



Um teste de verificação lexical de português brasileiro (TVLPB) como língua adicional: criação e validação

A Vocabulary Level Test for Brazilian Portuguese as a Second Language: Design and Validation

Thaís Máira Machado de Sá

Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Belo Horizonte, Minas Gerais / Brasil

thaismaira@gmail.com

<https://orcid.org/0000-0003-4828-4340>

Alexandre Alves Santos

University of Massachusetts Amherst (UMass-Amherst), Massachusetts / Estados Unidos da América

alexandresan@umass.edu

<https://orcid.org/0000-0003-0074-2189>

Ricardo Augusto de Souza

Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais / Brasil

ricsouza.ufmg@gmail.com

<https://orcid.org/0000-0001-6690-3948>

Luiz Amaral

University of Massachusetts Amherst (UMass-Amherst), Massachusetts / Estados Unidos da América

amaral@umass.edu

<https://orcid.org/0000-0001-6007-6603>

Victor Nascimento Almeida

Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais / Brasil

vnavictor92@gmail.com

<https://orcid.org/0000-0003-4493-1424>

Resumo: Dado o aumento de falantes de português como língua adicional, de acolhimento e de herança, é de grande importância ter uma ferramenta padronizada de medida de proficiência que seja de fácil e rápida administração. Neste trabalho, mostramos a criação e validação interna de um Teste de Verificação Lexical do Português Brasileiro (TVLPB). O teste foi criado com cinco bandas de frequência que correspondem à distribuição de frequência lexical encontrada em diferentes corpora. Os resultados com falantes nativos mostram convergência de respostas em relação às rubricas selecionadas e às bandas de frequência utilizadas, indicando uma validade interna. O teste foi aplicado também com falantes de herança, em que também percebemos um aumento da dificuldade em relação às bandas do teste.

Palavras-chave: VLT; português como língua adicional; teste de proficiência; avaliação.

Abstract: Given the growing number of speakers of Portuguese as a host, second, and heritage language, a proficiency level test that can be easily and reliably administered is necessary. In this work, we show the design and validation of the Brazilian Portuguese Vocabulary Level Test (TVLPB). The test comprises five different frequency bands that correspond to the lexical distribution found in different corpora. The results show that native speakers' responses tend to follow the same pattern across the different bands of the test which indicates internal validity. The test was also administered with heritage speakers of Portuguese, and the results show that accuracy rates decline as the frequency of words is decreased.

Keywords: VLT; Portuguese as a second language; proficiency test; evaluation.

1 Introdução

A construção de instrumentos padronizados para a avaliação da competência e das habilidades de uso de língua adicional por seus falantes ainda é um desafio no português brasileiro como língua adicional, de acolhimento e de herança¹ (SOUZA; SÁ; AMARAL, 2020). O CELPE-Bras² é o certificado de referência para o nivelamento de usuários³ de

¹ De forma simplista, o termo “adicional” substitui os termos “estrangeira” e “segunda língua”, para mais informações, ler Leffa e Irala (2014). O termo “acolhimento”, segundo Miranda e Lopez (2019), é utilizado a partir do ensino de português aos refugiados. O ensino de português para filhos de imigrantes brasileiros deu visibilidade ao português brasileiro como língua de herança (SOUZA, 2016).

² Mais informações sobre o Celpe-Bras: <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/celpe-bras>.

³ O termo “usuário” foi escolhido em detrimento do termo “falante” tendo-se em vista a conotação de oralidade que, em nosso entender, é acarretada pelo termo preterido, o que excluiria a comunidade surda do escopo do presente trabalho. Agradecemos à revisão anônima por chamar nossa atenção para a importância deste esclarecimento.

português brasileiro, que consiste em uma bateria de testes chancelada pelo INEP (Instituto Nacional de Estudos e Pesquisa Educacionais Anísio Teixeira) e pelo Ministério da Educação, e possui grande revisão bibliográfica sobre a confiabilidade de seus resultados (DELL'ISOLA; SCARAMUCCI; SCHLATTER; JUDICE, 2003, FERREIRA, 2018, 2020). Por ser uma bateria de testes, o CELPE-Bras apresenta uma grande complexidade em sua administração, com procedimentos de muitas horas para sua aplicação, fazendo com que seja um instrumento de perfilamento com restrições em seu uso.

Com o intuito de criar um teste mais acessível e prático para um público profissional cujos trabalhos dependem ou se beneficiam de testes de perfilamento, criamos o Teste de Verificação Lexical do Português Brasileiro (TVLPB) baseado, especificamente, no Vocabulary Level Test (VLT) de Nation (1990). O VLT é considerado uma medida objetiva psicométrica de proficiência e tem uma rápida aplicação, em que o participante realiza o teste em 11 minutos. O teste criado por Nation avalia o conhecimento de vocabulário de aprendizes não-nativos de língua inglesa a partir do acerto de questões que associam significados a palavras que são apresentadas em 5 níveis de acordo com sua frequência na língua. A proficiência é medida no teste como um correlato de conhecimento de itens lexicais cuja frequência é progressivamente menor, ou seja, quanto menos frequente é a palavra acertada pelo participante do teste, maior é a proficiência desse participante. Assim, o TVLPB criado e descrito neste artigo pretende também correlacionar proficiência à amplitude lexical em um teste objetivo, acessível, de rápida e fácil aplicação.

Além de ser um teste de base psicométrica e, por isso, levar em consideração teorias e aspectos psicolinguísticos de aquisição e processamento da linguagem, exploradas na próxima seção, há em psicolinguística uma necessidade contínua de se entender como medidas subjetivas se correlacionam aos dados coletados nos métodos experimentais utilizados pela área (GLEASON; RATNER, 1998, TRAXLER, 2011, WARREN, 2013). Mais especificamente, a área do bilinguismo constantemente precisa de um método que profile seus participantes em relação à sua proficiência linguística. Em um dos últimos artigos de meta-análise publicados sobre quais medidas de proficiência eram utilizadas em estudos de bilinguismo, Hulstijn (2012) afirma que mais de 50% dos estudos nessa área usavam a autoavaliação como a principal ou única medida de proficiência linguística, mesmo com estudos que demonstram que há baixa confiabilidade nesse tipo de medida (GOLLAN *et al.*, 2012;

SOUZA; SILVA, 2015). Tais dados demonstram a necessidade da criação de um teste de proficiência em língua portuguesa que seja objetivo e prático, com aplicação não somente em sala de aula, mas também com aplicação em pesquisas sobre português como língua não-nativa.

Destarte, nosso objetivo foi construir um teste psicométrico que assegure a confiabilidade das informações inferidas por meio da observação do desempenho de seus participantes. A psicolinguística nos auxiliou no alinhamento entre as teorias sobre o léxico, seus construtos embaixadores e os itens utilizados para a testagem desses construtos em nosso instrumento de avaliação (TVLPB).

