análise dos dados, que resultou em uma hierarquização das preferências.

A investigação voltou-se, então, para a segunda pergunta de pesquisa, que tinha por objetivo determinar, dos critérios elencados por meio da primeira questão, quais eram passíveis, quantitativa ou qualitativamente de mensuração. Das técnicas de mensuração de aspectos relativos à fala em língua estrangeira levantadas por meio de análise da literatura em avaliação de proficiência e presentes no questionário aplicado aos usuários do exame, pôde-se prosseguir para a terceira pergunta, que inquiria a respeito de como mensurar a proficiência linguístico-comunicativo-pedagógica do professor de línguas. Nesta etapa, a pesquisa preocupava-se com a garantia dos princípios de validade, confiabilidade e praticidade das avaliações, bem como

com a agregação dos valores do grupo de usuários do exame sobre os critérios, visando adequar os resultados da avaliação às exigências contextualmente democráticas impostas pelo momento sócio-histórico e, mais especificamente, por seus usuários, garantindo maior transparência e apoio às decisões avaliativas.

Percebeu-se, portanto, a necessidade de se elaborar uma escala de mensuração com finalidade de orientar a avaliação dos desempenhos de fala separada da escala de proficiência com finalidade informativa. Isso porque, verificou-se que os critérios relacionados à organização textual eram dependentes das tarefas do exame, ou seja, eram limitados ao contexto da tarefa e, como tais, não deveriam integrar a escala que tem finalidade inferencial para situações de uso da língua no mundo real. A escala de mensuração foi elaborada por meio da opinião dos usuários, mas também, por meio da criterialidade investigada nos desempenhos potencializados pelas tarefas e nos desempenhos reais dos usuários em cada tarefa. A escala de proficiência foi desenvolvida paralelamente, considerando-se essas variações e os princípios do CEFR (COE, 2018) e CETF (CAE, 2018).

O tratamento da interatividade dos critérios foi realizado por meio da aplicação do método multicritério de análise da decisão *Fuzzy TOPSIS* a aproximadamente 10% dos dados de fala que compõem a investigação. As amostras foram selecionadas por julgamento, ou seja, os candidatos foram selecionados intencionalmente visando à avaliação da capacidade de o método manter avaliações desejáveis e corrigir inconsistências, especialmente com relação ao desempenho dos candidatos nas tarefas de uso da língua para finalidades pedagógicas. Os resultados da reavaliação dos dados de fala revelaram consistência para mais de 80% dos candidatos – havendo, inclusive, a correção de uma avaliação originalmente inconsistente – em relação à escala de proficiência proposta, apontando para a existência de força preditiva positiva por parte dos critérios e dos valores de desempenho que integram a escala de mensuração.

Apesar da necessidade de preparação prévia e, em grande parte, manual dos dados, o que acaba por aumentar o tempo para fornecimento de resultados aos candidatos, é válido destacar que essa preparação garante que a decisão do avaliador seja mais informada, além de garantir uma disponibilidade maior dos dados para pesquisas que alimentem melhorias contínuas no próprio exame. Além disso, a abordagem avaliativa proposta neste trabalho provê níveis maiores de validade dos resultados das avaliações do desempenho de candidatos no exame EPPL ao propiciar maior consideração de variações de desempenho em seu modelo de mensuração e oferecer escalas de avaliação em acordo com as tarefas do exame e com os descritores de proficiência internacionalmente reconhecidos, garantindo, ainda, a minimização de ocorrências de inconsistências graças à informatividade das escalas e à consideração do peso

de preferências de outros decisores, e maior transparência na divulgação dos resultados. Além disso, propicia uma maior compreensão do conceito de proficiência-linguístico-pedagógica graças à análise dos dados de fala dos candidatos e da redefinição dos critérios que o integram.

Deste modo, espera-se que a abordagem de avaliação proposta represente um avanço na integração das opiniões de *stakeholders* do exame e na busca pela (semi-)automatização da correção de dados de fala.

Identifica-se como limitações metodológicas desta pesquisa a quantidade de amostras extraídas do corpus do teste oral do exame EPPL e consideradas na análise deste estudo, as quais, apesar de significativas, impedem a realização de generalizações confiáveis não apenas pelo número total de participantes, mas também por representarem apenas contextos de formação de professores privilegiados; e a quantidade e a variedade alcançada de participantes respondentes dos questionários, que, novamente, impedem a realização de generalizações acuradas e têm suas representatividades concentradas nas regiões Sul e Sudeste do Brasil. Além disso, pode-se questionar a orientação teórico-metodológica dos modos e das unidades de avaliação empregados com relação aos critérios. Isso porque, apesar de as unidades de mensuração utilizadas terem sua eficácia nacional e internacionalmente atestadas na literatura em Linguística Aplicada – como exibido Brown, Iwashita e McNamara (2005), Rimmer (2009), Lim e Galaczi (2010), Manosso (2013), De Jong *et al.* (2013) para citar alguns –, não há consenso a respeito dos melhores indicadores para a proficiência em língua estrangeira de brasileiros.

Desse modo, visando ao aprimoramento das contribuições que podem ser extraídas da realização deste trabalho, sugere-se que a escala de mensuração proposta seja alimentada com uma quantidade maior de dados do corpus do teste oral do exame EPPL de modo a se ampliar a confiabilidade dos indicadores de perfis de desempenho e a capacidade de generalização desses indicadores. Igualmente, sugere-se que as opiniões de um número maior e mais diversificado de usuários do exame seja investigada, evitando-se possíveis enviesamentos avaliativos oriundos de contextos específicos e regionalizados de pesquisas em avaliação. Sugere-se, ainda, conforme apontado pelos resultados da reavaliação dos dados de fala por meio do método multicritério de análise da decisão a revisão dos modelos de tarefas que compõem o teste oral do exame EPPL diante da existência de um foco pronunciado dessas tarefas em situações de uso da língua estrangeira para finalidades gerais em detrimento de situações de uso especificamente pedagógicas, o que faz com que desempenhos de fala pedagógica muito discrepantes da fala geral tenham sua importância reduzida. Sugere-se, por fim, que a aplicação

do método vertex na agregação dos pesos dos decisores de modo a se verificar possíveis diferenças no ordenamento sugerido.

Sugere-se, por fim, o prosseguimento da investigação do processo de validação do exame, especialmente considerado-se a escala de proficiência proposta, segundo os princípios determinados pelo Conselho Britânico (COE, 2009), a saber: a determinação da frequência de variação da classificação dos itens em relação ao CEFR, a determinação de consistência intra-avaliador e o nível de discordância por item.

REFERÊNCIAS

- ALDERSON, J. C. Bands and scores. *In*: ALDERSON, J. C.; NORTH, B. **Language testing in the 1990s: the communicative legacy**. London: Modern English Publications, British Council, 1991.
- ALMEIDA FILHO, J. C. P. **Dimensões comunicativas no ensino de línguas**. Pontes: Campinas-SP, 1993.
- ALMEIDA FILHO, J. C. P. O professor de língua estrangeira sabe a língua que ensina? A questão da instrumentalização linguística. **Contexturas**, v. 1, n. 1, p. 77-85, 1992.
- ALMEIDA, A. T.; COSTA, A. P. C. S. **Aplicações com métodos de apoio multicritério à decisão**. Recife: Editora Universitária da UFPE, 2003.
- ALMEIDA, S.; MORAIS, D. C.; ALMEIDA, A. T. Agregação de pontos de vista de stakeholders utilizando o Value-Focused Thinking associado a mapeamento cognitivo. **Production**, v. 24, n. 1, p. 144-159, 2014.
- AMIRIAN, S. M. R.; MOQADDAM, H. H.; MOQADDAM, Q. J. Critical Analysis of the models of language proficiency with a focus on communicative models. **Theory and Practice in Language Studies**, v. 7, n. 5, p. 400-407, 2017.
- ANDRELINO, P. J. **Análise da estrutura genérica das instruções na fala do professor de inglês: contribuições para o teste oral do EPPE**. 2014. 359 f. Tese (doutorado em Estudos Linguísticos) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, 2014.
- ANG-AW, H. T.; GOH, C. C. M. Understanding discrepancies in rater judgement on national-level oral examination tasks. **RELC Journal**, v. 42, n. 1, p. 31-51, 2011.
- ARESE, M. C.; RANGEL, L. A. D.; HALL, J.; ZOTES, L. P.; BONINA, N.; MEIRIÑO, M. J. Aplicação do método TOPSIS na avaliação dos critérios utilizados na seleção de docentes em uma instituição de ensino superior. **Conhecimento e Diversidade**, v. 9, n. 19, p. 47-58, 2017.
- AYRES, M.; AYRES JR., M.; AYRES, D. L.; SANTOS, A. A. S. **Bioestat: aplicações estatísticas nas áreas das ciências biomédicas**. Belém, 2007.
- BACHMAN, L. F. **Fundamental considerations in language testing**. Oxford: OUP, 1990.
- BACHMAN, L. F.; LYNCH, B. K.; MASON, M. Investigating variability in tasks and rater judgments in a performance test of foreign language speaking. **Language Testing**, v. 12, p. 238-257, 1995.
- BACHMAN, L. F.; PALMER, A. S. **Language assessment in practice**. Oxford: OUP, 1996.
- 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.

366 f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, 2010.

BAHABADI, M. D. **Fuzzy TOPSIS Software** – Chen (2000). [S.l.], 2018. Disponível em: https://www.researchgate.net/publication/327468165_fuzzy_topsis_software_-Chen_2000. Acesso em 10 jun. 2019.

BANA E COSTA, C. A. **Structuration, construction et exploitation d’un modèle multicritère d’aide à la décision**. Tese (Doutorado em Engenharia de Sistemas). 1992 – Universidade Técnica de Lisboa, Portugal, 1992.

BARBOSA, S. M. A. D. **Perfis variados de competência linguístico-comunicativa numa LE (inglês) e seu impacto no ensino de línguas**. 2007. 153 f. Dissertação (Mestrado em Linguística Aplicada) – Universidade de Brasília, Instituto de Letras, Brasília, 2007.

BEHZADIAN, M.; OTAGHSARA, S. K.; YAZDANI, M.; IGNATIUS, J. A state-of-the-art survey of TOPSIS applications. **Expert Systems with Applications**, v. 39, p. 13051-13069, 2012.

BORGER, L. **Looking beyond scores: A study of rater orientations and ratings of speaking**. 2014. 177 f. Licenciante thesis (Licenciante degree in Subject Matter Education) – University of Gothenburg, Faculty of Education, Gothenburg, 2014.

BORGES-ALMEIDA, V. **Precisão e complexidade gramatical na avaliação de proficiência oral em inglês do formando em Letras: implicações para a validação de um teste**. 2009. 272 f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, 2009.

BORGES-ALMEIDA, V.; CONSOLO, D. A. Testing oral proficiency: What does pronunciation tell us? **Ilha do Desterro**, n. 60, p. 247-275, 2011.

BRAGA, S. R. P. **Avaliação de proficiência oral nos cursos de Letras: a visão de professores e alunos**. 2013. 167 f. Dissertação (Mestrado em Linguística Aplicada) – Universidade de Brasília, Instituto de Letras, Brasília-DF, 2013.

