

Análise de interações simuladas da prova oral do exame Celpe-Bras

Analysis of simulated interactions of the oral test of the Celpe-Bras exam

Alex Sandro Beckhauser  

asbeckhauser@uefs.br

Universidade Estadual de Feira de Santana, BA, Brasil.

Rianne de Souza e Sousa  

riannesousas@gmail.com

Universidade Estadual de Feira de Santana, BA, Brasil.

Resumo

Este artigo tem o objetivo de analisar as justificativas das notas de avaliadores interlocutores e avaliadores observadores em um simulado da interação face a face do exame Celpe-Bras. Os avaliadores são 3 estudantes de cursos de licenciatura que atuam no Núcleo de Línguas do Idiomas sem Fronteiras da Universidade Estadual de Feira de Santana (NuLi-IsF/UEFS) em um curso preparatório para o exame Celpe-Bras. Cada professor realizou 4 entrevistas semiestruturadas ao final de cada interação simulada, a fim de garantir os encontros possíveis com cada um dos examinandos. Os resultados revelam um alto índice de semelhança na atribuição das notas, mas critérios de classificação distintos. Evidenciou-se também incertezas quanto à pontuação apresentada, ensejando um debate em torno da subjetividade da avaliação e da importância de refinar melhor os critérios para aumentar a confiabilidade do processo.

Palavras-chave: Celpe-Bras; Prova oral; Simulado.

Abstract

This article aims to analyze the justifications of the grades of interlocutor evaluators and observer evaluators in a simulated face-to-face interaction of the Celpe-Bras exam. The evaluators are 3 undergraduate students who work at the Nucleus of Languages without Borders of the State University of Feira de Santana (NuLi-Isf/UEFS) in a preparatory course for the Celpe-Bras exam. Each teacher conducted 4 semi-structured interviews at the end of each simulated interaction, in order to ensure possible meetings with each of the examinees. The results show a high level of similarity in grading, but distinct criteria of judgment in at least half of the meetings. The results reveal a high level of similarity in the attribution of grades, but different classification criteria. There was also uncertainty regarding the score presented, giving rise to a debate around the subjectivity of the evaluation and the importance of better refining the criteria to increase the reliability of the process.

FLUXO DA SUBMISSÃO

Submissão do trabalho: 26/12/2023

Aprovação do trabalho: 08/07/2024

Publicação do trabalho: 02/09/2024



10.23925/2318-7115.2024v45i4e64984

Distribuído sob Licença Creative Commons



Keywords: *Celpe-Bras; Oral exam; Simulated.*

1. Introdução

Os testes de proficiência alcançaram um enorme prestígio em nível mundial, tornando-se parte de uma política linguística de Estado. Seus efeitos retroativos produzem implicações de natureza política, social, econômica, pedagógica e linguística. É política quando Estados-nação obrigam a população imigrante a atestar um nível satisfatório de proficiência para obter acesso a determinadas políticas públicas; é social porque o grau de proficiência implica pertencer a uma coletividade e assim participar das decisões sobre suas próprias vidas; é econômica na medida em que os vários testes conformam um produto de uma economia de mercado; é pedagógica ao impactar os métodos aplicados e conteúdos a serem estudados; e é linguística porque controla e manipula os comportamentos linguísticos para determinar a prioridade de línguas específicas na sociedade e educação (Shohamy, 2006).

[...] os testes linguísticos são considerados poderosos porque são utilizados pelas autoridades centrais como ferramentas para impor o que será aprendido nas escolas, como será aprendido, quem é elegível para aprender e continuar a aprender e quais serão os critérios de correção. A escolha das línguas que estão sendo testadas na sociedade determina a sua importância. Aqueles que tomam estas decisões, os ministérios da educação e outras agências governamentais e grupos de autoridade, controlam o conhecimento e as classes da sociedade que determinam os critérios de desempenho e as medidas de sucesso, e tudo através de testes linguísticos. Assim, os testes linguísticos afetam a escolha das línguas, os métodos e conteúdos estudados, bem como os critérios de padrões e correção linguística (Shohamy, 2006, p. 105, tradução nossa).

No contexto brasileiro, o certificado de proficiência de língua portuguesa para estrangeiros (Celpe-Bras) tem incorporado algumas das características aludidas acima, tornando-se parte de uma política linguística estatal (Scaramucci, 2012), com crescente capilaridade em várias partes do mundo. Firmou-se como único dispositivo político-linguístico reconhecido pelo Governo nacional para aferir proficiência na variedade brasileira do português.

Qualquer pessoa interessada em realizar o exame Celpe-Bras deverá cumprir duas etapas de avaliação, a saber: prova escrita e oral. Com relação a esta última, denominada interação face a face, a avaliação é feita por duas pessoas, que ocupam o papel de examinador-observador (AO) e examinador-interlocutor (AI). O primeiro realiza um diagnóstico analítico do candidato, avaliando-

o sob seis critérios: interação, fluência, compreensão, pronúncia, adequação gramatical e adequação lexical. O segundo, por sua vez, conduzirá a interação durante 20 minutos, sendo os 5 primeiros dedicados a perguntas de cunho pessoal e os outros 15 minutos distribuídos entre três tópicos de ampla discussão na sociedade brasileira como, por exemplo, educação, cultura, saúde, tecnologia etc.

Assim como em todo teste com as características do Celpe-Bras, componentes subjetivos não estão totalmente isentos das etapas avaliativas, impactando diretamente na confiabilidade do processo. A esse respeito, Sakamori (2006) e Bottura (2014) nos dão uma amostra de como os avaliadores do Celpe-Bras elaboram e reelaboram as suas perguntas, como atuam e o que levam em conta quando estão avaliando um candidato, haja vista que a conduta dos avaliadores afeta o desempenho do examinando.

Considerando a importância das questões pontuadas e os poucos estudos a respeito da prova oral do supramencionado exame, o objetivo deste artigo é analisar as justificativas apresentadas pelos examinadores na atribuição das notas de avaliação da interação face a face durante um simulado do Celpe-Bras, realizado com os estudantes do Pré-PEC-G. Ao analisar as justificativas, traremos também algumas amostras sobre eventuais dificuldades na atribuição das notas, a fim de que possamos compreender em que medida os critérios adotados pelos avaliadores revelam uma subjetividade avaliativa.