1.1 A avaliação da proficiência em língua adicional através da amplitude lexical

Não há apenas um único método de mensuração das competências e habilidades comunicativas do falante de línguas adicionais. Usualmente tal mensuração se apoia em modelos globais de tais competências e habilidades, sendo que um modelo dessa natureza que tem exercido contínua influência nos estudos e na prática da avaliação e testagem linguística é o de Bachman (1990). O modelo contempla tanto dimensões próprias da organização linguística quanto procedurais de seu uso. Assim, no tocante às dimensões propriamente linguísticas, ele orienta o avaliador a incorporar a observação sistemática da totalidade dos aspectos da organização formal da gramática da língua-alvo e dos aspectos sociopragmáticos típicos de seu uso em contextos diversos. Por sua vez, no tocante às dimensões procedurais, ele enfatiza o aspecto estratégico da competência em uma língua adicional, orientando avaliadores a atentar-se aos requisitos e demandas das tarefas e à configuração global dos eventos comunicativos que constituem os estímulos para eliciação de amostras de habilidades, nas situações de avaliação de habilidades e competências em segunda língua.

A longevidade do modelo de Bachman (1990) pode ser atribuída ao seu sucesso em capturar uma ampla gama de facetas das habilidades e competências de uso de línguas adicionais. Contudo, instrumentos de avaliação (usualmente testes) que implementam fidedignamente o modelo exigem dos avaliadores a elaboração de baterias com múltiplos componentes e subcomponentes. Esse tipo de bateria exige esquemas complexos para sua administração, comumente demandando longos períodos de testagem e a participação de equipes. Os testes de proficiência

em inglês como língua adicional da Universidade de Cambridge são exemplos da complexidade de administração ora citada, pois tratam-se de baterias com até 5 subtestes componentes, sendo que há, dentre eles, subtestes que isoladamente chegam a demandar duas horas apenas para a aplicação. Essa situação leva à necessidade de que sejam buscados instrumentos que viabilizem avaliações mais rápidas e econômicas do ponto de vista da administração, mas que ao mesmo tempo tenham a validade como métricas de proficiência adequadamente comprovadas.

Uma alternativa para a avaliação rápida de habilidades linguísticas de falantes de L2 frequentemente estudada é a mensuração da competência lexical desses falantes (NATION, 1990, READ, 2000). Tal competência é definida como o estado de conhecimento dos itens do vocabulário mental ao longo de três dimensões (HAASTRUP; HERINKSEN, 2007): 1 - a precisão da compreensão do significado de um dado item lexical; 2 - a capacidade de reconhecimento e de emprego do item lexical; e 3 - o conhecimento da relação do item com outros itens do léxico mental, tanto no nível paradigmático (sinonímia, antonímia, gradação, etc.), quanto no nível sintagmático (restrições colocacionais e combinatórias). Assim, os testes que medem explicitamente o conhecimento léxico da L2 podem se concentrar no tamanho do vocabulário. Alternativamente, tais testes podem se concentrar na quantidade de detalhes representacionais, exemplificadas nas relações de hiperonímia, sinonímia e antonímia; conotação, pertinência a contextos formais ou informais, etc., e também na conectividade no léxico mental, ou seja, a tipicidade de combinação dos itens com outros itens e com construções para a configuração de frases, dimensões geralmente abarcadas na denominação “profundidade do vocabulário”.

Há métodos de testagem de atestada robustez para a mensuração da amplitude lexical em L2 (READ, 2000, SCHMITT, 2010). Por outro lado, não há pleno consenso na literatura sobre a operacionalização da dimensão da profundidade lexical e sobre a existência de métodos e técnicas de exame psicométrico plenamente confiáveis para a captura dessa dimensão (SCHMITT, 2014). Segundo Meara e Alcoy (2010), um dos instrumentos mais aceitos para medir o tamanho, ou amplitude do vocabulário em segunda língua é o Vocabulary Levels Test (VLT), de Nation (1990), desenvolvido inicialmente para o inglês como língua adicional.

O VLT foi desenvolvido como um teste de nivelamento para aprendizes de inglês como L2 em contexto instrucional universitário e conta com estudos de validação tais como Schmitt, Schmitt e Clapham

(2001) e Souza e Silva (2015), o segundo, um trabalho sobre a adequação do teste para a população brasileira. Trata-se de um teste de rápida aplicação e apuração de escores, composto por um único tipo de tarefa de associação entre palavras estímulo e possíveis definições para elas. Os itens estão distribuídos por cinco níveis, ou partes, cada um dos quais apresenta ao sujeito 18 palavras a serem associadas corretamente com as definições apresentadas. Os níveis do VLT correspondem à progressiva diminuição da frequência dos itens lexicais que compõem o teste, frequência derivada do banco de dados do Corpus Brown, de amostras de discurso do inglês usado nos EUA.

Há estudos que relatam evidências sugestivas de correlações confiáveis entre o tamanho (amplitude) do vocabulário em língua adicional e outras dimensões observáveis da competência e da habilidade de uso dos falantes dessas línguas. Por exemplo, Milton (2010) realizou um estudo empírico com o objetivo de analisar o impacto do vocabulário como uma dimensão dentro dos níveis de proficiência em L2 no CEFR (Common European Framework of Reference for Languages), confirmando a talvez intuitiva noção de que amplitude de vocabulário da L2 aumenta constantemente à medida que seu usuário é classificado em níveis progressivamente mais altos do CEFR. Estudos de Alderson (2005) e de Milton (2013) demonstram que escores em testes de amplitude lexical são preditores confiáveis de habilidades linguísticas receptivas e produtivas na oralidade e na escrita em segunda língua. Souza *et al.* (2015), por sua vez, apresentam evidências de que os escores do VLT, obtidos com falantes brasileiros do inglês como L2, apresentam correlações positivas significativas tanto com escores de um teste de nivelamento que recobre habilidades globais na segunda língua quanto com sua capacidade de emitir julgamentos de aceitabilidade convergentes com o estatuto de gramaticalidade de sentenças, em uma tarefa com restrição temporal para a emissão de julgamentos.

São as evidências supracitadas, somadas à questão de que instrumentos de mensuração de proficiência em línguas adicionais de administração e apuração eficiente e econômica, que nos motivam a propor o TVLPB à comunidade científica e profissional.

1.2 A avaliação em português brasileiro como língua adicional

O Celpe-Bras tem ocupado, desde sua criação, o papel central como bateria avaliativa de proficiência do português brasileiro e um papel de destaque no processo contemporâneo do ensino e da aprendizagem

como língua adicional do português brasileiro. (BULLA; COSTA, 2020). Podemos observar que hoje grande parte das discussões sobre a avaliação do português como língua adicional gravitam entre o Celpe-Bras e debates educacionais sobre o uso de notas como instrumento de controle e regulação, os problemas que isso acarreta, o desejo de uma maior autonomia e autorregulação dos aprendizes.

Apesar de tais reflexões serem importantes para o desenvolvimento de métodos avaliativos, muitas vezes falta, no contexto brasileiro, uma discussão que compreenda a questão objetiva dos testes de base psicométrica. Repercute-se uma sensação de mal-estar, como descrita por Souza, Sá e Amaral (2020), em que testes objetivos e métodos de mensuração e diferenciação de habilidades trazem um incômodo quando enquadrados. Contudo, ao pensarmos que os testes são, muitas vezes, responsáveis por hierarquizações e que fazem parte do dia a dia dos professores de línguas adicionais, percebemos o quanto é urgente um alinhamento entre as teorias da natureza da linguagem, sua aquisição e seu processamento, e os itens utilizados para a testagem dos aprendizes.