BROWN, J. D. Open-response items in questionnaires. *In*: HEIGHAM, J.; CROCKER, R. (ed.). **Qualitative research in Applied Linguistics: A practical introduction**. New York: Palgrave Macmillan, 2009.

BROWN, A. An investigation of the rating process in the IELTS Speaking Module. *In*: TULLOH, R. (ed.). **Research reports** – Vol. 3. Canberra, Australia: IELTS Australia, 2000.

BOWN, A.; IWASHITA, N.; MCNAMARA, T. An examination of rater orientations and test-taker performance of English-for-Academic-Purposes speaking tasks. **Monograph Series**. ETS, 2005.

BUEDE, D. M. Decision-making and decision analysis. *In*: GASS, S. I.; FU, M. C. (ed.). **Encyclopedia of operations research and management science**. Boston: Springer, 2013.

- CAE – Cambridge Assessment English. **Framework Components**. Cambridge: CET, 2018.
- CAI, H. Weight-based classification of raters and rater cognition in an EFL speaking test. **Language Assessment Quarterly**, v. 12, p. 262–282, 2015.
- 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.
- CARROLL, J. B. Fundamental considerations in testing for English language proficiency of foreign students. *In*: ALLEN, H.B. (ed.). **Teaching English as a Second Language**. New York: McGraw Hill, 1961.
- CARVALHO E SILVA, A. L. B. A Análise de necessidades e sua relação com cursos e avaliação de proficiência em inglês para um grupo de pilotos militares brasileiros. **Revista Estudos Linguísticos**, v. 45, n. 2, p. 490-503, 2016.
- CASANAVE, C. P. Case studies. *In*: PALTRIDGE, B.; PHAKITI, A. **Continuum companion to research methods in Applied Linguistics**. London: Continuum International Publishing Group, 2010.
- CERTO, S. C. Tomada de decisões. *In*: CERTO, S. C. **Administração moderna**. São Paulo: Pearson, 2005.
- CHAMBLESS, K. S. Teacher's oral proficiency in the target language: Research on its role in language teaching and learning. **Foreign language annals**, v. 45, n. S1, p. S141-S162, 2012.
- CHAPELLE, C. A. Validity argument for language assessment: The framework is simple... **Language Testing**, v. 29, p. 19–27, 2012.
- CHEN, C. T. Extension of the TOPSIS for group decision-making under fuzzy environment. **Fuzzy Sets and Systems**, v. 114, p. 1-9, 2000.
- CHOMSKY, N. **Aspects of the theory of syntax**. Cambridge: MIT Press, 1965.
- CLEMEN, R. T.; REILLY, T. **Making hard decisions with Decision Tools®**. Mason: Cengage Learning, 2013.
- CLAPHAM, C. The Assessment of metalinguistic knowledge. *In*: ELDER, C. *et al.* **Experimenting with uncertainty: Essays in honor of Alan Davies**. Cambridge: CUP, 2001.
- COE – Council of Europe. **Common European Framework of Reference for Languages: learning, teaching and assessment**. Strasbourg: Language Policy Division, 2001.
- COE – Council of Europe. **Relating language examinations to the Common European Framework of Reference for Languages: learning, teaching and assessment (CEFR)**. Cambridge: Cambridge University Press, 2018.
- COLOMBO, C. S. **O insumo linguístico oral em aulas de inglês como língua estrangeira para crianças: a fala do professor em foco**. 2014. 226 f. Dissertação (Mestrado em Estudos

Linguísticos) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, 2014.

CONSOLO, D. A. A competência oral de professores de língua estrangeira: A relação teoria-prática no contexto brasileiro. *In*: CONSOLO, D. A.; TEIXEIRA DA SILVA, V. L. (ed.). **Olhares sobre competências do professor de língua estrangeira: da formação ao desempenho profissional**. São José do Rio Preto: Editora HN, 2007.

CONSOLO, D. A. A construção de um instrumento de avaliação oral da proficiência do professor de língua estrangeira. **Trabalhos em Linguística Aplicada**, 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. **Perspectivas em avaliação no ensino e na aprendizagem de línguas: pesquisas e encaminhamentos na formação e na prática docente**. Campinas-SP: Pontes, 2017.

CONSOLO, D. A. **Avaliação da proficiência linguístico-comunicativo-pedagógica do professor de línguas**: operacionalização de construto no Exame de Proficiência para Professores de Língua Estrangeira (EPPL). Projeto trienal de pesquisa – ETAPA I. São José do Rio Preto: UNESP, 2011a.

CONSOLO, D. A. **Avaliação da proficiência linguístico-comunicativo-pedagógica do professor de línguas**: operacionalização de construto no Exame de Proficiência para Professores de Língua Estrangeira (EPPL). Projeto trienal de pesquisa – ETAPA II. São José do Rio Preto: UNESP, 2011b.

CONSOLO, D. A. **Exame de Proficiência para Professores de Língua Estrangeira (EPPL)**: definição de construto, tarefas e parâmetros para avaliação em contextos brasileiros. Projeto Trienal de Pesquisa. São José do Rio Preto: UNESP, 2008.

CONSOLO, D. A. Revendo a oralidade no ensino/aprendizagem de línguas estrangeiras. **Revista de Estudos Universitários**, v. 26, n. 1, p. 59-68, 2000.

CONSOLO, D. A.; PORTO, C. F. C. Competências do professor no processo de ensino-aprendizagem de língua estrangeira. **Horizontes de Linguística Aplicada**, a. 10, n. 2, p. 65-86, 2011.

CONSOLO, D. A.; TEIXEIRA DA SILVA, V. L. Em defesa de uma formação linguística de qualidade para professores de línguas estrangeiras: o exame EPPL. **Horizontes de Linguística Aplicada**, a. 13, n. 1, p. 63-87, 2014.

CONSOLO, D. A.; TEIXEIRA DA SILVA, V. L. Foreign language teacher's proficiency: The implementation of the EPPL examination in Brazil. *In*: GITSAKI, C.; COOMBE, C. **Current issues in language evaluation, assessment and testing: Research and practice**. Cambridge: Cambridge Scholars Publishing, 2016.

CONSOLO, D. A.; TEIXEIRA DA SILVA, V. L. **Olhares sobre competências do professor de língua estrangeira**: da formação ao desempenho profissional. São José do Rio Preto: Editora HN, 2007a.

CONSOLO, D. A.; TEIXEIRA DA SILVA, V. L. The TEPOLI test: Construct, updated tasks and new parameters to assess EFL teachers' oral proficiency. In: CONGRESSO INTERNACIONAL DA ABRAPUI, 1, 2007, Belo Horizonte. **Anais [...]**. Belo Horizonte: Universidade Federal de Minas Gerais, 2007b. p. 1-12.

CUOGHI, K. G. **Avaliação das alternativas na construção da Usina Hidrelétrica Belo Monte para aplicação de métodos multicritério de análise da decisão**. 2015. 74 f. Dissertação (Mestrado) – Faculdade de Economia, Contabilidade e Administração de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2015.

CUOGHI, K. G.; LEONETI, A. B. A group MCDA method for aiding decision-making of complex problems in public sector: The case of Belo Monte Dam. **Socio-Economic Planning Sciences**, p. 1-29, 2018. DOI: <https://doi.org/10.1016/j.seps.2018.04.002>.

DAMASIO, A. Neuroscience and the emergence of Neuroeconomics. In: GLIMCHER, P. W.; FEHR, E. (ed.). **Neuroeconomics**: Decision making and the brain. Cambridge, Massachusetts: Academic Press, 2014.

DAVIES, A. Ethics, professionalism, rights and codes. In: SHOHAMY, E.; OR, I. G.; MAY, S. **Language Testing and Assessment**. Cham: Springer International Publishing AG, 2017.

DE JONG, N. Fluency in second language testing: insights from different disciplines. **Language Assessment Quarterly**, v. 15, n. 3, p. 237-254, 2018.

DIAS, E. **Avaliação**: meio para intervenção pedagógica na formação do professor de língua estrangeira e (re)construção da sala de aula. 2011. 278 f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, 2011.

DIAS, L.; CLIMACO, J. ELECTRE TRI for groups with imprecise information on parameter values. **Group Decision and Negotiation**, v. 9, n. 5, p. 355-377, 2000.

DOUGLAS, D. **Assessing languages for specific purposes**. Cambridge: CUP, 2000.

DOUGLAS, D. Performance consistency in second language acquisition and language testing research: A conceptual gap. **Second Language Research**, v. 17, n. 4, p. 442-456, 2001.

DOUGLAS, D. **Understanding language testing**. New York: Routledge, 2010.

ECKES, T. Operational rater types in writing assessment: Linking rater cognition to rater behavior. **Language Assessment Quarterly**, v. 9, p. 270–292, 2012.

ECKES, T. Rater types in writing performance assessments: A classification approach to rater variability. **Language Testing**, v. 25, p. 155–185, 2008.

EDEN, C.; SIMS, D. Management science process – Subjectivity in problem identification. **Interface**, v. 11, n. 1, p. 68-74, 1979.

ELDER, C. Assessing the language proficiency of teachers: Are there any border controls? **Language Testing**, v. 2, n. 1, p. 149-170, 2001.

ELDER, C.; MCNAMARA, T.; KIM, H.; PILL, J.; SATO, T. Interrogating the construct of communicative competence in language assessment contexts: What the non-language specialist can tell us. **Language & Communication**, v. 57, p. 14-21, 2017.

ELLIS, R. **Instructed second language acquisition**. Oxford: Basil Blackwell, 1990.

ENSSLIN; L.; MONTIBELLER, G. B.; NORONHA, S. M. **Apoio à decisão: metodologias para estruturação de problemas e avaliação multicritério de alternativas**. Florianópolis: Insular, 2001.

FARHADY, H. The disjunctive fallacy between discrete-point and integrative tests. **TESOL Quarterly**, v. 13, n. 3, p. 347-357, 1979.

FARRELL, T. S.; RICHARDS, J. Teachers' language proficiency. In: FARRELL, T. S. (ed.). **Reflective language teaching: From research to practice**. London: Continuum, 2007.

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

FISCHER, A.; GREIFF, S.; FUNKE, J. The process of solving complex problems. **The Journal of Problem Solving**, v. 4, n. 1, p. 19-42, 2012.

FRANCO, L. A.; MONTIBELLER, G. **Problem structuring for multi-criteria decision analysis interventions**. London: The London School of Economics and Political Sciences, 2009.

FULCHER, G. **Testing second language speaking**. New York: Routledge, 2003.

FULCHER, G.; DAVIDSON, F. **Language testing and assessment: An advanced resource book**. New York: Routledge, 2007.

GARCÍA-CASCALES, M. S.; LAMATA, M. T. On Rank Reversal and TOPSIS Method. **Mathematical and Computer Modelling**, v. 56, p. 123-132, 2012.

GENESE, F. Evaluation. In: CARTER, R.; NUNAN, D. (ed.). **The Cambridge guide to teaching english to speakers of other languages**. Cambridge: Cambridge University Press, 2001.

GLIMCHER, P. W. **Decisions, uncertainty and the brain: The science of Neuroeconomics**. Cambridge, Massachusetts: The MIT Press, 2003.