A pesquisa foi realizada com três professores bolsistas que atuam no ensino de português como língua estrangeira (PLE) no Núcleo de Línguas do Idiomas sem Fronteiras da Universidade Estadual de Feira de Santana (NuLi-IsF/UEFS). Gravamos entrevistas semiestruturadas com cada um dos professores, tanto na condição de AO quanto de AI. O simulado foi realizado com seis alunos em setembro de 2023, que, à época, se preparavam para realizar o exame Celpe-Bras em outubro do referido ano.

2. O construto do Celpe-Bras

Em termos regulamentares, o Celpe-Bras tem uma história de 30 anos e, desde então, tem se firmado como um dos instrumentos da política linguística do Estado brasileiro (Scaramucci, 2012). Em 1993, foi instituída uma comissão para elaborar o exame, com o objetivo de atender às necessidades de seleção de estudantes que participariam do Programa de Estudantes Convênio

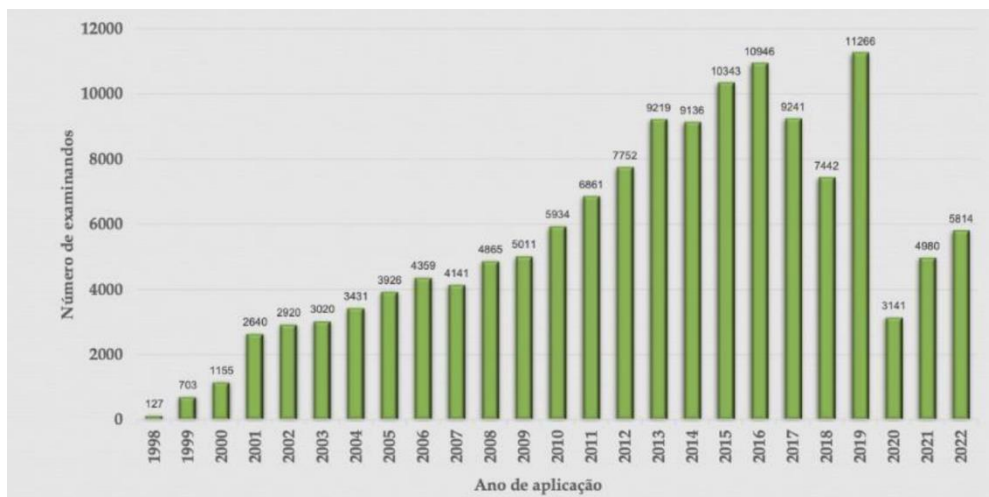
de Graduação (PEC-G). Em 1994, a Portaria nº 1.787 institucionaliza a iniciativa e, em 1998, é realizada a primeira edição em oito postos aplicadores (Brasil, 2020).

Além do PEC-G, a certificação no Celpe-Bras passou a ser requisito para o ingresso de estrangeiros no Programa de Estudantes Convênio de Pós-Graduação (PEC-PG), revalidação de diplomas de médicos formados no exterior e, mais recentemente, para o processo de naturalização (Anúnciação; Camargo, 2019).

Até 2008, a referida Comissão era responsável pela elaboração da prova. Porém, a partir do ano seguinte, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) assume a responsabilidade pela elaboração, aplicação, correção e divulgação dos resultados do Exame (Brasil, 2020). Além do Inep, o Ministério das Relações Exteriores participa indiretamente do processo, na medida em que auxilia instituições fora do Brasil a aplicarem a prova. Essa parceria entre o Inep e o MRE tem ajudado a consolidar o Celpe-Bras em outros países, viabilizando as condições para que a prova alcançasse um maior número de interessados.

O gráfico 1 nos dá uma dimensão da demanda pela prova ao longo do tempo.

Gráfico 1. Número de examinandos anuais do Exame Celpe-Bras



Fonte: Acervo Celpe-Bras.

Podemos observar uma curva ascendente, que expressa o aumento de pessoas interessadas em realizar a prova desde 1998. Não obstante, notamos também uma queda em 2017 e 2018, uma elevação no ano seguinte e novamente uma queda. Essa redução em comparação às edições anteriores pode ser justificada pelo fato de que até 2020 eram realizadas

duas aplicações anuais, vigorando, desde então, uma única edição. Outros motivos podem justificar o quadro atual, sobre os quais não nos debruçaremos no momento.

A respeito do construto do Celpe-Bras, cabe dizer que a prova está ancorada em uma proposta baseada em tarefas, fundamentada, dessa maneira, em "usar adequadamente a língua para desempenhar ações no mundo, em diferentes contextos, e sempre com um propósito social" (Brasil, 2020, p. 29). Em outras palavras, a tarefa implica, segundo Schlatter *et al* (2009), um enunciador, com um papel definido, que diz algo a alguém com propósito específico em um gênero discursivo.

Em consonância com as ideias do parágrafo anterior, afirma-se que:

O exame, por ser de natureza comunicativa, não busca aferir seus conhecimentos a respeito da Língua Portuguesa por meio de questões sobre gramática e vocabulário, mas sim avaliar sua capacidade de usar essa língua, independentemente das circunstâncias em que você a aprendeu. Sua proficiência será avaliada a partir de seu desempenho em tarefas e em uma interação face a face que exige compreensão escrita e/ou oral e produção escrita. Ou seja, práticas de uso da língua que possam ocorrer no cotidiano de um/a estrangeiro/a que pretende interagir em português. (Brasil, 2015, p. 9).

Ao considerar a autenticidade no processo avaliativo, o exame desconsidera os aspectos individualizados da língua e avalia, portanto, as competências linguísticas de forma integrada, exigindo do examinando "o conhecimento de práticas letradas diversas e específicas" (Anúnciação; Camargo, 2019, p. 15). Recursos lexicais e gramaticais mobilizados pelos examinandos na escrita e oralidade são avaliados em sua relação com o todo e não por meio artificial de sua manipulação.

O candidato deve demonstrar ser capaz de compreender e produzir textos de gêneros diversos (entendendo sua organização e propósito), além de posicionar-se e relacionar ideias, adequando a linguagem a situações diversas – capacidades essas não usualmente desenvolvidas pela escola mas fundamentais para a vida das pessoas e exercício da cidadania, que dependem, antes de mais nada, de letramento e não apenas de conhecimentos sobre a língua (Rodrigues, 2006, p. 54).