Assim como é discutido por McNamara (2000), quando não há um alinhamento entre técnicas de ensino, teorias de aprendizagem e mecanismos de testagem, há um problema nas relações entre a validade do construto e a validade do conteúdo. Dessa forma, a construção de testes, que não façam necessariamente parte de uma bateria e que alinhem tais questões, ainda é de grande necessidade no cenário avaliativo do português como língua adicional e, logo, propomos a criação do Teste de Verificação Lexical do Português Brasileiro (TVLPB) na tentativa de trazer uma medida válida e objetiva.

2 Metodologia

Com o intuito de criar a métrica de verificação lexical para o português brasileiro, baseada no VLT de Nation (1990), uma série de procedimentos, descrita nesta seção, foi estipulada.

2.1 A seleção das rubricas

A seleção das rubricas foi realizada a partir de uma busca em três corpora: ptTenTen (JAKUBÍCEK *et al.*, 2013), Linguateca (PB escrito)

(SARDINHA *et al.*, 2008) e C-Oral (PB oral) (RASO; MELLO, 2012)⁴. O ptTenTen é um corpus disponibilizado pela plataforma *Sketch Engine*⁵ e é composto por mais de 4 bilhões de palavras extraídas e processadas automaticamente da internet por um *web crawler*. No corpus, que também contém dados do português de Portugal, foi selecionada a opção de pesquisa em somente português brasileiro.

Os outros dois corpora contêm somente dados do português do Brasil. O C-Oral é um corpus de fala espontânea com cerca de 300 mil palavras e o Linguateca é um corpus de diferentes gêneros escritos e contém em média 1 bilhão de palavras. Foram extraídas as 5000 palavras mais frequentes de cada corpus e separadas por 5 bandas de frequência. A primeira banda continha palavras até a posição 1000, a segunda de 1001 a 2000, a terceira de 2001 a 3000, a quarta de 3001 a 4000 e a última de 4001 a 5000.

Selecionamos, em cada banda, 12 adjetivos, 12 substantivos e 12 verbos, dando preferência por palavras que se encontravam na mesma faixa de frequência nos três corpora. A escolha de quais palavras fariam parte das questões foi lematizada⁶. A palavra de maior frequência era a selecionada e seus lemas descartados. A partir do momento que não encontrávamos mais palavras que pertenciam aos três corpora na mesma banda de frequência, escolhemos palavras com o mesmo nível de frequência nas variantes oral (a partir do corpus C-Oral) e escrita (Linguateca). Para a quarta e a quinta banda, faltaram palavras que compusessem dois corpora (oral e escrito) ao mesmo tempo, dessa forma, optamos por reduzir para 06 adjetivos, 12 substantivos e 06 verbos (Tabela 1).

⁴ O ptTenTen está disponível em: <https://www.sketchengine.eu/pttenten-portuguese-corporus/>

O Linguateca em: <https://www.linguateca.pt/CETEMPUBLICO/>

E o C-Oral em: <http://www.c-oral-brasil.org/>

⁵ SketchEngine é um Corpus Query System, ou sistema de pesquisa de corpus. Está disponível no site: <http://www.sketchengine.co.uk/>.

⁶ Segundo Harrington (2018), o lema é a forma de uma palavra que é escolhida por convenção para representar a forma canônica de palavras que compartilham o mesmo radical. Ao se selecionarem as palavras de um teste de verificação lexical, a escolha é lematizada, pois o usuário pode inferir o lema através de seus lemas.

Tabela 1 – Questões e classes de palavras por banda

Banda	Frequência	Adjetivos	Substantivos	Verbos	Total de Questões
1	0-1000	12	12	12	18
2	1001-2000	12	12	12	18
3	2001-3000	12	12	12	18
4	3001-4000	6	12	6	12
5	4001-5000	6	12	6	12

Fonte: Elaborada pelos autores.

A partir das palavras selecionadas, enquanto algumas palavras foram escolhidas como as respostas corretas da questão, as outras, da mesma banda de frequência, eram colocadas como distratoras. Assim como em Nation (1990), todas as palavras apareciam em três questões (Figura 1), compondo 1 item. Foram criadas 78 questões.

A escolha dos significados ou sinônimos para as questões ocorreu a partir da seleção da primeira definição associada à rubrica no dicionário on-line Priberam⁷ na modalidade português brasileiro.

2.2 A demarcação do tempo

O VLT de língua inglesa (NATION, 1983, 1990; SCHMITT; SCHMITT; CLAPHAM, 2001) é um teste temporalizado, em que o sujeito tem um tempo máximo de 11 minutos para responder às questões. O primeiro piloto do TVLPB foi realizado no *Easy Test Maker*⁸ e o tempo disponibilizado para os sujeitos era de 11 minutos, como o VLT de língua inglesa. 5 participantes brasileiros, alunos do programa de pós-graduação em estudos linguísticos da UFMG, realizaram o teste e nenhum foi capaz de realizá-lo no tempo estipulado. Dessa forma, realizamos um segundo teste para determinar qual seria o tempo do teste em português brasileiro.

⁷ Disponível em: <https://dicionario.priberam.org/>.

⁸ Disponível em: <https://www.easytestmaker.com/>.

Figura 1 – Exemplo de três questões da primeira banda do TVLPB, que compõem 1 item

Nome de uma habitação.

- casa
- pessoa
- estado
- trabalho
- vez
- público

Indivíduo, ser humano.

- casa
- pessoa
- estado
- trabalho
- vez
- público

Condição em que alguém ou algo se encontra.

- casa
- pessoa
- estado
- trabalho
- vez
- público

Fonte: Elaborado pelos autores.

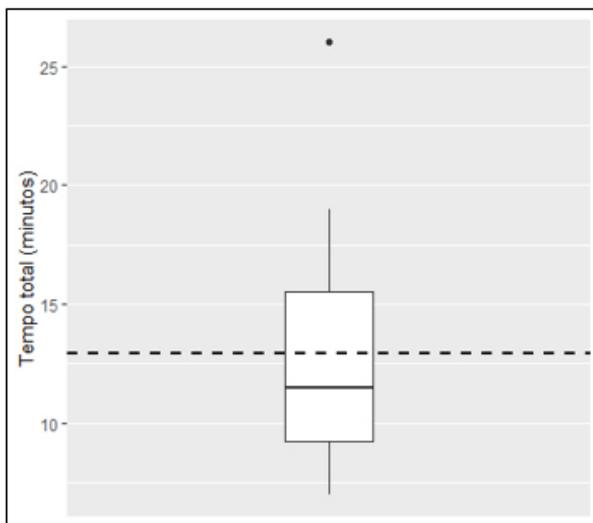
O segundo teste apresentava as mesmas questões, mas não era temporalizado. 18 sujeitos falantes nativos de português brasileiro, todos com ensino superior completo, realizaram o teste no *Google Forms*.