GLIMCHER, P. W. Neuroeconomics: History. *In*: SQUIRE, L. R. **Encyclopedia of Neuroscience**. Cambridge, Massachusetts: Academic Press, 2009.

GLIMCHER, P. W.; FEHR, E. Introduction: A brief history of neuroeconomics. *In*: GLIMCHER, P. W.; FEHR, E. (ed.). **Neuroeconomics: Decision making and the brain**. Cambridge, Massachusetts: Academic Press, 2014.

GOMES, L. F. A. M.; GOMES, C. F. S.; ALMEIDA, A. T. de. **Tomada de decisão gerencial: o enfoque multicritério**. Rio de Janeiro: Ed. Atlas, 2009.

GRANFELDT, J.; ÅGREN, M. SLA developmental stages and teachers' assessment of written French: Exploring Direkt Profil as a diagnostic assessment tool. **Language Testing**, v. 31, n. 3, p. 285-305, 2014.

GROSS, C. G. Early history of Neuroscience. *In*: ADELMAN, G. (ed.). **Encyclopedia of Neuroscience**. Basel: Birkhäuser, 1987.

GROSS, C. G. History of Neuroscience: Early Neuroscience. *In*: SQUIRE, L. R. **Encyclopedia of Neuroscience**. Cambridge, Massachusetts: Academic Press, 2009.

GUITOUNI, A.; MARTEL, J. M. Tentative guidelines to help choosing an appropriate MCDA method. **European Journal of Operational Research**, v. 109, p. 501-521, 1998.

HAMP-LYONS, L. Holistic writing assessment of LEP students. *In*: OFFICE of Bilingual Education and Minority Languages Affairs (ed.). **Proceedings of the Second National Research Symposium on Limited English Proficient Student Issues: Focus on evaluation and measurement**, Vol. 2. Washington, D.C.: US Government Office of Bilingual Education and Minority Languages Affairs, 1992.

HAWKINS, J. A; BUTTERY, P. Using learner language from corpora to profile levels of proficiency: Insights from the English Profile Programme. *In*: TAYLOR, L.; WEIR, C. J. (ed.). **Language Testing Matters: Investigating the Wider Social and Educational Impact of Assessment**. Cambridge: CUP, 2009.

HAWKINS, J. A; BUTTERY, P. Criterial features in learner corpora: Theory and illustrations. **English Profile Journal**, v. 1, n. 1, p. 1-23, 2010.

HOUT, R.; VERMEER, A. Comparing measures of lexical richness. *In*: DALLER, H.; MILTON, J.; TREFFERS-DALLER, J. (ed.). **Modelling and assessing vocabulary knowledge**. Amsterdam: Benjamins, 2007

HYMES, D. Two types of linguistic relativity. *In*: BRIGHT, W. **Sociolinguistics**. The Hague: Mouton, 1966.

HUHTA, A.; ALANEN, R.; TARNANEN, M; MARTIN, M.; HIRVELÄ, T. Assessing learners' writing skills in a SLA study: Validating the rating process across tasks, scales and languages. **Language Testing**, v. 31, n. 3, p. 307-328, 2014.

IWASHITA, N.; BROWN, A.; McNAMARA, T.; O'HAGAN, S. Assessing levels of second language speaking proficiency: How distinct? **Applied Linguistics**, v. 29, n. 1, p. 24-49, 2008.

JACOBY, S.; MCNAMARA, T. Locating competence. **English for Specific Purposes**, v. 18, n. 3, p. 213-241, 1999.

JAMIESON, J.; POONPON, K. Developing analytic rating guides for TOEFL iBT integrated speaking tasks. **Research Report**. ETS, jul. 2013.

JENKINS, J. **The phonology of English as an International Language**. Oxford: Oxford University Press, 2000.

JIN, T.; MARK, B. Distinguishing features in scoring L2 Chinese speaking performance: How do they work? **Language Testing**, v. 30, n. 1, p. 23-47, 2012.

JOERIN, F.; THERIAULT, M.; MUSY, A. Using GIS and outranking multicriteria analysis for land-use suitability assessment. **International Journal of Geographical Information Science**, v. 15, n. 2, p. 153-174, 2001.

KANE, M. An argument-based approach to validation. **Psychological Bulletin**, v. 112, p. 527-535, 1992.

KANE, M. Certification testing as an illustration of argument-based validation. **Measurement: Interdisciplinary Research and Perspective**, v. 2, n. 3, p. 135-170, 2004.

KANE, M. Validating high-stakes testing programs. **Educational Measurement: Issues and Practices**, v. 21, n. 1, p. 31-, 2002.

KEENEY, R. L. Creativity in decision-making with Value-Focused Thinking. **Sloan Management Review**, v. 35, n. 4, p. 33-41, 1994.

KEENEY, R. L. **Value-Focused Thinking: A path to creative decision-making**. Cambridge: Harvard University Press, 1992.

KEENEY, R. L. Value-Focused Thinking: Identifying decision opportunities and creating alternatives. **European Journal of Operational Research**, v. 92, p. 537-548, 1996.

KEENEY, R. L.; MCDANIELS, T. L. Value-Focused Thinking about strategic decisions at BC Hydro. **Interfaces**, v. 22, n. 6, p. 94-109, 1992.

KEENEY, G. L.; VON WINTERFELD, D. Identifying and structuring the objectives of terrorists. **Risk Analysis**, v. 30, n. 12, p. 1803-1816, 2009.

KOL, J. P. T. **Avaliando as competências dos professores de língua estrangeira (inglês): um estudo sobre a validade de uma certificação internacional**. 2009. 296 f. Dissertação (Mestrado em Linguística Aplicada) - Universidade de Brasília, Instituto de Letras, Brasília, 2009

KROHLING, R. A.; CAMPANHARO, V. C. Fuzzy TOPSIS for group decision making: A case study for accidents with oil spill in the sea. **Expert Systems with Applications**, v. 38, p. 4190-4197, 2011.

KUIKEN, F.; VEDDER, I. Rater's decisions, rating procedures and rating scales. **Language Testing**, v. 31, n. 3, p. 279-284, 2014.

KUNNAN, A. J. High-Stakes Language Testing. *In*: CHAPELLE, C. (ed.). **The encyclopedia of Applied Linguistics**. Malden: Wiley, 2012.

LIMA JÚNIOR, F. R. **Comparação entre os métodos Fuzzy TOPSIS e Fuzzy AHP no apoio à tomada de decisão para seleção de fornecedores**. 2013. 150 f. Dissertação (Mestrado em Engenharia de Produção) – Universidade de São Paulo, Escola de Engenharia de São Carlos, São Carlos, 2013.

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

LUOMA, S. **Assessing speaking**. Cambridge: CUP, 2004.

MAROCHI, T. B. Testes orais individuais e em pares: um estudo das diferenças de desempenho. **DELTA**, v. 24, n. 1, p. 23-49, 2008.

MAROSSO, P. Z. **O emprego do acento nuclear em língua inglesa por falantes brasileiros: um estudo comparativo na era do Inglês como língua franca**. 2013. 63 f. Trabalho De Conclusão de Curso (Graduação em Licenciatura em Letras) – Universidade Tecnológica Federal do Paraná, Curitiba, 2013.

MCNAMARA, T. Assessment of second language proficiency. *In*: BROWN, K. (Ed.). **Encyclopedia of language and linguistics**. Amsterdam: Elsevier, 2006.

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

MCNAMARA, T. Language testing. *In*: DAVIES, A.; ELDER, C. **The handbook of Applied Linguistics**. Oxford: Blackwell, 2004.

MCNAMARA, T. **Measuring second language performance**. London: Longman, 1996.

MCNAMARA, T.; ROEVER, C. **Language testing: The social dimension**. Oxford: John Wiley & Sons, 2006.

MCWHITNEY, B. **Tools for analyzing talk: Part 2 – The CLAN Program**. Carnegie Mellon: Carnegie Mellon University, 2019.

MESSICK, S. The interplay of evidence and consequences in the validation of performance assessments. **Educational Researcher**, v. 23, n. 2, p. 13-23, 1994.

MESSICK, S. Validity. *In*: LINN, R. L. (ed.). **Educational measurement**. New York: Macmillan, 1989.

MESSICK, S. **Validity and washback in language testing**. New Jersey: ETS, 1996.

MINGERS, J.; ROSENHEAD, J. Problem structuring methods in action. **European Journal of Operational Research**, v. 152, p. 530-554, 2004.

MIRANDA, C. M. G.; ALMEIDA, A. T. de. Avaliação de pós-graduação com método ELECTRE TRI – o caso de Engenharias III da CAPES. **Revista Produção**, v. 13, n. 3, p. 101-112, 2003.

MORAIS, D. C.; ALENCAR, L. H.; COSTA, A. P. C. S.; KEENEY, R. L. Using Value-Focused Thinking in Brazil. **Pesquisa Operacional**, v. 33, n. 1, p. 73-88, 2013.

MORROW, C. K. Communicative language testing. *In*: LIONTAS, J. I. **The TESOL Encyclopedia of English Language Teaching**. New York: John Wiley & Sons, 2018.

MUNDA, G. Multiple-criteria decision aid: Some epistemological considerations. **Journal of Multi-Criteria Decision Analysis**, v. 2, n. 1, p. 41-55, 1993.

MYFORD, C. M.; WOLFE, E. W. **Monitoring sources of variability within the Test of Spoken English assessment system**. New Jersey: ETS, 2000.

NICHOLLS, D. The Cambridge Learner Corpus – error coding and analysis for lexicography and ELT. *In*: ARCHER, D.; RAYSON, P.; WILSON, A.; MCENERY, T. (ed.). **Proceedings of the Corpus Linguistics**, 2003, Conference, UCREL technical paper number 16, UCREL, Lancaster University, 2003.

NORTH, B. The Common European Framework of Reference: development, theoretical and practical issues. *In*: A NEW DIRECTION IN FOREIGN LANGUAGE EDUCATION: The Potential of the Common European Framework of Reference for Languages, 2006. **Proceedings [...]**. Japan: Osaka University of Foreign Studies, mar. 2006.

NORTH, B.; SCHNEIDER, G. Scaling descriptors for language proficiency scales. **Language Testing**, v. 5, n. 2, p. 217-262, 1998.

O'SULLIVAN, B. Language testing. *In*: SIMPSON, J. (ed.). **Routledge handbook of Applied Linguistics**. Oxford: Routledge, 2011.

PLOUGH, I. C.; BRIGGS, S. L.; BONN, S. V. A multi-method analysis of evaluation criteria used to assess the speaking proficiency of graduate student instructors. **Language Testing**, v. 27, n. 2, p. 235-260, 2010.

POLLITT, A.; MURRAY, N. L. What raters really pay attention to. *In*: MILANOVIC, M.; SAVILLE, N. (ed.). **Language Testing 3: Performance, testing, cognition and assessment**. Cambridge: Cambridge University Press, 1996.

RAJAGOPALAN, K. 'World English' or 'World Englishes'? Does it make any difference? **International Journal of Applied Linguistics**, v. 22, n. 3, p. 374-391, 2012.

RAJAGOPALAN, K. O inglês como língua internacional na prática docente. *In*: LIMA, D. C. (org.). **Ensino e aprendizagem de língua inglesa**. São Paulo: Editora Parábola, 2009.