O exame Celpe-Bras está dividido em uma parte coletiva e outra individual (Dell'Isola *et al*, 2003). Com relação à segunda, objeto de nossa pesquisa, o candidato é submetido a uma avaliação oral com duração de vinte minutos. Nos 5 minutos iniciais, os candidatos têm a oportunidade de falar sobre si, com base nas respostas que foram preenchidas no formulário de inscrição. Esse momento inicial é utilizado para criar uma situação de interação menos tensa,

buscando dirimir a assimetria que há entre o AI e examinando (Brasil, 2020). Na segunda etapa, o AI conduzirá a interação com base em 3 materiais denominados elementos provocadores (EPs). Esses elementos são selecionados pelo interlocutor a partir de uma lista de 20 opções, utilizando, como critério, a trajetória de vida do candidato, isto é, a relação dos tópicos selecionados com a profissão, gostos, estudos e outras afinidades pontuadas pelo examinando no formulário de inscrição.

Cabe ao AI criar as condições de interação com o examinando por meio de perguntas sugeridas pelo manual de aplicação. Ao final, o AI atribui uma nota de 0 a 5 para o candidato de acordo com a grade de avaliação holística da proficiência oral do Celpe-Bras.

Em um processo de avaliação paralela, o AO atém-se a uma análise pormenorizada do candidato, buscando, como comentado na introdução deste trabalho, avaliá-lo à luz de seis critérios. Cada critério deve ser avaliado também de zero a cinco, cada um com descritores de desempenho específicos.

Vale destacar que não é esperado do examinando um desempenho nos moldes de um falante nativo ideal, pois sua proficiência apresenta níveis conforme “as necessidades de uso futuro da língua” (Brasil, 2020, p. 27).

O Celpe-Bras emite o certificado em quatro níveis de proficiência: intermediário, intermediário superior, avançado e avançado superior. No entanto, a prova a ser feita para qualquer um dos níveis é a mesma, haja vista que é por meio do uso real da língua portuguesa que será analisado o desempenho de cada candidato.

3. Metodologia

Esta é uma pesquisa de abordagem qualitativa de cunho interpretativista. Nesse tipo de abordagem, segundo Crocker (2009), os pesquisadores se posicionam próximos aos participantes, no sentido de ver o mundo a partir de sua ótica. Criam a sensação de entrar no mundo desses sujeitos e compartilhar a experiência de estar lá com eles.

Os participantes da pesquisa são 3 professores bolsistas que atuam no ensino de PLE na UEFS pelo Programa Pré-PEC-G. Esses três professores realizam atividades que visam ao desenvolvimento da competência comunicativa e ao preparatório para o Celpe-Bras. Uma das

etapas de preparação é justamente um simulado de interação face a face, realizado em setembro de 2023, constituindo-se o objeto de nosso estudo.

Utilizamos como técnica de coleta de dados 12 entrevistas semiestruturadas, que foram realizadas com cada um dos professores, no papel de avaliadores, após o término de cada simulado de interação. As entrevistas tiveram uma duração de até 8 minutos. Conforme Triviños (1987, p. 146):

Podemos entender por entrevista semi-estruturada, em geral, aquela que parte de certos questionamentos básicos, apoiados em teorias e hipóteses, que interessam à pesquisa, e que, em seguida, oferecem amplo campo de interrogativas, fruto de novas hipóteses que vão surgindo à medida que se recebem as respostas do informante. (Triviños, 1987, p. 146).

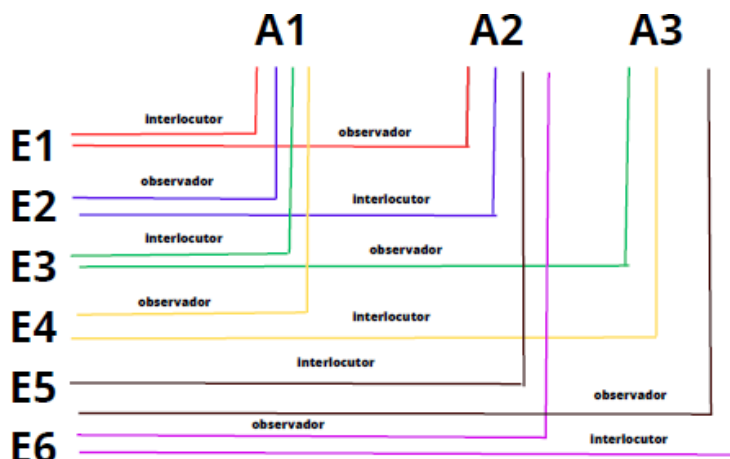
Para as entrevistas, elaborou-se um mesmo roteiro de perguntas, utilizado tanto para o AI quanto para o AO. Embora este segundo informante tenha uma grade de avaliação fragmentada com notas de 0 a 5 em cada um dos critérios, as perguntas da entrevista foram elaboradas com o objetivo de analisarmos eventuais subjetividades avaliativas, conforme Quadro 1:

Quadro 1. Roteiro de perguntas aos avaliadores do simulado do Celpe-Bras

Perguntas ao avaliador-interlocutor e ao avaliador observador
<ol style="list-style-type: none">1. Com base em sua avaliação, que nota você atribuiria ao futuro candidato? Por quê?2. Que aspecto da grade de avaliação você considerou mais importante para a atribuição da nota?3. Em algum momento, você teve dúvida quanto à nota atribuída ao futuro candidato(a)? Por quê?

Fonte: os autores.

Durante a etapa do simulado, cada professor foi entrevistado 4 vezes: em duas ocasiões como AI e em duas como AO. Essa configuração permitiu que todos os encontros se tornassem possíveis e assim evitar que um mesmo avaliador-interlocutor repetisse esse papel com o mesmo avaliador-observador. A figura 1 nos dá uma amostra da configuração dos encontros e, conseqüentemente, das entrevistas feitas ao final de cada interação simulada.

Figura 1. Configuração das situações de interação do simulado do Celpe-Bras**Fonte:** os autores.

Respeitando os princípios éticos da pesquisa, os avaliadores e examinandos foram identificados por letras e números. No caso dos informantes da pesquisa, o avaliador 1 assumiu o papel de AI 1 e AO 1; o avaliador 2, o papel de AI 2 e AO 2; e o terceiro avaliador, o papel de AI 3 e AO 3, conforme configuração apresentada no quadro 2. Destaca-se que o estudo foi aprovado pelo Comitê de Ética e Pesquisa sob o número 71249523.3.0000.0053.