Como procedimento para calcular o tempo que seria necessário para que um falante nativo terminasse, os participantes respondiam a pergunta “Quantas horas são?” no início do teste e no final de cada uma das 5 partes. Os participantes levaram em média treze minutos (Gráfico 1), que consideramos como o tempo limite do nosso TVLPB.

2.3 A escolha do software

O teste é de múltipla escolha e o participante deve combinar um significado ou um sinônimo a uma resposta em um tempo limitado (Figura 2). Apesar de um grande número de ferramentas permitir a criação de questionários, como nosso intuito era que o teste fosse facilmente acessado e de forma on-line e gratuita, acabamos com um número limitado de plataformas para a programação do teste. Um outro fator de dificuldade foi a necessidade de temporalizar o teste, que acabou reduzindo nossas opções à plataforma *JotForm*⁹.

Gráfico 1 – Boxplot do tempo gasto pelos sujeitos, com a média tracejada



Fonte: Elaborado pelos autores.

⁹ Disponível em: <https://www.jotform.com/>.

Figura 2 – Printscreen do teste na plataforma JotForm



13:00
minutes seconds

VLT - Parte 01

Associe a palavra mais adequada para cada significado.

Confirme seu email *

Recente.

- grande
- só
- novo
- social
- bom
- mesmo

Fonte: Elaborado pelos autores.

A plataforma JotForm é uma plataforma desenvolvida para a criação e hospedagem de formulários, que não exige nenhum tipo de instalação nem para a criação dos testes nem para sua aplicação. Os testes podem ser criados com diferentes ferramentas, incluindo a possibilidade de serem temporalizados, como necessitávamos. Além disso, a plataforma permitiu a aleatorização das questões dentro de uma mesma banda e dos seus itens. O aplicativo apresenta uma versão gratuita e diferentes versões

pagas, nosso teste foi hospedado em uma versão gratuita que permite a resposta de 100 diferentes participantes por mês¹⁰.

3 Validação com nativos

Falantes nativos realizaram o TVLPB para que verificássemos a taxa de acerto por rubrica e por banda do teste. Nossa hipótese era que os falantes nativos apresentariam um alto nível de acerto, sendo uma forma de validação interna do teste¹¹. Outra hipótese era que a escolha dos itens e a demarcação do tempo seriam válidos se os falantes nativos conseguissem terminar o teste temporalizado e se as questões de todos os itens conduzissem à adesão clara a uma única opção de resposta.

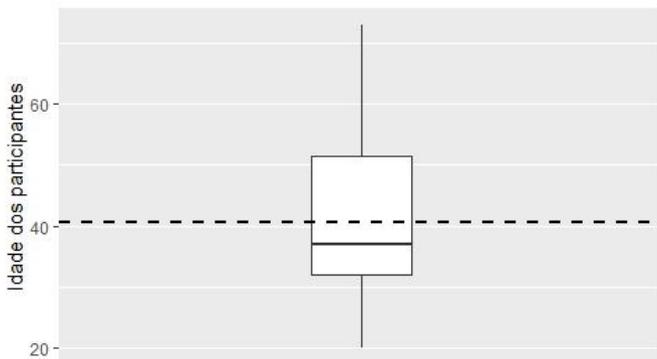
3.1 Participantes

115 participantes (23 homens) falantes nativos de português brasileiro realizaram a tarefa. O participante mais jovem tinha 20 anos e o mais velho 73, a idade média era 40 anos (Gráfico 2). Com relação à escolaridade (Gráfico 3), todos tinham, no mínimo, ensino médio completo, sendo que 36 tinham ensino superior completo, 38 alguma especialização e 17 tinham mestrado ou doutorado.

¹⁰ Como instrumento de hospedagem para o processo de validação e exploração do TVLPB, a plataforma se demonstra suficiente. Contudo, sabemos que, para uma ampla utilização do teste, seria interessante que sua hospedagem fosse feita institucionalmente, com programação própria.

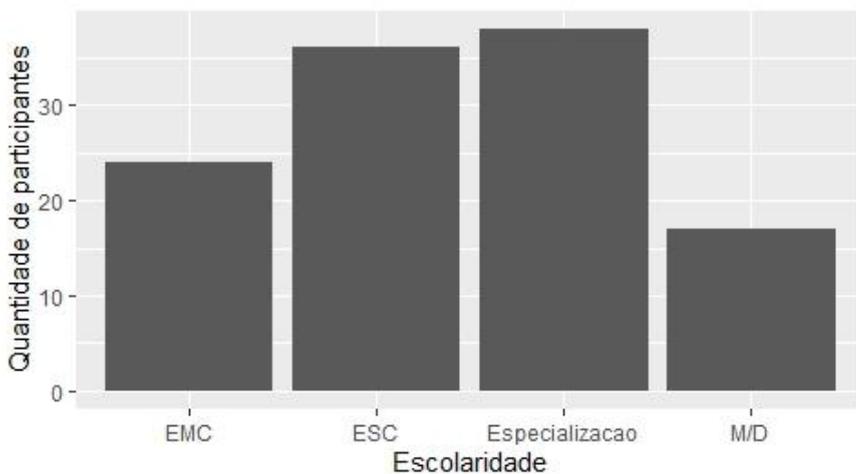
¹¹ Sabemos que o termo “validade” apresenta diferentes nuances, como “validade de conteúdo”, “validade de critério” e “validade de construto”. Nossa intenção é uma validação do construto “proficiência” e sua operacionalização através de um teste de vocabulário. Como um primeiro passo desse processo, realizamos uma validação interna a partir dos dados de falantes nativos.

Gráfico 2 – Boxplot das idades dos participantes (a linha marca a média)



Fonte: Elaborado pelos autores.

Gráfico 3 – Gráfico de barras com as escolaridades dos participantes:.



Fonte: Elaborado pelos autores.

Nota: A escolaridade máxima deveria ser marcada (somente uma opção era possível).

Todos possuem ensino médio completo (EMC), mas as escolaridades alcançadas poderiam ser ensino superior completo (ESC), especialização (Especialização), mestrado ou doutorado (M/D).