RANGANATHAN, P.; PRAMESH, C. S.; AGARWAL, R. Common pitfalls in statistical analysis: Measures of agreement. **Perspectives in Clinical Research**, v. 8, p. 187-191, 2017.

REA-DICKINS, P. Fossilization or evolution: The case of grammar testing. In: ELDER, C. *et al.* **Experimenting with uncertainty**: essays in honor of Alan Davies. Cambridge: CUP, 2001.

RIAZI, A. M. **The Routledge encyclopedia of research methods in Applied Linguistics**: Quantitative, qualitative and mixed-methods research. New York: Routledge, 2016.

RICHARDS, H. M.; CONWAY, C.; ROSKVIST, A.; HARVEY, S. Foreign language teachers' language proficiency and their language teaching practice. **The Language Learning Journal**, v. 41, n. 2, p. 231-246, 2013.

RIMMER, W. Can what counts in complexity be counted? **Language Studies Working Papers**, v. 1, p. 25-34, 2009.

ROBBINS, S.; JUDGE, T.; SOBRAL, F. **Comportamento organizacional**: teoria e prática no contexto brasileiro. São Paulo: Pearson, 2010.

ROSENHEAD, J. What's the problem? An introduction to problem structuring methods. **Interfaces**, v. 26, n. 6, p. 117-131, 1996.

ROSENHEAD, J.; MINGERS, J. (ed.). **Rational analysis for a problematic world revisited**. Chichester: Wiley, 2001.

RUTISCHINI, A. Neuroeconomics: What have we found, and what should we search for. **Current Opinion in Neurobiology**, v. 19, n. 6, p. 672-677, 2009.

SAWAKI, Y. Construct validation of analytic rating scales in a speaking assessment: Reporting a score profile and a composite. **Language Testing**, v. 24, p. 355-390, 2007.

SCARAMUCCI, M. V. R. Proficiência em LE: considerações terminológicas e conceituais. **Trabalhos em Linguística Aplicada**, v. 36, p. 11-22, 2000.

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

SCHOFFEN, J. R. **Gêneros do discurso e parâmetros de avaliação de proficiência em português como língua estrangeira no exame CELPE-BRAS**. 2009. 192 f. Tese (Doutorado em Linguística Aplicada) – Universidade Federal do Rio Grande do Sul, Instituto de Letras, Porto Alegre, 2009.

SCHULZ, L. O.; CUSTÓDIO, M. M. C.; VIAPIANA, S. Concepções de língua, linguagem, ensino, aprendizagem e suas repercussões na sala de aula de língua estrangeira. **Pensar Línguas Estrangeiras**, v. 1, n. 1, p. 1-21, 2012.

SHEPARD, L.A. Evaluating test validity. **Review of Research in Education**, v. 19, p. 405-450, 1993.

SHOHAMY, E. **Language policy**: Hidden agendas and new approaches. London: Routledge, 2006.

SHOHAMY, E. Language tests as language policy tools. **Assessment in Education**, v. 14, n. 1, p. 117-130, 2007.

SHOHAMY, E. **The power of tests**. London: Longman, 2001.

SHOHAMY, E.; OR, I. G. Volume editors' introduction to "Language Testing and Assessment". *In*: SHOHAMY, E.; OR, I. G.; MAY, S. **Language testing and assessment**. Cham: Springer International Publishing AG, 2017.

SIEBERT, J.; VON WINTERFELD, D.; JOHN, R. S. Identifying and structuring the objectives of the "Islamic State of Iraq and the Levant" (ISIL) and its followers. **Decision Analysis**, v. 13, n. 1, p. 26-50, 2016.

SIMON, H. A. **Comportamento administrativo**: estudo dos processos decisórios nas organizações administrativas. Rio de Janeiro: Fundação Getúlio Vargas, 1965.

SMITH, J. K. Reconsidering reliability in classroom assessment and grading. **Educational Measurement: Issues and Practices**, v. 22, n. 4, p. 26-33, 2003.

SMITH, D. V.; HUETTEL, S. A. Decision neuroscience: Neuroeconomics. **WIREs Cognitive Science**, v. 1, p. 854-871, 2010.

SPOLSKY, B. History of language testing. *In*: SHOHAMY, E.; OR, I. G.; MAY, S. **Language testing and assessment**. Cham: Springer International Publishing AG, 2017.

STASSENKO, I.; SHOPINSKAJA, L.; LIIV, S. Investigating cultural variability in rater judgements of oral proficiency interviews. **Eesti Rakenduslingvistika Ühingu Aastaraamat**, v. 10, p. 269-281, 2010.

TAGUCHI, N. Rater variation in the assessment of speech acts. **Pragmatics**, v. 21, n. 3, p. 453-471, 2011.

TASKIN, H.; ÜSTÜN, Ö; DELIKTAS, D. Fuzzy MCDAM approach for oral examination in erasmus student selection process. **Journal of the Institute of Science & Technology of Dumlupinar University**, v. 32, p. 21-40, 2013.

TAYLOR, D. The meaning and using of the term competence in linguistics and applied linguistics. **Applied Linguistics**, v. 9, p. 148-168, 1988.

TAYLOR, L.; JONES, N. Cambridge ESOL exams and the Common European Framework of Reference (CEFR). **Research Notes**, v. 24, p. 2-5, 2006.

TEIXEIRA DA SILVA, V. L. **Fluência oral**: imaginário, construto e realidade num curso de Letras. 2000. 218 f. Tese (Doutorado em Linguística Aplicada) – Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas-SP, 2000.

THOMSON, R. I. Measurement of accentedness, intelligibility and comprehensibility. *In*: KANG, O.; GINTHER, A. **Assessment in second language pronunciation**. New York: Routledge, 2018.

TOFFOLI, S. F. L.; SIMON, C. V. B. A utilização do modelo multifacetado de Rasch na análise das influências dos avaliadores sobre as avaliações com itens abertos. **Ensaio: avaliação e políticas públicas em educação**, v. 26, n. 1, p. 1303-1323, 2018.

WALKER, R. **Teaching the Pronunciation of English as a Lingua Franca**. Oxford: Oxford University Press, 2010.

WaTRÓBSKI, J.; JANKOWSKI, J.; ZIEMBA, P.; KARZMARCZYK, A.; ZIOLO, M. Generalised framework for multi-criteria method selection. **Omega**, p. 1-18, 2018.

WEIR, C. J. **Language testing and validation: An evidence-based approach**. New York: Palgrave Macmillan, 2005a.

WEIR, J. P. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. **Journal of Strength and Conditioning Research**, v. 19, n. 1, p. 231-240, 2005b.

WIDDOWSON, H. Communicative language testing: The art of the possible. In: ELDER, C. *et al.* **Experimenting with uncertainty: Essays in honor of Alan Davies**. Cambridge: CUP, 2001.

XI, X.; SAWAKI, Y. Methods of test validation. In: SHOHAMY, E.; OR, I. G.; MAY, S. **Language testing and assessment**. Cham: Springer International Publishing AG, 2017.

YAN, X. An examination of rater performance on a local oral English proficiency test: A mixed-methods approach. **Language Testing**, v. 31, n. 3, p. 1-27, 2014.

YOUN, S. J. Rater variability across examinees and rating criteria in paired speaking assessment. **Papers in Language Testing and Assessment**, v. 7, n. 1, p. 32-60, 2018.

ZADEH, L. A. Fuzzy set. **Information and Control**, v. 8, n.3, p. 338-353, 1965.

ZADEH, L. A. The concept of a linguistic variable and its application to approximate reasoning – I. **Information Sciences**, v. 8, p. 199-249, 1975.

ZAVADSKAS, E. K.; MARDANI, A.; TURSKIS, Z.; JUSOH, A.; NOR, K. M. D. Development of TOPSIS method to solve complicated decision-making problems: An overview on developments from 2000 to 2015. **International Journal of Information Technology & Decision-Making**, v. 15, p. 1-38, 2016.

ZOPOUNIDIS, C.; PENTARAKI, K.; DOUMPOS, M. A review of country risk assessment approaches: A new empirical evidence. In: ZOPOUNIDIS, C.; PARDALOS, P. M. **Managing in uncertainty: Theory and practice**. Dordrecht: Springer Science+Business Media B. V., 1998.

APÊNDICES

APÊNDICE A – Características distintivas da escala de proficiência oral do EPPLE

	Faixas A e B	Faixas B e C	Faixas C e D	Faixas D e E
D1	-Acuidade no uso de estruturas linguísticas; necessidade de autocorreção.	-Grau em que os objetivos da comunicação são atingidos; -Fluência; -Acuidade no uso de estruturas linguísticas.	-Grau em que os objetivos da comunicação são atingidos; -Fluência; -Acuidade no uso de estruturas linguísticas.	-Grau em que os objetivos da comunicação são atingidos; -Acuidade no uso de estruturas linguísticas.
D2	-Pronúncia: proximidade com falante competente.	-Pronúncia: grau de proximidade com falante competente; -Interferência da LM.	-Pronúncia: grau de proximidade com falante competente;	-Pronúncia: grau de proximidade com falante competente; -Interferência da LM.
D3	-Qualidade do fornecimento de informações.	-Acuidade no uso de estruturas linguísticas; -Propriedade no emprego lexical.	(não há)	- Acuidade no uso de estruturas linguísticas.
D4	-Capacidade de narrar a cena; -Capacidade de estabelecer associações; -Capacidade de fornecer contribuição ao ensino-aprendizagem de línguas.	-Capacidade de formular hipóteses; -Capacidade de estabelecer associações; -Capacidade de fornecer contribuição ao ensino-aprendizagem de línguas.	(não há)	-Capacidade de descrever a cena.
D5	-Capacidade de selecionar informações relevantes; -Capacidade de explicar; -Uso de terminologia específica; -Compreensibilidade da explicação.	-Conhecimento de regras gramaticais; -Capacidade de explicar; -Capacidade de selecionar informações relevantes; -Uso de terminologia específica; -Compreensibilidade da explicação.	-Conhecimento de regras gramaticais; -Capacidade de selecionar informações relevantes; -Compreensibilidade da explicação.	-Conhecimento de regras gramaticais; -Capacidade de explicar.

D6	(não há)	-Capacidade de compreensão da fala; -Interferência no desempenho oral.	-Capacidade de compreensão da fala; -Interferência no desempenho oral.	-Interferência no desempenho oral.
----	----------	---	---	------------------------------------

APÊNDICE B - Questionário de levantamento de interesses sobre o exame EPPLE I

Stakeholders específicos: especialistas do EPPLE

Página 1

Este questionário consiste na primeira etapa de uma coleta de dados realizada para uma pesquisa de Doutorado desenvolvida pela Profa. Camila Sthéfanie Colombo junto ao PPGEL IBILCE/UNESP. A pesquisa visa a reestruturar a escala de proficiência do exame EPPLE. Por favor, responda ao questionário a seguir com base em suas crenças e experiências enquanto membro e pesquisador da equipe do exame EPPLE. Sua participação é imprescindível para o desenvolvimento da pesquisa.