Conforme observado na Figura 1, foram consideradas 6 simulações da parte oral. Dos alunos que participaram do simulado, 3 são de países anglófonos e 3 são de países francófonos. Eles foram os únicos aprendizes no curso de PLE ofertado pela UEFS, e compuseram o conjunto ideal para todas as situações de interação.

4. Atribuição do nível pelos avaliadores

Conforme comentado, um mesmo par de avaliadores se encontrou em duas ocasiões, ocupando papéis distintos nas interações. Na primeira pergunta das entrevistas, o AI e o AO tinham que atribuir uma nota para o examinando e explicar o motivo. O quadro abaixo sintetiza as classificações atribuídas aos examinandos durante o simulado da interação face a face.

Quadro 2. Atribuição dos níveis pelos AI e AO

EXAMINANDOS	AVALIADOR 1		AVALIADOR 2		AVALIADOR 3	
	AI 1	AO1	AI 2	AO 2	AI 3	AO 3
E1	3			4		
E2		3	4			
E3				1	1	
E4			4			4
E5	4					4
E6		4			4	

Fonte: os autores.

O quadro 2 elucida informações importantes sobre uma das etapas do processo avaliativo do exame Celpe-Bras. Conforme dissemos acima, a avaliação da supracitada prova está desenhada em uma escala de desempenho de 0 a 5, sendo as notas de 0 e 1 atribuídas aos examinandos considerados não proficientes, a nota 2 caracterizada como nível intermediário, a 3 como intermediário superior, a 4 para o nível avançado e a nota 5 avançado superior.

Observamos divergências nas notas atribuídas pelos avaliadores ao E1 e E2. Na primeira delas, o AI 1 atribuiu nota 3 e o AO 2 nota 4. De acordo com o Manual do candidato (Brasil, 2020), a discrepância em evidência é considerada aceitável, no sentido de que não haveria a necessidade de uma reavaliação do desempenho oral do participante. Se necessário fosse, outros dois avaliadores entrariam em cena para avaliar o desempenho oral do examinando.

Quanto às justificativas, o AI 1 destaca uma melhor desenvoltura do examinando nos primeiros 10 minutos da interação, ao falar de si e no primeiro elemento provocador:

AI 1: *“Principalmente nas duas primeiras entrevistas, na conversa livre sobre os interesses dele e no primeiro elemento provocador, mas talvez 2, alguma coisa entre 2 e 3.”*

Com base na resposta, é possível inferir que, à medida que outros EPs foram introduzidos na interação, o examinando reduziu seu desempenho, implicando aumento de inadequações.

Fatores de natureza individual e/ou operacional podem ter influenciado o desempenho. É individual quando o candidato parece não reunir as condições linguístico-comunicativas para interagir em português. É operacional quando algum aspecto da organização da prova dificulta o desempenho. Sobre esse último, podemos citar mudanças contextuais como, por exemplo, a seleção do tópico (Niederauer, 2014) e sua incoerência em relação à imagem do elemento provocador (Faria; Gattolin, 2021).

Ao ser perguntado se em algum momento teve dúvida quanto ao nível atribuído ao futuro candidato e por quê, o AI 1 apresentou a seguinte justificativa:

AI 1: "Sim, exatamente porque ele tem uma boa fluência e compreendeu muito bem, mas a pronúncia em alguns momentos e a interrupção do fluxo da conversa para buscar palavras me deixou com dúvidas."

A resposta do AI 1 apresenta uma contradição argumentativa no critério fluência. Ao passo que a alega positivamente para não reduzir a nota, afirma que o examinando produz interrupção no fluxo da fala, razão pela qual essa medida de desempenho pesa na atribuição da pontuação. De acordo com a grade holística do Celpe-Bras, o examinando considerado nota 2 demonstra algum grau de dificuldade no fluxo natural da fala, o que não parece ser o caso do E1. Por esse descritor, a dúvida do avaliador estaria resolvida. A nota 3, por outro lado, é atribuída ao candidato com alguma interrupção no fluxo da fala, respaldando a classificação atribuída pelo AI 1 (Brasil, 2020).

Percebemos também o impacto da pronúncia na nota atribuída pelo AI 1, que a justificou com base em algumas inadequações do E1. O descritor referente à pronúncia na grade holística parece respaldar a pontuação aferida, haja vista que o examinando realizou algumas inadequações.

Ainda referente ao E1, o AO 2 atribuiu-lhe nota 4 sob a seguinte alegação:

AO 2: "Porque eu percebi que ele conseguiu fazer o que tava proposto, conseguiu interagir com o que AI 1 trouxe. Percebi que ele conseguiu colocar o diálogo pra frente, que foi o principal objetivo da prova."[...] "A parte interação e compreensão pesaram mais na minha escolha."

Observa-se que o AO 2 considera dois critérios como os mais importantes para a nota 4 do E1: interação e compreensão. A partir da resposta, pressupõe-se que a interação tem um peso muito significativo para o AO 2, de tal maneira que prováveis inadequações em outros critérios

não pesaram proporcionalmente.

Ao ser questionado se teve alguma dúvida quanto ao nível atribuído ao E1 e o porquê, o AO 2 expressou a seguinte justificativa:

AO 2: *“Eu tive dúvidas com relação a quando atribuía 3 ou 4.” [...] “Essas dúvidas surgiram porque ele tinha algumas inadequações, como eu falei, mas, às vezes, eu pensava que podia ser algo natural, pontual e não da fluência.”*

Na resposta, vemos um momento de dúvida do AO 2, ou seja, se o E1 deveria receber nota 3 ou 4. O avaliador considerou, portanto, que as inadequações poderiam ser algo pontual e/ou natural do fluxo da conversa. O AO 2 entendeu que algumas inadequações apresentadas não prejudicaram a fluência, mas se tornaram relevantes para não atribuir nota máxima ao candidato.