3.2 Materiais e procedimentos

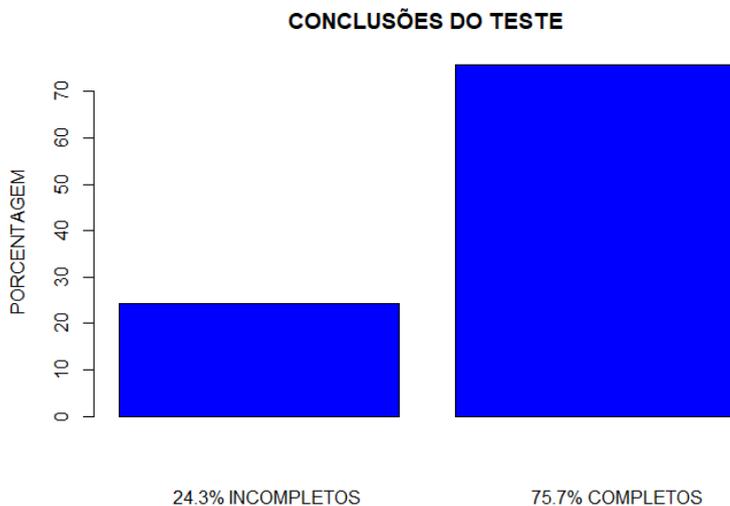
Os participantes responderam as 78 questões do TVLPB em até 13 minutos por meio da plataforma JotForm. Antes de entrarem no teste, eles leram o Termo de consentimento livre e esclarecido e marcaram se desejavam ou não participar do teste; todos consentiram. Em seguida, responderam questões demográficas e passaram por 3 questões similares às experimentais em que a resposta correta era fornecida, a título de treinamento. Após a fase de treinamento, era dado início à contagem regressiva e eles realizavam o teste. Uma banda de cada vez era disponibilizada, sendo que para acessar a próxima parte, clicava-se em próximo. Nenhuma questão tinha resposta obrigatória.

3.3 Resultados

A primeira observação feita foi em relação à quantidade de sujeitos que completaram em 13 minutos. 75,7% dos participantes completaram o teste dentro de 13 minutos, ou seja, não deixaram respostas em branco (Gráfico 4), o que é uma diferença significativa dos que não completaram ($\chi^2=42,2$, $df=1$, $p<.05$).

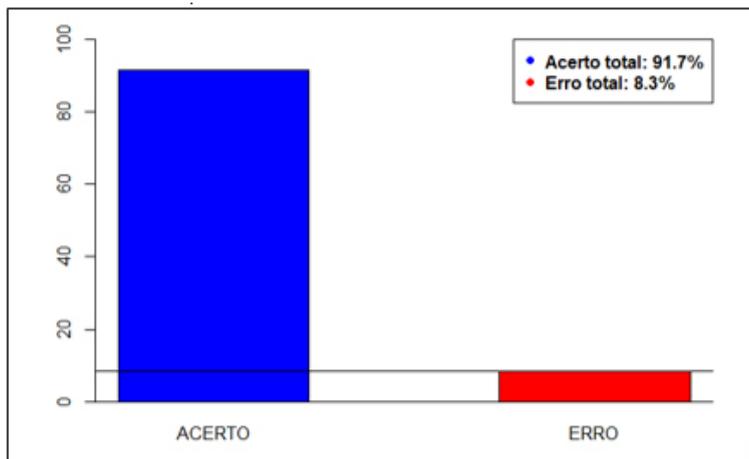
Em seguida, observamos a porcentagem de acerto das questões que não foram deixadas em branco, com uma taxa de acerto de 91,7% (Gráfico 5) ($\chi^2=6358,2$, $df=1$, $p<.05$). Também observamos a distribuição do acerto por idade (Gráfico 6) e escolaridade (Gráfico 7). Em todas as faixas etárias houve mais de 75% de acerto e em todas as escolaridades mais de 89%.

Gráfico 4 – Porcentagem de falantes nativos que completaram e que não completaram o teste em 13 minutos



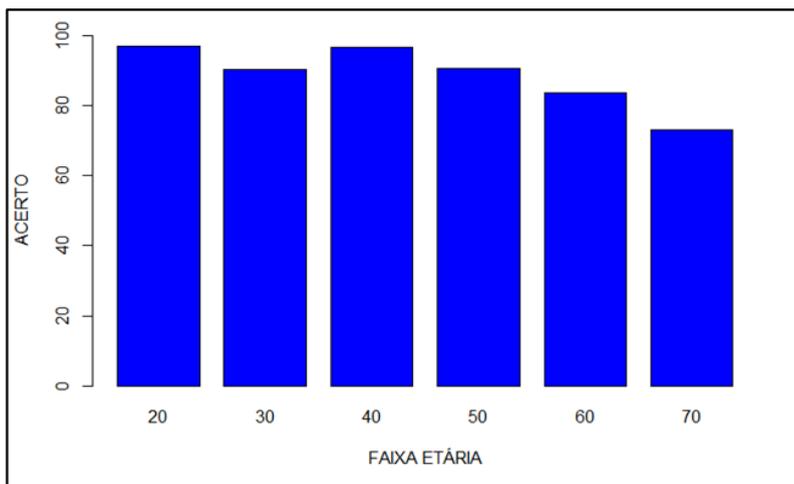
Fonte: Elaborado pelos autores.

Gráfico 5 – Porcentagem de acerto das questões do teste pelos falantes nativos



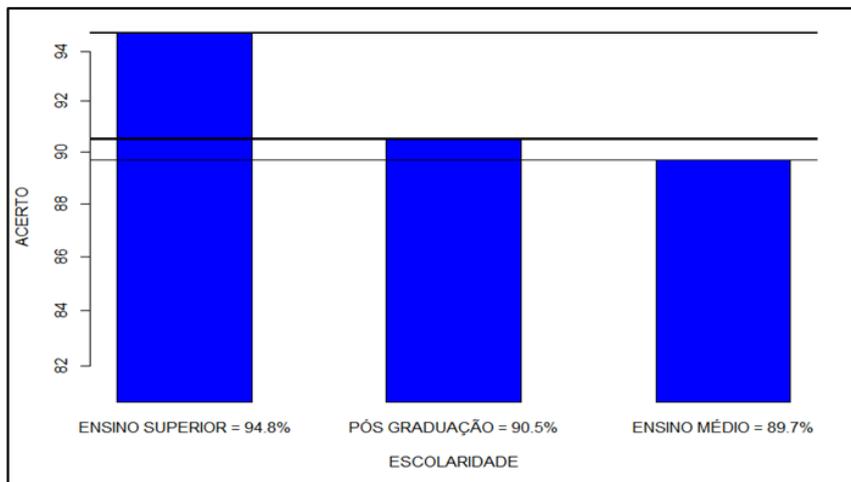
Fonte: Elaborado pelos autores.

Gráfico 6 – Porcentagem de acerto das questões do teste em relação à idade dos falantes nativos



Fonte: Elaborado pelos autores.

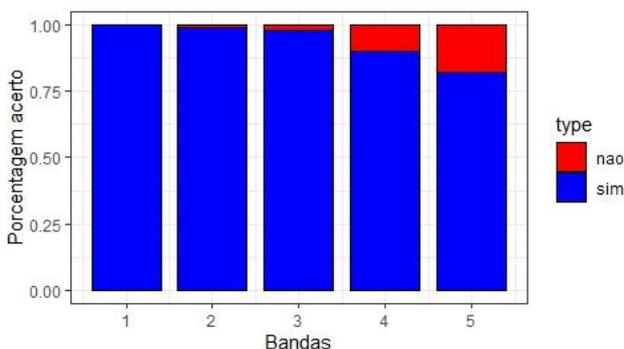
Gráfico 7 – Porcentagem de acerto das questões do teste em relação à escolaridade dos falantes nativos



Fonte: Elaborado pelos autores.