Agradeço, antecipadamente, sua colaboração,
Camila Sthéfanie Colombo
colombo_camila@hotmail.com

Página 2

Termo de Consentimento Livre e Esclarecido - TCLE (Conselho Nacional de Saúde, Resolução 466/2012)

Você está sendo convidado a participar como voluntário do projeto de pesquisa “Avaliação de Proficiência Oral em Língua Estrangeira: Uma Proposta de Modelo de Validação de um Exame para Professores baseada na Teoria da Tomada de Decisão” sob responsabilidade da pesquisadora Camila Sthéfanie Colombo. O estudo será realizado com base em dados coletados por meio de questionários que visam à reestruturação da escala de proficiência do exame EPPLE. Haverá um risco mínimo para sua saúde caracterizado por eventual constrangimento ao responder as questões. 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. 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 a reestruturação e, eventual, adequação da escala de proficiência do teste oral do exame EPPLE.

Diante das explicações, se você concorda em participar deste projeto, forneça os dados solicitados e selecione a caixa “Li o termo e aceito participar desta pesquisa”.

Nome: _____ R.G. _____
Endereço: _____ Fone: _____
_____, _____ de _____ de 2016

() Li o termo e aceito participar desta pesquisa.

Nome da pesquisadora: Camila Sthéfanie Colombo	Cargo/Função: Discente do PPGEL
Instituição:	
Endereço:	
São José do Rio Preto – fone 17-3221.2480 e 3221.2482	

Página 3

Por favor, preencha as informações a seguir.

Observação: Informações marcadas com asterisco (*) prestam-se somente à obtenção de um panorama da amplitude da diversidade de participantes alcançada pelo questionário e permitem à pesquisadora entrar em contato para a realização da segunda etapa de coleta de dados, na qual será realizada uma consulta de opinião acerca dos resultados compilados nesta primeira fase. O nome do participante e de sua instituição de ensino, bem como quaisquer informações que permitam a identificação de ambos, serão eticamente preservados e, portanto, sob hipótese alguma serão divulgados.

1. Nome do participante*

2. E-mail do participante para contato*

3. Você autoriza a pesquisadora a entrar em contato com você, por meio do endereço fornecido, para a realização da segunda etapa de coleta de dados desta pesquisa?*

Marcar apenas uma oval.

() Sim

() Não

4. Qual sua formação acadêmica (graduação e pós-graduação)?

5. Em que instituição de ensino você atua (Nome-Sigla*/Estado)?

6. A qual âmbito sua instituição se vincula?

Marque todas que se aplicam.

() Estadual

() Federal

() Municipal

() Particular

() Outro: _____.

7. Qual cargo você ocupa na instituição?

Marque todas que se aplicam.

() Aluno de Letras

() Aluno de cursos de línguas

() Professor de línguas

() Diretor de escola

() Proprietário de escola

() Secretário de ensino

() Coordenador de ensino

() Selecionador de professores

() Treinador de professores

() Formador de professores

() Aplicador de exames de proficiência

() Corretor de exames de proficiência

() Elaborador de exames de proficiência

() Outro: _____.

8. Se você é professor de línguas, para quais níveis educacionais você leciona? Se atua em outros cargos, para quais níveis sua instituição presta serviços educacionais?

Marque todas que se aplicam.

- Educação infantil
- Ensino Fundamental I
- Ensino Fundamental II
- Ensino Médio
- Cursos pré-vestibulares
- Graduação (Letras)
- Graduação (Outros cursos)
- Pós-graduação lato sensu
- Pós-graduação stricto sensu
- Cursos preparatórios para concursos
- Cursos de línguas para crianças
- Cursos de línguas para jovens
- Cursos de línguas para adultos
- Cursos de línguas para a terceira idade
- Cursos preparatórios para exames de proficiência
- Outro: _____.

9. Se você é professor de línguas, como você ingressou nessa instituição? Se atua em outros cargos, como são feitos a seleção e o ingresso de professores nessa instituição?

Marque todas que se aplicam.

- Concurso público
- Entrevista
- Análise de currículo
- Prova didática
- Resultado de exame de proficiência consagrado (Cambridge Exams, TOEFL, TOEIC, ...)
- Resultado de exame de proficiência elaborados pela instituição
- Outro: _____.

10. Para a contratação de professores de línguas, é exigida comprovação de nível de proficiência mínimo na língua-alvo em que o professor irá atuar?

Marcar apenas uma oval.

- Sim
- Não

11. Se a resposta para a questão anterior for afirmativa, qual nível de proficiência mínimo é exigido para contratação dos professores de línguas?

Marque todas que se aplicam.

- Iniciante
- Básico
- Pré-intermediário
- Intermediário
- Intermediário superior
- Avançado
- Fluente
- O professor deverá ter proficiência equivalente ao nível para o qual irá lecionar
- O professor deverá ter proficiência superior ao nível para o qual irá lecionar
- Outro: _____.

12. Há quanto tempo você se encontra envolvido com o exame EPPLE?

13. Em quais fases de aplicação do exame EPPLE você tem experiência?

Marcar apenas uma oval.

- Elaboração de tarefas
- Elaboração de escala de proficiência
- Aplicação do exame
- Correção do exame e atribuição de conceitos
- Outro: _____.

14. Esclarecimentos e Informações ou comentários adicionais.

Página 4

Questionário de levantamento de interesses sobre o exame EPPLE

As questões apresentadas a seguir são referentes à profissão de professor de línguas e ao exame EPPLE. Caso seus conhecimentos sejam insuficientes para responder alguma questão, escreva apenas "Não tenho conhecimento sobre esse aspecto do exame". Por favor, responda às perguntas de acordo com sua opinião, justificando suas asserções.

1. O que você valoriza em um exame que avalie a proficiência de professores de línguas, como, por exemplo, o EPPLE?
2. Quais são os objetivos principais de um exame como o EPPLE?
3. Quais são os elementos fundamentais a serem avaliados em um exame como o EPPLE?
4. Como os elementos listados na questão anterior podem ser mensurados?
5. O que é bom e o que é ruim em cada faixa de proficiência apresentada pelo EPPLE?
6. Qual o número ideal de faixas de proficiência e que características devem conter cada faixa de um exame como o EPPLE para diferenciar um bom professor de língua estrangeira?
7. O que está certo e o que está errado na administração (organização, elaboração, aplicação, operacionalização, administração, correção, ...) do exame EPPLE? Cite exemplos de experiências boas e ruins e comente o que pode ser corrigido.
8. Quais limitações das diferentes fases de administração (organização, elaboração, aplicação, operacionalização, administração, correção, ...) do exame EPPLE são impostas sobre você na condição de elaborador/aplicador/examinador?
9. Levando em consideração o modo como o exame EPPLE encontra-se administrado hoje, quais efeitos oriundos dessa administração podem preocupar você em algum momento futuro?
10. Que efeitos a aplicação do exame EPPLE pode gerar nos âmbitos social, educacional, econômico e político?

Página 5

Questionário de levantamento de interesses sobre o exame EPPLE

Muito obrigada!

Stakeholders gerais: Atores do cenário de ensino-aprendizagem de línguas

Página 1

Este questionário consiste na primeira etapa de uma coleta de dados realizada para uma pesquisa de Doutorado desenvolvida pela Profa. Camila Sthéfanie Colombo junto ao PPGEL IBILCE/UNESP. A pesquisa visa a reestruturar a escala de proficiência do exame EPPLÉ. O EPPLÉ consiste em um exame de proficiência para professores de línguas estrangeiras - esta pesquisa interessa-se, mais especificamente, pela língua inglesa - que tenham ou não a língua-alvo como língua materna, mas que atuem no contexto brasileiro. Mais informações sobre o exame podem ser obtidas por meio do endereço <<http://epplebrasil.org/>>. Por favor, responda ao questionário a seguir com base em suas crenças e experiências profissionais. Sua participação é imprescindível para o desenvolvimento da pesquisa.

Agradeço, antecipadamente, sua colaboração,
Camila Sthéfanie Colombo
(colombo_camila@hotmail.com)

Página 2

Termo de Consentimento Livre e Esclarecido - TCLE
(Conselho Nacional de Saúde, Resolução 466/2012)

Você está sendo convidado a participar como voluntário do projeto de pesquisa “Avaliação de Proficiência Oral em Língua Estrangeira: Uma Proposta de Modelo de Validação de um Exame para Professores baseada na Teoria da Tomada de Decisão” sob responsabilidade da pesquisadora Camila Sthéfanie Colombo. O estudo será realizado com base em dados coletados por meio de questionários que visam à reestruturação da escala de proficiência do exame EPPLÉ. Haverá um risco mínimo para sua saúde caracterizado por eventual constrangimento ao responder as questões. 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. 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 a reestruturação e, eventual, adequação da escala de proficiência do teste oral do exame EPPLÉ.

Diante das explicações, se você concorda em participar deste projeto, forneça os dados solicitados e selecione a caixa “Li o termo e aceito participar desta pesquisa”.

Nome: _____ R.G. _____
Endereço: _____ Fone: _____
_____, _____ de _____ de 2016

() Li o termo e aceito participar desta pesquisa.

Nome da pesquisadora: Camila Sthéfanie Colombo	Cargo/Função: Discente do PPGEL
Instituição:	
Endereço:	
São José do Rio Preto – fone 17-3221.2480 e 3221.2482	

Página 3

Por favor, preencha as informações a seguir.

Observação: Informações marcadas com asterisco (*) prestam-se somente à obtenção de um panorama da amplitude da diversidade de participantes alcançada pelo questionário e permitem à pesquisadora entrar em contato com os participantes para a realização da segunda etapa de coleta de dados, na qual será realizada uma consulta de opinião acerca dos resultados compilados nesta primeira fase. O nome do participante e de sua instituição de ensino, bem como quaisquer informações que permitam a identificação de ambos, serão eticamente preservados e, portanto, sob hipótese alguma serão divulgados.

1. Nome do participante*

2. E-mail do participante para contato*

3. Você autoriza a pesquisadora a entrar em contato com você, por meio do endereço fornecido, para a realização da segunda etapa de coleta de dados desta pesquisa?*

Marcar apenas uma oval.

() Sim

() Não

4. Qual sua formação acadêmica (graduação e pós-graduação)?

5. Em que instituição de ensino você atua (Nome-Sigla*/Estado)?

6. A qual âmbito sua instituição se vincula?

Marque todas que se aplicam.

() Estadual

() Federal

() Municipal

() Particular

() Outro: _____.

7. Qual cargo você ocupa na instituição?

Marque todas que se aplicam.

() Aluno de Letras

() Aluno de cursos de línguas

() Professor de línguas

- Diretor de escola
- Coordenador de ensino
- Proprietário de escola
- Secretário de ensino
- Seleccionador de professores
- Treinador de professores
- Formador de professores
- Aplicador de exames de proficiência
- Corretor de exames de proficiência
- Elaborador de exames de proficiência
- Outro: _____.

8. Se você é professor de línguas, para quais níveis educacionais você leciona? Se atua em outros cargos, para quais níveis sua instituição presta serviços educacionais?

Marque todas que se aplicam.