O simulado realizado com o E2 mantém os mesmos avaliadores, porém em papéis opostos. Novamente, os mesmos avaliadores atribuem notas distintas ao examinando, conforme apontado no Quadro 2. O AI 2 justifica sua nota com base no critério de interação:

AI 2: *“Ele contribuiu para o diálogo, mas o que eu percebi foi a questão mais de respostas um pouco mais breves [...] Mas eu percebi bastante fluência com relação ao uso da língua.”*

Retomando a grade de avaliação holística, o único diferencial entre as notas 4 e 5 para a interação é o acréscimo do advérbio “muito”, na descrição do desempenho. A nota máxima é para o examinando que “contribui muito para o desenvolvimento da interação” (Brasil, 2020, p. 52), o que não parece ser o caso do E2, com base no argumento do interlocutor.

Respostas breves podem ser um indicativo de estilos conversacionais entre quem avalia e quem se submete à avaliação. Há uma tendência de a pessoa avaliada transferir tais estilos de sua primeira língua para a língua estrangeira (Niederauer, 2014). Muitas vezes, o examinando alcança um desempenho semelhante àqueles com respostas longas, mas é penalizado por apresentar respostas breves (Young, 2000). Essas respostas podem ser também um reflexo da dinâmica de interação do avaliador, que acaba influenciando as tomadas de decisão do examinando (McNamara, 1997).

Neste momento, cabe retomar Niederauer (2014), segundo a qual aspectos da competência interacional como, por exemplo, respostas breves carecem de definições nas áreas de estudos dedicadas à interação face a face. A nosso juízo, a ausência de definições impacta nas tomadas de decisão do avaliador, estando mais suscetível a subjetividades.

Ao ser questionado se teve alguma dúvida quanto ao nível atribuído ao examinando, o AI

2 sinalizou apenas que se sentiu mais confortável em atribuir nota com base nos critérios da grade holística do que da analítica. Consideramos que é curiosa a justificativa, haja vista que a descrição do desempenho da segunda é mais completa do que a da primeira, pois fornece descritores mais objetivos.

Com relação à segunda avaliação da mesma interação face a face, o AO 1 fez a seguinte afirmação para justificar a nota 3 do E2:

AO 1: *"Porque eu tentei fazer a média de todas".*

O AO 1 evidencia uma justificativa com base no peso relativo dos critérios utilizados na grade analítica. Esse avaliador considerou principalmente a compreensão, adequação lexical e fluência como os critérios que mais pesaram positivamente na nota atribuída, atendendo ao que está proposto no Documento Base do Celpe-Bras (Brasil, 2020). Por outro lado, embora não mencionada explicitamente, podemos inferir que a interação, a exemplo AI 2, também foi determinante para reduzir a pontuação de E2.

Ao ser questionado se em algum momento teve dúvida quanto ao nível atribuído ao examinando e o porquê, o AO 1 parece bastante inseguro em sua pontuação:

AO 1: *"Sim, em todos os momentos, porque é um processo contínuo, a gente está vendo quatro momentos e o candidato pode variar nessas competências de acordo com o tema ou outras variáveis e foi isso que aconteceu."*

Esse avaliador destaca a influência dos EPs no percurso avaliativo. Algumas pesquisas têm apontado que a escolha do tema exerce um papel importante na desenvoltura do candidato durante a prova, suscitando dúvidas quanto ao nível de proficiência do examinando (Faria; Gattolin, 2021) ou criando uma situação de desconforto para o candidato, que pode se recusar a interagir quando o tema não lhe agrada (Schoffen, 2003). Faria e Gattolin (2021) destacam a importância de selecionar EPs que suscitem discussões e estimulem a produção oral do examinando em diversos contextos.

Os examinandos mencionados a seguir tiveram as mesmas notas atribuídas, tanto pelos AI quanto pelos AO. No que tange às entrevistas feitas após o simulado com o E3, O AI 3 justificou sua nota com base nas interrupções e as muitas pausas feitas pelo examinando. Esses descritores afetam a fluência na língua estrangeira, tornando-se determinantes para a nota 1 dada ao E3. De

acordo com o Documento base do exame Celpe-Bras (Brasil, 2020), o examinando com essa pontuação é considerado não proficiente e, portanto, não recebe certificação de comprovação de nível.

O argumento apresentado pelo A1 3 parece corroborar outros estudos a respeito da sensibilidade dos avaliadores no critério fluência, no sentido de que o bom desempenho nesse critério conduz a classificações mais altas (Duijim; Schoonem; Hulstjin, 2018).

Interrupções no fluxo da fala podem ser consideradas um descritor objetivo de avaliação e podem ser medidas quantitativamente, razão pela qual tendem a reduzir o grau de subjetividade do avaliador (Duijim; Schoonem; Hulstjin, 2018). Amostras de fala manipuladas foneticamente com falantes nativos e não-nativos revelam que os avaliadores tendem a igualar a pontuação quando pausas e velocidades de fala também são igualadas (Bosker *et al*, 2014). A nosso juízo, observar e avaliar a fluência, à luz desses elementos objetivos, é uma estratégia para reduzir a subjetividade da avaliação e assim garantir simetria nas pontuações.

O AO 2, que também atribuiu a mesma nota para o E3, justificou-a com base no seguinte argumento:

AO 2: *"Porque ele apresentou diversas inadequações com relação à língua portuguesa em todas as categorias que temos de análise".*

A justificativa baseada na inadequação suscita ampla interpretação, haja vista que três componentes da grade analítica a mencionam como descritor avaliativo: léxico, gramática e pronúncia. O AO 2 usa o advérbio "diversas" para classificar o grau de inadequações cometidas pelo E3 e assim respaldar o nível atribuído. Esse advérbio se assemelha à descrição do desempenho do examinando com nota 0 e 1, que, de acordo com a grade analítica, é aquele que apresenta "muitas inadequações de estruturas básicas e complexas" ou "vocabulário muito inadequado" (Brasil, 2020, p. 50-51). O termo "inadequado" consta também no descritor avaliativo pronúncia, porém este não foi mencionado pelo AO 2 para atribuição da nota.

Conseguimos ter clareza em relação às dúvidas acima quando perguntado que aspecto da grade de avaliação AO 2 considerou mais importante para a atribuição do nível:

AO 2: *"Adequação gramatical, porque eu acho que a estrutura gramatical seria o que mais daria sentido ao que o participante está falando."*

Podemos observar a gramática como um componente decisivo para o mencionado avaliador. Segundo o AO 2, um bom desempenho está condicionado à capacidade de o examinando mobilizar recursos gramaticais de maneira eficaz para, então, produzir sentido. Em outro contexto investigativo, profissionais que atuam com exames de proficiência também focaram em inadequações gramaticais para avaliar a oralidade em língua estrangeira (Duijim; Schoonem; Hulstjin, 2017).