Com relação às bandas, também analisamos se houve diferença entre a taxa de acerto entre bandas. Nossa hipótese era que as últimas bandas, por apresentarem palavras com menor nível de frequência do teste, teriam menor taxa de acerto. Houve um aumento significativo de erros (Gráfico 8) de acordo com a banda de frequência dos itens ($\chi^2=492.53$, $df=4$, $p<.05$).

Gráfico 8 – Percentagem de acerto das questões do teste em relação à banda de frequência em que os itens se encontravam



Fonte: Elaborado pelos autores.

Além de uma análise de qui-quadrado para observar se há adesão clara às rubricas e uma diferença significativa entre bandas, realizamos o teste Alfa de Cronbach¹² para a verificação da consistência interna da escala proposta. O teste indica uma alta consistência interna quando consideramos as bandas ($\alpha = .73$, $N = 5$) e quando consideramos os itens ($\alpha = .95$, $N = 78$). Dessa maneira, verificamos que os itens do teste podem estar aferindo o mesmo conceito subjacente já que apresentam correlação em sua covariância.

3.4 Discussão dos resultados

Os nossos resultados demonstraram que mais de 75% dos falantes nativos, em diversas faixas etárias e níveis de escolaridade, conseguiram terminar o teste temporalizado nos treze minutos propostos, sendo assim

¹² Agradecemos à sugestão do parecer anônimo que indicou o teste de Alfa de Cronbach por ser um coeficiente muito utilizado nas medidas de consistência interna.

considerado tempo suficiente para o TVLPB. Além disso, as rubricas conduziram à adesão clara a uma única opção de resposta, sendo a taxa total de acerto do teste de 91,7%. Dessa forma, a partir de uma percepção de que os falantes nativos poderiam ser considerados o padrão ouro da performance esperada no teste, acreditamos que o tempo seja suficiente e que as rubricas estejam adequadas, com uma única opção de resposta. Além disso, o teste teve um acréscimo de dificuldade entre bandas, o que é esperado em um teste de nivelamento de proficiência.

Contudo, o VLT de Nation (1990) faz uma distinção entre níveis de baixa proficiência e alta proficiência, sendo o aprendiz de alta proficiência aquele que chega à banda 04 do teste. Como são falantes nativos, é esperado que nossos participantes sejam de alta proficiência. Como uma forma de explorar a diferença entre as bandas, nosso planejamento inicial era aplicar nosso teste com falantes não nativos previamente nivelados em imersão no Brasil. Devido à pandemia, nossos planos foram modificados e aplicamos o teste em falantes de herança de diversas proficiências para averiguar sua possível funcionalidade com falantes de herança do português brasileiro, como descrevemos na próxima seção.

4 Uma aplicação com falantes de herança

No atual cenário educacional e de pesquisa, principalmente fora do Brasil, pesquisadores e professores que trabalham com português como língua de herança utilizam ferramentas não validadas que variam em sua estrutura e na escolha de construtos para medir a proficiência. Poucos pesquisadores e educadores têm acesso à bateria do Celpe-Bras ou que dispõem de mais de três horas para sua aplicação. Logo, se o TVLPB se mostrar eficiente, poderia se tornar uma ferramenta com o potencial de unificar a maneira de como medimos a proficiência de falantes de herança.

4.1 Métodos avaliativos para falantes de herança

Os trabalhos que tratam sobre a testagem de proficiência de falantes de herança são escassos e um dos possíveis motivos se deve ao fato de que esses falantes, na maioria dos casos, apresentam heterogeneidade em suas habilidades linguísticas, sendo que alguns apresentam grande domínio da habilidade oral e outros apresentam, exclusivamente, um extenso vocabulário receptivo da língua (POLINSKY, 2018). Logo,

pesquisadores e professores frequentemente têm que escolher um teste que se foque na medida de proficiência oral, por exemplo, fazendo uma entrevista que segue a escala da ACTFL (*American Council on the Teaching of Foreign Languages*)¹³, ou um teste de vocabulário que foi desenvolvido primariamente para falantes de segunda língua. Assim, o pesquisador sempre terá que ter cautela porque esses testes não foram validados com falantes de herança, não levam em consideração particularidades das línguas de herança e também não são sensíveis às diferenças dialetais e de registro.

Por exemplo, Kagan e Friedman (2003) fizeram uma entrevista seguindo a escala indicada pelo ACTFL. Eles reportam que o teste foi capaz de diferenciar as proficiências dos 11 falantes de herança de russo que participaram de seu estudo e encontraram também uma correlação entre o teste e medidas de autoavaliação. No entanto, os autores notam que tal teste só foi eficaz porque todos os participantes tinham acesso a um registro uniforme, reconhecendo que se os falantes de herança têm diferentes dialetos ou registros, a escala da ACTFL não seria a melhor escolha, porque, nesse caso, o avaliador da entrevista deve conhecer as particularidades da língua de herança para evitar que avaliações recebam notas baixas indevidamente.

Potowski *et al.* (2011) apresentam a criação de um teste de proficiência para diferenciar falantes de herança de espanhol e falantes de espanhol como segunda língua, pois acreditam que testes separados são problemáticos. Para os autores, falantes de herança, talvez por temerem um curso mais difícil ou por não saberem certamente se são falantes de herança, podem se identificar como aprendizes de segunda língua. Assim, os autores recomendam que apenas um teste, que seja capaz de discriminar proficiência, seja usado. O teste criado por eles é dividido em 4 partes e cada parte é focada em específicas construções gramaticais e vocabulário. Apesar das quatro partes do teste não funcionarem da mesma forma para discriminar perfis de proficiência, os autores tiveram êxito em mostrar que um mesmo teste escrito pode ser utilizado para diferenciar falantes de herança de falantes de segunda língua.

¹³ A escala de proficiência oferecida pela ACTFL para entrevistas pode ser acessada através da página: <https://www.actfl.org/assessment-research-and-development/actfl-assessments/actfl-postsecondary-assessments/oral-proficiency-interview-opi>.

O teste desenvolvido por Potowski *et al.* (2011) apresenta importantes vantagens, dado sua fácil aplicação. No entanto, como descrito anteriormente, não são todas as partes do seu teste que podem diferenciar os falantes. Talvez isso aconteça devido à mescla de diferentes elementos linguísticos em cada parte do teste: vocabulário, gramática e expressões, ou seja, os aspectos testados foram, implicitamente, amplitude e profundidade de vocabulário. Logo, vemos uma diferença entre essas propostas e a proposta do TVLPB, já que o aspecto testado no último é uniforme. O TVLPB testa a amplitude de vocabulário de um falante, em que cada parte é especificada pela distribuição de frequência encontrada na língua.

4.2 Participantes

Aplicamos o teste com 23 falantes de herança¹⁴ universitários do português brasileiro, recrutados nos cursos de português oferecidos pelo Departamento de Espanhol e Português da Universidade de Massachusetts, Amherst. Os participantes estavam nos dois últimos níveis do curso de língua e em cursos avançados, mas tinham proficiências variadas, entre as classificações avançado e intermediário.