- Educação infantil
- Ensino Fundamental I
- Ensino Fundamental II
- Ensino Médio
- Cursinhos pré-vestibulares
- Graduação (Letras)
- Graduação (Outros cursos)
- Pós-graduação lato sensu
- Pós-graduação stricto sensu
- Cursos preparatórios para concursos
- Cursos de línguas para crianças
- Cursos de línguas para jovens
- Cursos de línguas para adultos
- Cursos de línguas para a terceira idade
- Cursos preparatórios para exames de proficiência
- Outro: _____.

9. Se você é professor de línguas, como você ingressou nessa instituição? Se atua em outros cargos, como são feitos a seleção e o ingresso de professores nessa instituição?

Marque todas que se aplicam.

- Concurso público
- Entrevista
- Análise de currículo
- Prova didática
- Resultado de exame de proficiência consagrado (Cambridge exams, TOEFL, TOEIC, ...)
- Resultado de exame de proficiência elaborados pela instituição
- Outro: _____.

10. Na instituição em que você atua, para a contratação de professores de línguas, é exigida comprovação de nível de proficiência mínimo na língua-alvo em que o professor irá atuar?

Marcar apenas uma oval.

() Sim

() Não

11. Se a resposta para a questão anterior for afirmativa, qual nível de proficiência mínimo é exigido para contratação dos professores de línguas?

Marque todas que se aplicam.

() Iniciante

() Básico

() Pré-intermediário

() Intermediário

() Intermediário avançado

() Avançado

() Fluente

() O professor deverá ter proficiência equivalente ao nível para o qual irá lecionar

() O professor deverá ter proficiência superior ao nível para o qual irá lecionar

() Outro: _____.

12. Esclarecimentos e informações ou comentários adicionais

Página 4

Questionário de levantamento de interesses sobre o exame EPPLE

As questões apresentadas a seguir são referentes à profissão de professor de línguas e ao exame EPPLE. Caso seus conhecimentos sejam insuficientes para responder alguma questão, escreva apenas "Não tenho conhecimento sobre esse aspecto". Se você não atua como professor de línguas, mas

(i.) ocupa cargos de gerência, por favor, responda às questões com base nas características que você gostaria para os professores por você gerenciados,

(ii.) é aluno, responda às questões com base no que você gostaria para os professores que lhe ministram aulas.

Por favor, responda às perguntas de acordo com sua opinião, justificando suas asserções.

1. Você já participou de algum exame que avalie a proficiência de professores de línguas para atuar no contexto brasileiro? Em caso afirmativo, informe a(s) data(s), o(s) nome(s) do(s) exame(s) e a(s) pontuação(ões) por você obtida(s).

2. Você participaria de um exame que avaliasse a proficiência de professores de línguas para atuar no contexto brasileiro, como o EPPLE? Em caso afirmativo, comente a utilidade desse exame para sua vida profissional. Em caso negativo, explique por quê.

3. Na sua opinião, quais características linguísticas, comunicativas e pedagógicas caracterizam um bom professor de língua estrangeira?

4. Pensando em sua atuação enquanto professor de línguas: Com quais características de sua proficiência seu empregador e seus alunos se importam?
5. Pensando em sua proficiência enquanto professor de línguas hoje: Em algum momento futuro, que efeitos oriundos de sua atuação enquanto professor de línguas poderão preocupar você?
6. Pensando em um exame que avalie a proficiência do professor de línguas para atuar no contexto brasileiro: Quais devem ser seus objetivos principais?
7. Pensando em um exame que avalie a proficiência do professor de línguas para atuar no contexto brasileiro: Que elementos linguísticos, comunicativos e pedagógicos ele deve avaliar?
8. Você conhece o exame EPPLE? Já participou de alguma aplicação do exame EPPLE? Se sim, em que ano e qual foi sua pontuação?
9. Se você respondeu afirmativamente à questão de número 7: O que já ocorreu de bom e de ruim nas aplicações do exame EPPLE das quais você participou?
10. Se você respondeu afirmativamente à questão de número 7: Sua participação no exame EPPLE atendeu as suas expectativas? Comente.

Página 5

Questionário de levantamento de interesses sobre o exame EPPLE

Muito obrigada!

APÊNDICE C - Questionário de levantamento de interesses sobre o exame EPPLE II

Stakeholders gerais e específicos

Página 1

1 Você concorda que todos os critérios listados a seguir são adequados e igualmente importantes para avaliar a proficiência do professor de línguas?

Sim

Não – siga para a página 2

Critérios/Subcritérios	
Capacidade de compreensão oral	
Capacidade de contribuir com a interação	
Capacidade de uso de elementos linguísticos	Vocabulário
	Gramática
	Entoação
	Fluência
Capacidade de contribuir com o ensino da língua estrangeira	Pronúncia
	Conhecimento de tópico/conteúdo do ensino
	Adequação da linguagem do professor de línguas
	Fornecimento de instruções para gerenciamento de sala de aula e introdução de atividades
	Uso de metalinguagem (uso de terminologia específica e fornecimento de explicações)

2 Você concorda que todos os critérios listados a seguir são passíveis de serem mensurados quantitativa ou qualitativamente para a avaliação da proficiência do professor de línguas?

Sim

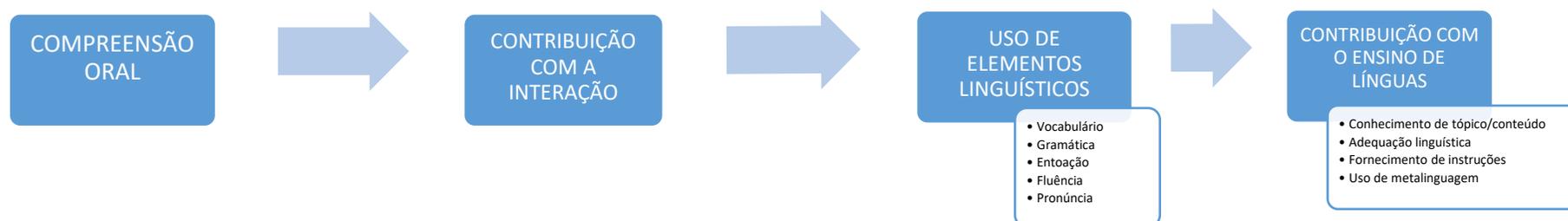
Não – siga para a página 2

Critérios/Subcritérios	
Capacidade de compreensão oral	
Capacidade de contribuir com a interação	
Capacidade de uso de elementos linguísticos	Vocabulário
	Gramática
	Entoação
	Fluência
Capacidade de contribuir com o ensino da língua estrangeira	Pronúncia
	Conhecimento de tópico/conteúdo do ensino
	Adequação da linguagem do professor de línguas
	Fornecimento de instruções para gerenciamento de sala de aula e introdução de atividades
Uso de metalinguagem (uso de terminologia específica e fornecimento de explicações)	

3 Você concorda com a ordem de mensuração dos critérios proposta a seguir?

Sim

Não – siga para a página 2



4 Você concorda com o modo de mensuração de todos os critérios proposto a seguir?

Sim

Não – assinale apenas aqueles com as quais você concorda e siga para a página 2

Critério		Modo de medição	Unidade de medida	Variação da unidade		
				Limite inferior	Limite superior	
Compreensão oral		<input type="checkbox"/> Cumprimento da tarefa	<input type="checkbox"/> Grau de completude da resposta	<input type="checkbox"/> Nenhum tópico	<input type="checkbox"/> Resposta completa	
Contribuição com a interação		<input type="checkbox"/> Extensão textual	<input type="checkbox"/> Tempo de resposta	<input type="checkbox"/> Inferior a 1 min	<input type="checkbox"/> 2 minutos ou mais	
		<input type="checkbox"/> Propriedade de ideias	<input type="checkbox"/> Grau de relevância das ideias	<input type="checkbox"/> Repetitivas e/ou limitadas ao insumo	<input type="checkbox"/> Ideias novas, não limitadas ao insumo	
			<input type="checkbox"/> Grau de relação com o tema	<input type="checkbox"/> Não relacionado (fuga ao tema)	<input type="checkbox"/> Completamente relacionado	
			<input type="checkbox"/> Quantidade de ideias	<input type="checkbox"/> Limitado a opiniões, sem oferta de razões	<input type="checkbox"/> Fornecimento de opiniões e razões	
		<input type="checkbox"/> Organização discursiva	<input type="checkbox"/> Grau de clareza das ideias	<input type="checkbox"/> Pouca coerência	<input type="checkbox"/> Coerência plena	
			<input type="checkbox"/> Grau de progressão textual	<input type="checkbox"/> Texto sem introdução e/ou conclusão	<input type="checkbox"/> Texto com introdução, desenvolvimento e conclusão	
Uso de elementos linguísticos	Lexicais	Vocabulário	<input type="checkbox"/> Precisão	<input type="checkbox"/> Grau de adequação ao significado	<input type="checkbox"/> Inadequado/pouco adequado	<input type="checkbox"/> Completamente adequado
	Semânticos		<input type="checkbox"/> Adequação	<input type="checkbox"/> Grau de adequação ao contexto	<input type="checkbox"/> Inadequado/pouco adequado	<input type="checkbox"/> Completamente adequado
	Pragmáticos		<input type="checkbox"/> Variedade	<input type="checkbox"/> Quantidade de tokens e types	<input type="checkbox"/> Baixa frequência de types e/ou reutilização frequente de insumo	<input type="checkbox"/> Alta frequência de tokens e types, expressões idiomáticas, escolhas estilísticas
	Sociolinguísticos					
	Gramaticais	Gramática	<input type="checkbox"/> Acurácia	<input type="checkbox"/> Quantidade de erros estruturais (nível da palavra/frase)	<input type="checkbox"/> Erros muito frequentes	<input type="checkbox"/> Erros pouco frequentes
		<input type="checkbox"/> Grau de variedade de erros estruturais (nível da palavra/frase)		<input type="checkbox"/> Erros muito variados	<input type="checkbox"/> Erros pouco variados	

			() Quantidade de AS-units	() Poucos elementos coesivos	() Muitos elementos coesivos	
			() Grau de variedade de AS-units	() Pouca variedade	() Muita variedade	
	Sintáticos		() Complexidade	() Produto da divisão da quantidade total de orações isoladas e de C-units	() Pouca subordinação	() Subordinação frequente
	Fonético-fonológicos	Entoação	() Tom	() Grau de variedade	() Alta constância	() Alta variedade
			() Acento	() Grau de tipicidade	() Pouco típico	() Muito típico
			() Ritmo	() Grau de tipicidade	() Pouco típico	() Muito típico
		Fluência	() Velocidade de fala	() Quantidade de sílabas/segundo	() Inferior ou superior a 4.7 sílabas/seg	() Média de 4.7 sílabas/seg
			() Motivo de hesitações	() Escolha de estruturas gramaticais, escolha vocabular, reparos estruturais	() Planejamento de expressão, elicitacão de exemplos para apoiar opiniões, reformulações para clarificação	
			() Fluidez discursiva	() Frequência e duração de pausas (preenchidas e silenciosas)	() Preenchidas: baixa () Silenciosas: acima de 0,6-0,4s	() Preenchidas: alta () Silenciosas: entre 0,6 e 0,4s
			() Frequência de repetição	() Alta	() Baixa	
			() Frequência de reparo	() Alta	() Baixa	
			() Frequência de falsos inícios	() Alta	() Baixa	
			() Frequência de rupturas	() Alta	() Baixa	
			() Frequência de extensão silábica	() Alta	() Baixa	
			Pronúncia	() Inteligibilidade	() Grau de efetividade da compreensão	() Inadequado
		() Compreensão		() Grau de dificuldade de compreensão	() Alto	() Baixo
		() Sotaque		() Grau de influência da LM	() Alto	() Baixo
Contribuição com o ensino da língua		Conhecimento de tópico/conteúdo de ensino	() Cumprimento da tarefa	() Grau de adequação da resposta	() Pouco adequada	() Completamente adequada
	() Grau de completude da resposta		() Insuficiente (poucos tópicos)	() Resposta completa		
	Adequação da linguagem docente	() Repetições	() Frequência	() Alta	() Baixa	
		() Substituição/Reformulação	() Frequência	() Alta	() Baixa	
		() Checagem de compreensão/ <i>Feedback</i>	() Frequência	() Alta	() Baixa	
		() Grau de adequação contextual	() Pouco adequado	() Muito adequado		