Cumpre-nos lembrar que a avaliação da competência gramatical e lexical não pode levar em conta o somatório de inadequações sem relação com o contexto, mas sim o impacto desses recursos na interação (Brasil, 2020). Em um estudo com hispanofalantes, Schoffen (2003) concluiu que o léxico é um dos principais critérios definidores do nível de proficiência de um candidato e, por isso, precisa-se um olhar cuidadoso sobre o modo como o examinador o avalia na interação face a face. Ao comparar o critério de AO 2 com a avaliação do AI 3, observamos uma subjetividade avaliativa..

Na quarta interação face a face, os examinadores coincidem também no nível atribuído. O AI 2 justifica a nota 4 com base no seguinte argumento:

AI 2: "Porque eu percebi que houve uma compreensão dos tópicos abordados e que houve uma interação que evoluiu [...] Porém, eu vi que o participante teve momentos de pensar bastante no que falar e isso, às vezes, dificultava um pouco".

Ao usar a expressão "teve momentos de pensar", o AI 2 faz alusão a interrupções no fluxo da fala, que, como já vimos, é considerado um componente central para avaliar a fluência. Novamente, uma variável quantitativa foi decisiva para a atribuição da pontuação, impedindo o E4 de alcançar a nota máxima.

É curioso observar que a resposta do AI 2 não contempla o vocabulário como um critério positivo para validar a nota atribuída. Só teremos ciência dessa informação quando perguntado qual critério da grade o avaliador considerou mais importante para a atribuição do nível:

AI 2: "Vocabulário, porque eu percebi que ele conseguiu falar dos assuntos mesmo que eles não fossem tão cotidianos."

Esse mesmo avaliador foi questionado se em algum momento teve dúvida quanto à classificação atribuída ao examinando e por quê. Sua resposta não faz referência a um critério

específico, mas a um fenômeno anteriormente observado, na justificativa apresentada pelo AO 1 na interação com o E2:

AI 2: *"Sim, porque pensei em comparação com outros candidatos e aí eu tive algumas dificuldades, mas nada com relação à grade."*

O supracitado avaliador traz à baila a comparação avaliativa, utilizando-a como um recurso para tentar manter a isonomia do resultado e assim diminuir a subjetividade. Esse recurso, por sua vez, não foi considerado pelo seu par homólogo, o AO 3, que pontuou sua incerteza entre atribuir nota 4 ou 5 ao E4 e informa que usou os descritores da grade analítica para sustentar o resultado final:

AO 3: *"Tive dúvida nas primeiras partes aqui, fiquei na dúvida se 5 ou se 4. Mas aí a minha dúvida foi tirada pelas palavras que eles colocam aqui na grade." [...] "Porque considerando o quadro geral de tudo que ele falou em alguns momentos poderia ser 5, mas no final houveram algumas pausas maiores."*

De acordo com a resposta, a fluência também teve um peso significativo na atribuição da nota, coincidindo com a justificativa do AI 2. Esse critério serviu como base para inviabilizar a nota máxima ao E4, haja vista algumas pausas no fluxo natural da fala.

Nas duas últimas interações face a face, feitas com E5 e E6, o critério interação foi avaliado positivamente pelo AI 1 e AI 3 respectivamente. A fluência foi mencionada também como um critério importante para a pontuação, embora tenhamos identificado uma contradição na justificativa do AI 1 para o E5:

AI 1: *"Porque ela demonstrou autonomia, desenvoltura, contribuiu bastante para o momento da interação. Eu acho que ela apresentou fluência e um vocabulário muito bom também, alguns momentos até específicos. Teve algumas inadequações e alguns espaços vazios para pensar, mas 4."*

AI 3: *"Porque considerando que o nível 4 é aquele que demonstra autonomia e desenvoltura que contribui para o desenvolvimento da conversa, com poucas inadequações, eu acho que esse é o que mais se encaixa devido ao desempenho do candidato."*

Conforme observado, o AI 1 não está completamente seguro do desempenho da fluência de E5, embora a tenha citado como um elemento para garantir uma pontuação alta e, concomitantemente, a considerou relevante para não lhe atribuir nota 5 ao dizer que o examinando teve "alguns espaços vazios para pensar".

O AI 3, por outro lado, não faz alusão direta aos descritores relacionados à fluência, mas a considera importante para a atribuição da nota. Temos clareza disso quando perguntado que elemento da grade de avaliação ele julgou importante para a pontuação.

Outro dado interessante a ser analisado é o das inadequações de E5 e E6. A despeito de ambos os examinandos terem recebido a mesma pontuação, o grau de suas inadequações foi distinto, razão pela qual ensejaria alteração em uma das notas. De acordo com a grade holística, o examinando com nota 4 é aquele que apresenta “poucas inadequações”, o que respalda a justificativa de AI 3, mas não a de AI 1.

Esses mesmos avaliadores no papel de observadores julgam igualmente os mesmos critérios como centrais para a classificação de E5 e E6, isto é, compreensão e adequação lexical. O AO 3 não apresentou detalhes dos elementos que considerou pertinente para não atribuir nota máxima ao E5. O AO 1, por outro lado, mencionou uma situação envolvendo a compreensão:

AO 1: “Porque ele teve uma fluência muito boa, compreendeu todas as perguntas, apenas pediu para reestruturar uma, o vocabulário também muito especial, contribuiu muito para a interação.”

O avaliador não nos fornece muitas evidências do que seria reestruturar uma pergunta, de tal maneira que não é possível afirmar se o evento em tela, efetivamente, corrobora algum problema na compreensão do fluxo natural da fala. Bottura (2014) demonstra que há situações no Celpe-Bras em que o avaliador realiza perguntas longas e confusas, aliadas a reformulações desnecessárias que dificultam a compreensão do examinando.

Ao ser questionado se em algum momento teve dúvida quanto à nota atribuída, o AO 3 responde positivamente e justifica sua resposta com base em uma eventual interferência de outra língua estrangeira no E6:

AO 3: “Eu tive dúvida por causa da questão da influência de outra língua na parte gramatical. Em alguns momentos, veja só, ele utiliza um vocabulário que poderia ter sido de outra língua. [...] E também porque o sotaque português, de outras variedades do português, também me deixaram assim: será que ele está falando dessa forma por causa do português de Portugal [...] ou por causa do inglês?”