Todos os participantes selecionados para esse estudo reportaram que foram expostos à língua portuguesa desde o nascimento e pelo menos um dos pais era falante nativo do português brasileiro. Ademais, outro fator comum a esses falantes, é que todos indicaram que a língua mais usada durante a infância era o português e a língua mais usada durante a adolescência e fase adulta era o inglês¹⁵. Em relação ao país de nascimento, 61% nasceram no Brasil e se mudaram para os Estados Unidos antes dos dois anos de idade e 39% nasceram nos Estados Unidos, todos sempre moraram no estado de Massachusetts. A idade média dos participantes

¹⁴ Para averiguar o status de falante de herança, utilizamos um questionário de uso de língua para bilíngues criado pelos autores e outro questionário sociodemográfico. Apesar de serem questionários diferentes, ambos continham as informações relevantes para este estudo, como: idade de exposição à língua e língua falada pelos pais.

¹⁵ Esse é um fenômeno comum aos falantes de herança já que usam a língua minoritária em casa e quando começam a expandir seu círculo social adotam a língua majoritária da sociedade (POLINSKY, 2018).

é de 19,4 anos com desvio padrão de 1,3 anos. Todos os participantes receberam créditos ou incentivo em dinheiro para participar no estudo¹⁶.

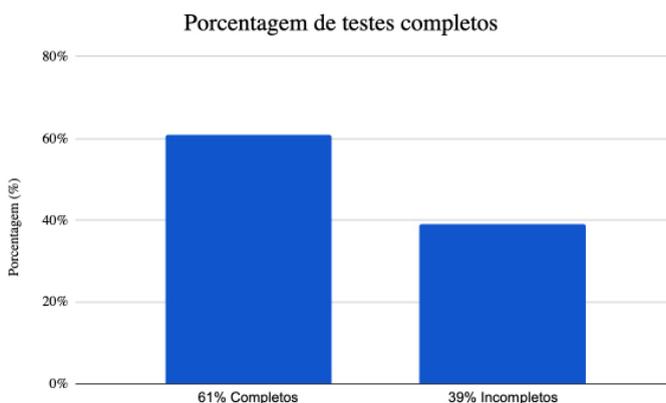
4.3 Materiais e procedimentos

O pesquisador recebeu cada participante em uma sala silenciosa na universidade e explicou em inglês e português os procedimentos do teste. Após assinar o termo de consentimento, o participante começava a fase de treinamento com a ajuda do pesquisador. Nessa etapa, o pesquisador averiguava se o participante havia entendido os procedimentos e, ao final, respondia qualquer pergunta que o participante pudesse ter. Após o treinamento, o participante era informado que teria 13 minutos para terminar o teste. Todos os testes foram administrados através da plataforma Jotform em um computador *MacBook Air*.

4.4 Resultados

Os resultados mostram que 61% dos participantes conseguiram terminar o teste dentro do prazo e 39% não conseguiram (Gráfico 9), uma diferença que não foi significativa ($\chi^2=1.08$, $df=1$, $p>.05$). A maioria dos testes incompletos pararam na banda 4.

Gráfico 9 – Porcentagem de testes completos

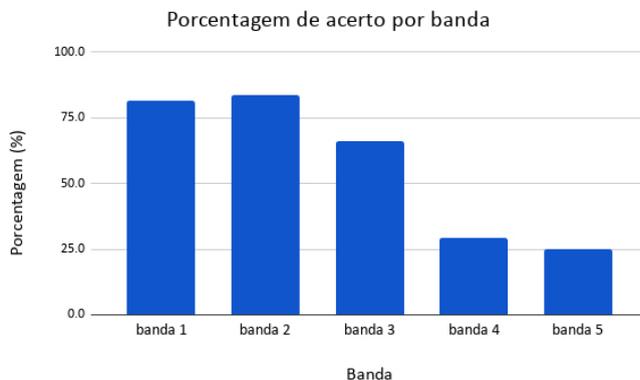


Fonte: Elaborado pelos autores.

¹⁶ A comissão de ética das universidades dos Estados Unidos permite que pessoas recebam pagamentos como incentivo para participação em estudos.

Ao analisarmos o índice de acerto por bandas (Gráfico 10), as bandas 1 e 2 apresentam um nível de acerto maior que 75%, a banda 3 um nível de acerto de 66% e as bandas 4 e 5, 30% e 26%, respectivamente, o que indica que houve um aumento de dificuldade entre algumas das bandas de palavras com menor frequência ($\chi^2=252.8$, $df=4$, $p<.05$).

Gráfico 10 – Porcentagem de acerto por banda de frequência

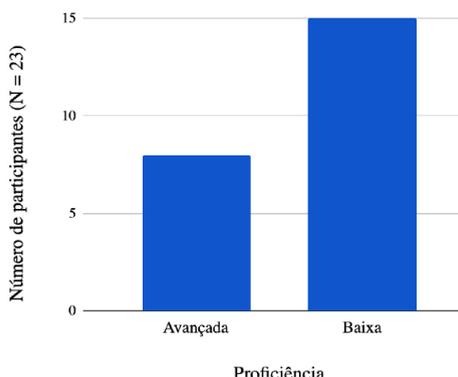


Fonte: Elaborado pelos autores.

Levando em consideração o nivelamento esperado do TVLPB, em que o participante tem que atingir a banda 04 para ser considerado um falante de alta proficiência, observamos que 8 participantes foram classificados como alta proficiência, 15 como baixa proficiência¹⁷ (Gráfico 11).

¹⁷ Como não colocamos o preenchimento obrigatório das questões, dois participantes pularam bandas e, por isso, foram considerados de baixa proficiência, por conseguirem atingir o ponto de corte em uma banda avançada, mas não conseguiram atingir o ponto de corte em uma banda anterior, o que é exigido no VLT.

Gráfico 11 – Classificação da proficiência dos falantes de herança



Fonte: Elaborado pelos autores.

4.5 Discussão dos resultados

Se levássemos em consideração apenas a nossa percepção da proficiência oral dos nossos participantes, falantes de herança, classificaríamos, erroneamente, 19 participantes como alta proficiência e os outros como intermediários.

Um primeiro componente de nivelamento seria o tempo, que se demonstrou mais uma vez como um importante fator de nosso teste. Percebemos que somente 14 dos falantes de herança conseguiram concluir o teste em 13 minutos, sendo que esperávamos que mais participantes seriam capazes de concluí-lo.

Como esperado, há uma distinção de acertos em relação às bandas, os resultados indicam a presença de um componente de dificuldade dado que as notas diminuem na medida em que a frequência das palavras também diminui. As bandas 1 e 2 mostram que os participantes foram capazes de chegar a um nível de acerto de maior que 75%, mas esse nível declina significativamente nas bandas subsequentes, chegando a 26% na última banda do teste. Assim, mais uma vez, vemos uma discrepância entre a percepção da proficiência que tínhamos dos falantes e a proficiência aferida através do conhecimento de vocabulário. Em outras palavras, falantes de herança apresentam domínio das palavras mais frequentes da língua, mas têm dificuldades quando as palavras são menos frequentes, principalmente quando essas palavras passam da faixa das três mil palavras mais frequentes.