	Fornecimento de instruções (gerenciamento de sala e atividades)		() Variedade vocabular	() Quantidade de tokens e types	() Baixa frequência de types e/ou reutilização frequente de insumo	() Alta frequência de tokens e types, expressões idiomáticas, escolhas estilísticas
			() Precisão vocabular	() Grau de adequação ao significado	() Pouco adequado	() Muito adequado
	Uso de metalinguagem	Uso de terminologia específica	() Precisão vocabular	() Grau de adequação ao significado	() Pouco adequado	() Muito adequado
			() Descrição do uso	() Grau de adequação da resposta	() Pouco adequado	() Muito adequado
		Fornecimento de explicação	() Exemplificação	() Grau de adequação da resposta	() Pouco adequado	() Muito adequado
			() Elicitação de participação	() Frequência	() Inexistente	() Satisfatória
			() Fornecimento de correção	() Grau de adequação vocabular	() Pouco adequado	() Muito adequado
			() Fornecimento de correção	() Grau de adequação	() Inadequado	() Adequado

5 Comentários:

Página 2

Apenas responda as questões desta página se você assinalou a opção “não” em alguma das questões da página anterior.

1 Marque todos os itens que você acredita que devam integrar a avaliação da proficiência do professor de línguas:

	Acento		Inteligibilidade
	Acurácia gramatical		Organização discursiva
	Adequação ao contexto		Pausas (preenchidas e silenciosas)
	Adequação da linguagem docente		Precisão vocabular
	Adequação da resposta		Progressão textual
	Checagem de compreensão/ <i>Feedback</i>		Pronúncia
	Clareza das ideias		Propriedade de ideias
	Completude da resposta		Quantidade de ideias
	Complexidade gramatical		Relação com o tema

	Compreensibilidade		Relevância das ideias
	Compreensão oral		Reparo
	Conhecimento de tópico/conteúdo de ensino		Repetição
	Contribuição com a interação		Ritmo
	Contribuição com o ensino da língua		Rupturas
	Cumprimento da tarefa		Sotaque
	Descrição do uso da língua		Substituição/Reformulação
	Elicitação de participação		Tempo de resposta
	Entoação		Tom
	Exemplificação		Uso de elementos fonético-fonológicos
	Extensão silábica		Uso de elementos gramaticais
	Extensão textual		Uso de elementos lexicais
	Falsos inícios		Uso de elementos pragmáticos
	Fluência		Uso de elementos semânticos
	Fluidez discursiva		Uso de elementos sintáticos
	Fornecimento de correção		Uso de elementos sociolinguísticos
	Fornecimento de explicação		Uso de metalinguagem
	Fornecimento de instruções		Uso de terminologia específica
	Hesitações		Velocidade de fala
	Influência da língua materna		Outros:

2 Agora, utilizando-se da escala Likert de 5 pontos a seguir, assinale o grau de importância dos critérios que você marcou na questão de número 1 desta página:

1 – imprescindível 2 – muito importante 3 – importante 4 – pouco importante 5 – sem importância

Observações: a) Números iguais devem ser atribuídos a critérios com a mesma importância / b) Critérios que você não assinalou na questão 1 desta página não devem integrar a hierarquização e, portanto, não se deve atribuir números a eles. Deixar em branco.

	Acento		Inteligibilidade
	Acurácia gramatical		Organização discursiva
	Adequação ao contexto		Pausas (preenchidas e silenciosas)
	Adequação da linguagem docente		Precisão vocabular
	Adequação da resposta		Progressão textual
	Checagem de compreensão/ <i>Feedback</i>		Pronúncia
	Clareza das ideias		Propriedade de ideias
	Completeness da resposta		Quantidade de ideias
	Complexidade gramatical		Relação com o tema
	Compreensibilidade		Relevância das ideias
	Compreensão oral		Reparo
	Conhecimento de tópico/conteúdo de ensino		Repetição
	Contribuição com a interação		Ritmo
	Contribuição com o ensino da língua		Rupturas
	Cumprimento da tarefa		Sotaque
	Descrição do uso da língua		Substituição/Reformulação
	Elicitação de participação		Tempo de resposta
	Entoação		Tom
	Exemplificação		Uso de elementos fonético-fonológicos
	Extensão silábica		Uso de elementos gramaticais
	Extensão textual		Uso de elementos lexicais
	Falsos inícios		Uso de elementos pragmáticos
	Fluência		Uso de elementos semânticos
	Fluidez discursiva		Uso de elementos sintáticos
	Fornecimento de correção		Uso de elementos sociolinguísticos
	Fornecimento de explicação		Uso de metalinguagem
	Fornecimento de instruções		Uso de terminologia específica
	Hesitações		Velocidade de fala
	Influência da língua materna		Outros:

3 Marque todos os itens que você acredita que são passíveis de serem efetivamente mensurados quantitativa ou qualitativamente na avaliação da proficiência do professor de línguas:

	Acento		Inteligibilidade
	Acurácia gramatical		Organização discursiva
	Adequação ao contexto		Pausas (preenchidas e silenciosas)
	Adequação da linguagem docente		Precisão vocabular
	Adequação da resposta		Progressão textual
	Checagem de compreensão/ <i>Feedback</i>		Pronúncia
	Clareza das ideias		Propriedade de ideias
	Completeness da resposta		Quantidade de ideias
	Complexidade gramatical		Relação com o tema
	Compresibilidade		Relevância das ideias
	Compreensão oral		Reparo
	Conhecimento de tópico/conteúdo de ensino		Repetição
	Contribuição com a interação		Ritmo
	Contribuição com o ensino da língua		Rupturas
	Cumprimento da tarefa		Sotaque
	Descrição do uso da língua		Substituição/Reformulação
	Elicitação de participação		Tempo de resposta
	Entoação		Tom
	Exemplificação		Uso de elementos fonético-fonológicos
	Extensão silábica		Uso de elementos gramaticais
	Extensão textual		Uso de elementos lexicais
	Falsos inícios		Uso de elementos pragmáticos
	Fluência		Uso de elementos semânticos
	Fluidez discursiva		Uso de elementos sintáticos
	Fornecimento de correção		Uso de elementos sociolinguísticos
	Fornecimento de explicação		Uso de metalinguagem
	Fornecimento de instruções		Uso de terminologia específica
	Hesitações		Velocidade de fala
	Influência da língua materna		Outros:

4 Indique como mensurar os critérios assinalados por você na questão de número 3 seguindo as orientações a seguir:

- a) escreva o critério escolhido;
- b) escreva O se o critério deve ser avaliado mais objetivamente e S se ele deve ser avaliado mais subjetivamente – se houver a possibilidade de avaliá-lo de ambas as formas, escolha apenas uma;
- c) indique a variação desejável para avaliação do critério, informando a quantidade de níveis (a maioria das escalas de proficiência variam em até 5 níveis) e os atributos que caracterizam e diferenciam cada nível; e
- d) indique a unidade de mensuração que deve ser empregada na avaliação do critério.

5 Comentários:

Apenas stakeholders específicos

Página 3

As transcrições a seguir representam dados de fala de candidatos ao exame EPPL, em uma de suas versões eletrônicas, no ano de 2015²⁶.

Candidato 1:

- 1063:** eh: I am from: ((nome da cidade)) + a c/ a town of the state of São Paulo + I am twenty-one years old
- 1065:** I started my experience + of earn/ earning english in the school + and I was: a kid + and my interested begin began + in that time + ah: I chose the + english language the university + and I like very much
- 1069:** I'll try to make + new friends + ah: s/ speak a: native english and maybe + to watch + hm + theather + movies an:d + go to places with lots of people [1] and a job + I retain to: put in/

²⁶ Aplicações UNESP/IBILCE 2015 (Versão 3.0) - Candidato 1 (61569): A1 (C), Candidato 2 (61577): A7 (D), Candidato 3 (78039): A2 (C), Candidato 4 (78041): A9 (B), Candidato 5 (78043): A6 (B).

- 1071: I never worked as a english teacher + but + I: + intend to when I finish the letters course + ah: I like the area + of working
- 1073: I would like to teach according to my: course with letters ah: + I like both languages portuguese and english + I: would like to: teach grammar
- 1075: the first extract video + ah: is about the accent + and the importance of your personal style on: o/ on your: accent you don't have to look like a native speaker + like: a british person + or american person + but you have to + to feel comfortable about your own [2] style of: accent
- 1081: I think mr. duncan's talking to: the students the + english language students + an:d because they are the people who most + worry about the + accent when they are learning
- 1084: the second video extract is about + a: a man who tries to: learn the + the english + language and he has a (INCOMP) that sentence probably + a foreigner + a non native speaker + and the teacher is trying in him to + to make him talk like a + (INCOMP) accent + just to: basic communication
- 1088: the: + two videos provide ah + informations eh: who talk + that we + can have an accent + a non standardisation accent + and we but we have to: + make our way to: express + in a good way + like second video + the man he can express himself in english + but he is trying to
- I chose the second + situation + ah: w/ will talk wi/ we use will when we are + talking about a prediction + we're not certain of about the future +
- 1097: I don't know if the the THING will will happen in a certain way + and you use + going to when you have + ah: + when the the + the thing will happen is closer + and you had a *suretainty* + more ah you are more (sure that that will) happen

Candidato 2:

- 1063: I'm I'm ((nome do candidato)) I'm from ((nome da cidade)) + an:d + my occupation now is student an:d I don't have + another + occupation [3]
- 1065: well + I start at at my college + in: two thousand eleven + beFORE this period I had ano/ others experience in english + english BUT I think eh this experience + it + was not very good for: + improve my english I had ah english in my secondary school and ah: + I I don't think that this/ maybe I I I would not go with brazilians people with brazilian people because + I don't think this w/ will improve and maybe I I will + ah enter in a + a high school for study there
- 1069: my experience with english eh wi/ eh AS a english teacher is only at the + *estágio* in: at school at at yes at a a + a school and: I don't think that
- 1071: now I'm very prepared for be a teacher + so I + would like to: + improve more for that work
- the video is about ah: + a maybe a question that a a student english asked for what what is the better accen/ accent to learn in english an:d ah mr.
- 1075: duncan explain this + this question saying that there is no accent a proper accent you: you say you: speak in english as you + as your own language you you have to to create your own own way to: to speak in other eh to be communicate
- 1081: eh: eh the person who + mr duncans is + speaking to is probably a student of english *e*: and probably a foreign *foreigned* student because he ah speak about accent and + this is a question that student *FOREIGNed* students asked eh usually asked [1] ask [10]
- 1084: the video is about pronunciation + english pronunciation and + this is a: a [1] a *methodo* to to learn english repeating what the professor said + and + I think ah this is quite common in some courses and videos to to learn english in + on the internet for example [7]
- we can + ah: + see that the english language is a language that you learn by by heart by + eh WITH your heart + no by heart and you need to to put your your feelings and emotions to to learn this language and the na/ the other one isn't/ ah: the many point is that + some + sounds an:d soun/ some sounds is difficult for a specific nationality
- 1088: that I choose is the second one and ah the difference the the main difference between going to and will is that we going to is ah a decision more planned and ah you have more you have more sure about the event or the situation that will have happened so and will is + more: you have we
- 1097:

don't have more: you have + as certain as you have with when you use going to so + this is a: I think the the main difference and also ah will we have + to + for: you ha/ you you use go/ will for instan/ instant decision and ah + situations that you don't have clue to to say what you are saying [10]

Candidato 3:

- 1063:** [1] my name is ((nome do candidato)) + and I'm a + professor + an:d also a student
[2] I started s/ ah tea/ ah + studying english + ah: in two thousand and nine around two thousand and nine and I've been studying english since
- 1065:** then + ah: I studied I started studying english in a private school + called ((nome da escola)) in in the city where I li/ where I was born in ah: and when I graduated there + I came to college and started studying english here + and now ah: I also teach english at a private school [1]
[2] I would go out + ah most of the time and try to talk to people ah: about many things such as touristic ah: places ah places to visit places to EAT
- 1069:** + ah things like that because I think when you are speaking the language ah in a a real context of use you can improve your + your language in the case your english better than just ah + studying in a classroom
[2] I've started s/ ah started teaching english ah: last year + with: ah: a class ah: + directed to: the entrances entrances exams and then I started
- 1071:** teaching english in a private school called ((nome da escola)) this year and I've been teaching there ah fo/ since august since ju/ JULY and ah: that's it
[2] my plans for my professional life are to become an english teacher at ah: + schools ah + in private schools maybe and ah: + private ENGLISH
- 1073:** schools but I want but what I really want is to become an english teacher at *cursinhos* because + that's what I like and that's where I love the most to work at
- 1075:** [1] the first eh video extract is about if you should speak like a british guy or if you should speak like an american guy + but + what is really important is that you should speak like you you should sound like yourself
- 1081:** [2] I think mr. duncan is talking to and speaking to + ah: anyone else who's learning english as a foreign language
- 1084:** [2] the second video extract is about a man who's trying to pronoun/ to PROperly pronounce a certain sentence in english
- 1088:** [2] I believe that the video + the videos that I just saw provide me the information that english language students and teachers shouldn't be worried about pronouncing ah: the language in a specific way whether they need to + speak it and make themselves understood ah: + with the language
- 1097:** [2] ah going to is used when you are talking about a situation in the future that you know + is going to happen + and will is used for predictions so youyou use will you are talking about a situation in the future that you are not sure that would happen

Candidato 4:

- 1063:** [2] my name is ((nome do candidato)) + I'm from ((nome da cidade)) + and my main occupation is studying [3]
I started studying english when I was [1] ah: fourteen + at a: in my city my town ((nome da cidade)) in a an english school ALTHOUGH I + had
- 1065:** been + I had had class/ had been having having classes in english ah: + at school since I was + ah: NIne I think [1] and I've been studying english ever since
- 1069:** I would probably try to: + talk + to + as many + people as I could [1] ah: + and also at different PLAcEs hm: [3] an:d [20] (INCOMP)

- 1071:** [4] I have I HAVEN'T been I have neer been an english teacher have never worked as an english teacher + and I: + do not consider + ah: being a teacher [2] now
- 1073:** [2] hm + I intend to take a test to became a [1] to to go to the public service [3] and maybe if I [2] ah: if it takes me too long to get a job + in that area + I might ah: have a job as an english teacher
- 1075:** the first video talks about + ah: what you should sound like when you are talking in english [1] and ah: + it says + that any/ everyone should sound like themselves they shouldn't + uh sound like somebody else just because they're speaking + some other language [2] and ah: [3] it/ it's really about accent
- 1081:** he's talking about + he's talking to + ah: people that were born in + non english speaking countries [2]
- 1084:** the second + the second video is about how difficult it is + for people whose first language is not english [2] ah: + to learn + english the videos + ah: [1] are ah make me think that + giv/ give the information + that + it is very hard for someone + ah: that wasn't born in a sp/ english speaking country + to learn english [1] and ah: + they also say + that ah as long as + you: as long as people understand you ah: + your accent doesn't matter th/ the way you sound doesn't matter as long as + as people can understand + yo/ ah: what you're TALKing + in english
- 1097:** [7] the problem with the first one is the: verb tense + ah: [2] since it's it's ah [1] eh eh the ver/ the verb tense should be: present + ah: [5] present perfect continuous instead of + present it should be I have been living in são paulo all my life

Candidato 5:

- 1063:** [3] my name is ((nome do candidato)) + and: I am twenty years old + and I'm from: ((nome da cidade)) Brazil but now I live in ((nome da cidade)) + and: + I believe my main occupation + right now is to: + finish my graduation and to take care of my son
- 1065:** well I started when I was: seven years old + I believe I started a few years before but just at school + but at seven I started in a: + in a language school + and: + I spen:t I believe five or six years studying there and + I think it helped me a lot to to chose my my course when I when I came to to to graduate and: I think that's it
- 1069:** [2] oh I think if I could I would find a job to: to to to work with people and to to know people + because + working I think we can: we can deal better with the language an:d + we can know people from different countries and with different accents maybe and I think + would be + maybe better than just trav/ travel
- 1071:** last year I worked as an assistant in a: bilingual school for for kids an:d + there we: we used to teach + children about like + three three + almost four years old kids an:d + we + we are: supposed to te/ to te/ teach them in english so: I think + I started there + my experience as an english teacher [1] not a/ as a as a real teacher but + almost something like that
- 1073:** I: I don't know I just don't know I: just want to finish my graduation an:d I think I will think about when I when I finish it so: + I'm still waiting for [1] for something to [1] maybe to god waiting for god to: + to guide my path
- 1075:** the first video is about:t different ways of speaking english and different accents and how: a pear/ a person who is learning + english as a second language + shoul:d + find + his or her + own way + of speaking + there is no need to copy or: something like that
- 1081:** I believe he's talking to + all: those people that + are interested in learning english [2]
- 1084:** well I believe the se:cond video is about:t learning english as a foreign language an:d how difficult it can be + so we have the + Accent + there's a difficulty of it + and [1] the teacher keeps repeating + time after time + hoping for the student to + repeat what she's saying + in a proper way

- 1088:** [1] I believe this information would be that pronunciation it's a very important aspect of in the process of learning in english + but + the the person also ha/ has to: [1] has to have his or her own way of learning and speaking so + it's it's also a personal thing [1] despite being + something + important for the language
- 1097:** [1] wow I ch/ I chose the: situation number two and my explanation would be that + you can use going to to talk about the future when you are really sure that the fact is going to happen + so it will happen in a near future and + you can use will when you are not so sure about the fact + so it can happen in the future but going to + would show more certain about the fact

Observe as transcrições, leia o prompt do teste oral, ouça os arquivos de áudio e, para cada candidato:

1 Descreva a linguagem empregada pelo candidato para responder a questão. Informe os pontos positivos e negativos do desempenho em cada transcrição com relação aos seguintes aspectos:

- a. capacidade de compreensão oral;
- b. capacidade de contribuir com a interação: adequação de ideias e organização discursiva;
- c. capacidade de uso de elementos linguísticos: vocabulário, gramática, entoação, fluência e pronúncia;
- d. capacidade de contribuição com o ensino da língua: conhecimento do conteúdo de ensino, adequação da linguagem docente, fornecimento de instruções e emprego de metalinguagem (uso de terminologia específica e fornecimento de explicações);
- e. outros aspectos que você considerar relevantes.

Candidato 1:

Candidato 2:

Candidato 3:

Candidato 4:

Candidato 5:

2 Utilizando-se dos critérios e dos modos de mensuração fornecidos por você nas questões das páginas anteriores:

a) avalie a linguagem dos candidatos individualmente;

b) indique se houve dificuldade para mensurar algum dos critérios e, em caso afirmativo, como você resolveu esse impasse.

Candidato 1:

Candidato 2:

Candidato 3:

Candidato 4:

Candidato 5:

3 Utilizando-se da escala Likert de 5 pontos apresentada a seguir, indique o nível de proficiência para ensinar línguas de cada candidato e justifique.

1 – extremamente proficiente 2 – muito proficiente 3 – satisfatoriamente proficiente 4 – pouco proficiente 5 – insuficientemente proficiente

Candidato 1:

Candidato 2:

Candidato 3:

Candidato 4:

Candidato 5:

4 Comentários:

ANEXOS

ANEXO A - Escala Global do CEFR

PROFICIENT USER	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
INDEPENDENT USER	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
BASIC USER	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

ANEXO B - A escala do IELTS

Band score	Skill level	Description
9	Expert user	The test taker has fully operational command of the language. Their use of English is appropriate, accurate and fluent, and shows complete understanding.
8	Very good user	The test taker has fully operational command of the language with only occasional unsystematic inaccuracies and inappropriate usage. They may misunderstand some things in unfamiliar situations. They handle complex and detailed argumentation well.
7	Good user	The test taker has operational command of the language, though with occasional inaccuracies, inappropriate usage and misunderstandings in some situations. They generally handle complex language well and understand detailed reasoning.
6	Competent user	The test taker has an effective command of the language despite some inaccuracies, inappropriate usage and misunderstandings. They can use and understand fairly complex language, particularly in familiar situations.
5	Modest user	The test taker has a partial command of the language and copes with overall meaning in most situations, although they are likely to make many mistakes. They should be able to handle basic communication in their own field.
4	Limited user	The test taker's basic competence is limited to familiar situations. They frequently show problems in understanding and expression. They are not able to use complex language.
3	Extremely limited user	The test taker conveys and understands only general meaning in very familiar situations. There are frequent breakdowns in communication.
2	Intermittent user	The test taker has great difficulty understanding spoken and written English.
1	Non-user	The test taker has no ability to use the language except a few isolated words.
0	Did not attempt the test	The test taker did not answer the questions.

ANEXO C - Escala de proficiência oral do EPPLE

Extraído de Consolo e Teixeira da Silva (2014)

FAIXA A

A1) Atinge plenamente os objetivos de comunicação verbal, apresentando fluência em 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, sem 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 autocorrigir.

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.

TERMO DE REPRODUÇÃO XEROGRÁFICA

Autorizo a reprodução xerográfica do presente Trabalho de Conclusão, na íntegra ou em partes, para fins de pesquisa.

São José do Rio Preto, ____/____/____

Assinatura do autor