A situação em tela nos remete a contextos de avaliação feitos com falantes de espanhol, no sentido de que a proximidade tipológica com o português enseja momentos em que o avaliador

pode suspeitar que alguns léxicos mobilizados durante a interação não são propriamente resultados de interferência, mas sim conhecimento da língua portuguesa.

Essa suspeita pode ser provocada por algum traço fonético que remete à língua de origem, que o examinando não adquiriu, mas nem por isso inviabiliza a utilidade do léxico na interação. Nesse sentido, distinções fonéticas para as mesmas palavras em ambas as línguas precisam ser mais bem analisadas, de modo que não sejam constantemente confundidas como inadequações lexicais. Estamos de acordo com Henriques, Barros e Madureira (2020, p. 791), segundo as quais:

[...] critérios de avaliação embasados foneticamente precisam ser testados em trabalhos de pesquisa de natureza experimental de tal modo a fornecer subsídios aos desenvolvedores de testes dessa natureza e gerar instrumentos aos avaliadores que os auxiliem a julgar a parte oral.

Com base no exposto, aperfeiçoar os descritores fonéticos nas grades avaliativas poderá reduzir o grau de subjetividade dos avaliadores e assim ampliar o grau de confiabilidade do teste. A subjetividade avaliativa identificada em outros estudos (Ferreira, 2018; Hayes-Harb; Hacking, 2015; Schoffen, 2003) precisa ser mais bem compreendida, em especial na interação face a face do Celpe-Bras.

Considerações finais

Este trabalho teve como objetivo analisar as justificativas das notas atribuídas pelos avaliadores-interlocutores e avaliadores-observadores em um simulado da prova oral do Celpe-Bras, a fim de que pudéssemos compreender em que medida os critérios considerados na atribuição das notas revelaram alguma subjetividade avaliativa.

De acordo com Neves (2018), o papel do entrevistador exerce importância fundamental no processo de aplicação do exame Celpe-Bras e também na atribuição de notas, haja vista que é esse ator que conduz a interação e avalia o candidato. Assim, tendo em vista que quem avalia possui um importante papel na indicação das notas, o problema que buscamos responder foi: como os professores que preparam os estudantes para o exame Celpe-Bras classificam a proficiência desses futuros candidatos?

Nossos resultados mostraram que, das seis interações simuladas, houve apenas duas discrepâncias nas notas, nos encontros entre o EI 1 e EO2 e EI2 e EO 1. Em um contexto real do

exame Celpe-Bras, a discrepância seria considerada aceitável, não havendo a necessidade de uma reavaliação de desempenho. Em que pese a diferença assinalada, os avaliadores coincidiram em quase todos os critérios adotados para justificar a pontuação, demonstrando pouca margem para a subjetividade avaliativa.

Observando todas as notas, podemos concluir que o índice de assimetria foi de 33%. Cumpre esclarecer, no entanto, que esse quantitativo não pode servir como parâmetro, pois nossas amostras são pouco representativas. Nossos dados convidam outros pesquisadores a reproduzirem o método com outra amostra populacional, a fim de averiguar as aproximações e dissonâncias em relação ao nosso trabalho.

Um olhar global sobre as justificativas na atribuição das notas nos permite observar poucas discrepâncias, razão pela qual o simulado demonstrou ser confiável. De todos os critérios, a fluência exerceu um peso bastante significativo nas pontuações, positiva e negativamente. Alguns examinandos que deixaram de alcançar a pontuação máxima foram “penalizados” por realizar algum grau de interrupção no fluxo da conversa. Esse mesmo critério, em determinadas situações, enseja uma reflexão mais cuidadosa, haja vista que foi identificado um paradoxo na avaliação. Enquanto justificado positivamente para o bom desempenho, foi avaliado negativamente por um mesmo avaliador.

A interação foi outro critério amplamente mencionado, em especial nos contextos de notas mais altas. De modo geral, os examinandos contribuíram para o desenvolvimento da interação, o que pode ter garantido maior simetria avaliativa e, assim, evitar discrepância nas classificações. Independente do quadro que se apresenta, estamos convencidos de que os descritores das grades avaliativas merecem constantes atualizações, de modo a garantir maior confiabilidade do teste, reduzindo eventuais dúvidas nas pontuações, conforme demonstrado em nossa pesquisa.

A despeito de os avaliadores, em sua maioria, terem apresentado simetrias nos critérios utilizados para a atribuição das notas, o fato é que quase todos tiveram dúvidas no momento de pontuar os examinandos. Três casos, em especial, merecem atenção porque foram explicitados nas amostras: referimo-nos ao AI 1 e AO 2 durante a interação com E1 e AO 3 no simulado feito com o E3. Depreendemos desse fato não propriamente uma subjetividade no processo avaliativo, mas o que comentamos nos parágrafos acima, isto é, a importância de atualizações nas grades avaliativas, de modo que os descritores possam ser o mais claro possível.

Finalizamos o trabalho retomando outro aspecto identificado em uma das entrevistas, a saber, a avaliação comparativa. Classificar um candidato na prova oral do Celpe-Bras comparando-o com interações precedentes pode colocar em contestação a confiabilidade da prova, haja vista que os primeiros entrevistados poderão ser prejudicados.

Referências

- ANUNCIÇÃO, R. F. M. de. CAMARGO, H. R. E. de. O exame Celpe-Bras como política gatekeeping para a naturalização no Brasil. **Muiraquitã: Revista de Letras e Humanidades**, São Paulo, v. 7, n. 2, p. 10-22, 2019. Disponível em: <https://teste-periodicos.ufac.br/index.php/mui/article/view/2764> . Acesso em: 20 de jul. 2023
- BOSKER H. R. et al. The perception of fluency in native and nonnative speech. **Language Learning**, Michigan, v. 64, n. 3, p. 579–614, 2014. Disponível em: <https://onlinelibrary.wiley.com/doi/epdf/10.1111/lang.12067> . Acesso em: 07 nov. 2023.
- BOTTURA, E. B. Exame Celpe-Bras: uma investigação sobre o papel do entrevistador na interação face a face. 2014. **Dissertação** (Mestrado em Linguística) - Universidade Federal de São Carlos, São Carlos, 2018.
- BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira — Inep. **Certificado de proficiência em Língua Portuguesa para Estrangeiros — Manual do Examinando**. Brasília: MEC/Inep/Daeb, 2015. Disponível em: <http://www.ufrgs.br/acervocelpebras/arquivos/manuais/manual-do-examinando-2015> . Acesso em: 24 nov. 2023.
- BRASIL. Ministério da Educação. Secretaria de Ensino Superior. **Documento base do exame Celpe-Bras**. Brasília, 2020. Disponível em: <https://www.ufrgs.br/acervocelpebras/wp-content/uploads/2021/12/Documento-base-do-exame-Celpe-Bras-2020.pdf> Acesso em: 21 de abril de 2023.
- BRASIL. Celpe-Bras. **Histórico**. Disponível em <https://www.gov.br/inep/pt-br/areas-deatuacao/avaliacao-e-exames-educacionais/celpe-bras/historico> Acesso em 31 out. 2023.
- CROCKER, R. A. An introduction to qualitative research. In: HEIGHAM, J.; CROCKER, R. A. **Qualitative research in Applied Linguistics: a practical introduction**. London: Palgrave Macmillan, 2009, p. 3-24.
- DELL'ISOLLA, R. L. P. et al. A avaliação de proficiência em português língua estrangeira: o exame CELPE-Bras. **Rev. Brasileira de Linguística Aplicada**, v. 3, n. 1, p. 153-184, 2003. Disponível em: <https://www.scielo.br/j/rbla/a/3NJgKW795HBhL7zbTNQrTWz/?format=pdf&lang=pt> . Acesso em: 14 nov. 2023.
- DUIJIM, K.; SCHOONEM, R.; HULSTJIN, J. H. Professional and non-professional raters' responsiveness to fluency and accuracy in L2 speech: An experimental approach. **Language Testing**, v. 35, n. 4, p. 501-527, 2018. Disponível em: <https://journals.sagepub.com/doi/epub/10.1177/0265532217712553> . Acesso em: 07 nov. 2023

- FARIA, E. R.; GATTOLIN, S. R. B. Avaliação da proficiência oral em língua portuguesa como língua estrangeira: um olhar para a interação face a face do exame Celpe-Bras. **Revista Linguagem**, São Carlos, v. 39, n.1, p. 213-233, 2021. Disponível em: http://www.ileel.ufu.br/anaisdosielp/wp-content/uploads/2014/07/volume_2_artigo_168.pdf. Acesso em: 29 de abril de 2023.
- FERREIRA, L. M. L. Avaliação da proficiência oral: uma análise fatorial e de discriminação dos itens do exame Celpe-Bras. 2018. **Tese**. (Doutorado em Estudo de Linguagens) – Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2018.
- HAYES-HARB, R.; HACKING, J. F. Beyond rating data: What do listeners believe underlies their accentedness judgments? **Journal of Second Language Pronunciation**, Utah, v. 1, n. 1, p. 43 -64, 2015. Disponível em: <https://benjamins.com/catalog/jslp.1.1.02hay> Acesso em: 30 out. 2023
- HENRIQUES, A. S.; BARROS, T.; MADUREIRA, S. Proficiência oral em português brasileiro: cenário, contextos de avaliação e de instrução de questionamentos. **Revista da Abralim**, v. 19, n. 3, p. 775-798, 2020. Disponível em: <https://revista.abralin.org/index.php/abralin/article/view/1755/1963> . Acesso em: 14 nov. 2023
- MCNAMARA, T. F. ‘Interaction’ in second language performance assessment: whose performance? **Applied Linguistics**, v. 18, n. 4, p. 446-466, 1997. Disponível em: <https://academic.oup.com/applij/article-abstract/18/4/446/206309?redirectedFrom=fulltext> Acesso em 01 set. 2023.
- NEVES, L. O. Confiabilidade e comportamento avaliativo na prova oral do exame Celpe-Bras: um estudo longitudinal. 2018. **Tese**. (Doutorado em Estudos de linguagens) - Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2018.
- NIEDERAUER, M. Competência interacional: critério para avaliação da produção oral em língua adicional. **Trabalhos em Linguística Aplicada**, Campinas, v. 53, n. 2, p. 403-424, jul./dez. 2014. Disponível em <https://www.scielo.br/j/TLA/a/TTg8q4tcQy6PKx7sJcyjTNN/> . Acesso em 30 out. 2023
- RODRIGUES, M. S. A. O Exame Celpe-Bras: reflexões teórico-didáticas para o professor de português para falantes de outras línguas. 2006. **Dissertação**. (Mestrado em Linguística Aplicada). Universidade Estadual de Campinas, Campinas, 2006.
- SAKAMORI, L. A atuação do entrevistador na interação face a face do exame Celpe-Bras. 2006. **Dissertação** (Mestrado em Linguística Aplicada) - Universidade Estadual de Campinas, Campinas. 2006.
- SCARAMUCCI, M. V. O Exame Celpe-Bras e a proficiência do professor de português para falantes de outras línguas. **Revista Digilenguas**, Córdoba, v. 12, p. 48-67, 2012. Disponível em: <https://www.lenguas.unc.edu.ar/uploads/DigilenguasN12.pdf#page=50> . Acesso em: 14 nov. 2023.
- SCHLATTER, M. *et al.* Celpe-Bras e Celu: impactos da construção de parâmetros comuns de avaliação de proficiência em português e em espanhol. SCHLATTER, Margarete *et al* (org). **O português do Brasil como língua transnacional**. 1ª ed. Campinas: RG Editora, Campinas, 2009, p. 195-222.
- SCHOFFEN, J. R. A avaliação de proficiência oral em língua estrangeira: descrição dos níveis de

candidatos falantes de espanhol no exame Celpe-Bras. 2003. **Dissertação.** (Mestrado em Estudos de Linguagens) - Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2003.

SHOHAMY, E. **Language policy: hidden agendas and the new approaches.** Routledge: Nova York, 2006.

TRIVIÑOS, A. N. S. **Introdução à pesquisa em ciências sociais.** 1. ed. São Paulo: Atlas, 1987.

YOUNG, R. Sociolingüística de la adquisición de segundas lenguas. In: PRESTON, D. R.; YOUNG, R. **Adquisición de segundas lenguas: variación y contexto social.** Tradução por Francisco Moreno Fernández. Madrid: Arco Libros, 2000, p. 17-38