5 Discussão geral dos resultados

Nosso objetivo era de criar um teste mais acessível e prático, com parâmetros aceitáveis de qualidade para um público profissional cujos trabalhos dependem ou se beneficiam de testes de perfilamento. O teste TVLPB foi criado e, a partir de sua validação interna com nativos, percebeu-se que as rubricas eram consistentes, tendendo a uma única opção de resposta. Além disso, esperava-se que a manipulação de frequência entre bandas interferisse na quantidade de acerto, sendo que quanto menor a frequência das palavras de uma banda, menor seria a taxa de acerto na mesma banda. A aplicação com nativos também demonstrou que tal manipulação foi efetiva e que o nível de dificuldade do teste variava de acordo com a banda. Dessa forma, conseguimos criar um teste mais acessível e prático que apresenta validade interna em suas análises com nativos, contudo, pelos nativos apresentarem um padrão ouro, em que sempre se espera uma alta proficiência, acreditamos que mais dados sejam importantes para assegurar a confiabilidade das informações inferidas por meio da observação do desempenho de seus participantes.

Em busca de atestar ainda mais a qualidade do nosso teste, realizamos uma aplicação com falantes de herança que indicou uma validade interna do teste, por meio do tempo e das bandas do TVLPB, que se mostraram componentes de nivelamento. Entretanto, acreditamos que mais estudos precisam ser feitos com falantes de herança, pois nossas expectativas sobre as proficiências dos alunos nem sempre corresponderam aos resultados dos testes. Precisamos da validação externa com esses falantes de herança, em que outras medidas de proficiência sejam correlacionadas aos nossos dados. Estamos, no momento, comparando a performance de falantes de herança no nosso TVLPB com uma medida de proficiência oral da ACTFL e medidas de autoavaliação.

Ainda com relação a uma validade externa de nosso teste, nossa ideia inicial era que o mesmo processo fosse realizado com aprendizes de português como segunda língua, imersos no Brasil, que tivessem alguma certificação de proficiência prévia, contudo, devido à pandemia, houve um regresso desses falantes às suas perspectivas casas e não conseguimos um número mínimo de respondentes para o nosso teste.

Referências

ALDERSON, J. C. *Diagnosing foreign language proficiency: the interface between learning and assessment*. London: Continuum, 2005.

BACHMAN, L. *Fundamental considerations in language testing*. Oxford: Oxford University Press, 1990.

BULLA, G. S.; COSTA, E. V. Português como língua adicional: uma entrevista com Margarete Schlatter. Entrevistada: Margarete Schlatter. *Revista virtual de estudos da linguagem*, [S. l.], v. 18, n. 35, p. 489-508, 2020.

DELL'ISOLA, R. L. P.; SCARAMUCCI, M. V. R.; SCHLATTER, M.; JUDICE, N. A avaliação de proficiência em português língua estrangeira: o exame CELPE-Bras. *Revista brasileira de linguística aplicada*, [S. l.], v. 3, n. 1, p. 153-184, 2003. ISSN: 1984-6398. DOI: <https://doi.org/10.1590/S1984-63982003000100010>.

FERREIRA, L. M. L. *Avaliação da proficiência oral: uma análise fatorial e de discriminação dos itens do exame Celpe-Bras*. 2018. Tese (Doutorado em Estudos de Linguagens) – Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2018. Disponível em: http://www.ufrgs.br/acervocelpebras/pesquisas/FERREIRA_2018_tese.pdf. Acesso em: 23 out. 2020.

FERREIRA, L. M. L. Componentes da habilidade oral: uma análise das propriedades dos itens analíticos do exame Celpe-Bras. *Revista da Abralin*, [S. l.], v. 19, n. 3, p. 799-824, 2020. Disponível em: <https://revista.abralin.org/index.php/abralin/issue/view/83/4>. Acesso em: 01 mar. 2021.

GLEASON, J. B.; RATNER, N. B. *Psycholinguistics*. Orlando: Harcourt Brace & Company, 1998.

HAASTRUP, K.; HENRISEN, B. Vocabulary acquisition: acquiring depth of knowledge through network building. *International journal of applied linguistics*, [S. l.], v. 10, n. 2, p. 221-240, Apr. 2007. DOI: <https://doi.org/10.1111/j.1473-4192.2000.tb00149.x>

HARRINGTON, M. *Lexical facility: size, recognition speed and consistency as dimensions of second language vocabulary knowledge*. London: Palgrave Macmillan, 2018.

HULSTIJN, J. H. The construct of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism: language and cognition*, Cambridge, v. 15, n. 2, p. 422-433, Apr. 2012. DOI: <https://doi.org/10.1017/S1366728911000678>.

JAKUBÍČEK, M. *et al.* The TenTen corpus family. In: INTERNATIONAL CONFERENCE ON CORPUS LINGUISTICS, 5., 2013, Alicante. *Proceedings* [...]. [S. l.]: [s. n.], 2013.

KAGAN, O.; FRIEDMAN, D. Using the OPI to place heritage speakers of Russian. *Foreign language annals*, Alexandria, v. 36, n. 4, p. 536-545, 2003. DOI: <https://doi.org/10.1111/j.1944-9720.2003.tb02143.x>.

LEFFA, V. J.; IRALA, V. B. O ensino de outra(s) língua(s) na contemporaneidade: questões conceituais e metodológicas. In: LEFFA, V. J.; IRALA, V. B. (org). *Uma espiadinha na sala de aula: ensinando línguas adicionais no brasil*. Pelotas: EDUCAT, 2014. p. 21-48.

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

MCNAMARA, T. *Measuring second language performance*. London: Addison-Welsely Longman, 1996.

MEARA, P.; ALCOY, J. Words as species: an alternative approach to estimating productive vocabulary size. *Reading in a foreign language*, Manoa, v. 22, n. 1, p. 222–236, Apr. 2010. ISSN: 1539-0578.

MILTON, J. Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In: BARDEL, C.; LINDQVIST, C.; LAUFER, B. (org.). *L2 vocabulary acquisition, knowledge and use: new perspectives on assessment and corpus analysis*. [S. l.]: European Second Language Association, 2013.

MILTON, J. The development of vocabulary breadth across the CEFR levels. In: BARTNING, I.; MARTIN, M.; VEDDER, I. (org). *Communicative proficiency and linguistic development: intersections between SLA and language testing research*. [S. l.]: European Second Language Association, 2010. (EuroSLA monograph series, v. 1)

MIRANDA, Y. C. C.; LOPEZ, A. P. A. Considerações sobre a formação de professores no contexto do português como língua de acolhimento. In: FERREIRA, L. C.; PERNA, C.; GUALDA, R.; LEURQUIN, E. V. L. F. (org.). *Língua de acolhimento: experiências no Brasil e no mundo*. Belo Horizonte: Mosaica, 2019.

NATION, I. S. P. Testing and teaching vocabulary. *Guidelines*, [S. l.], v. 5, p. 12-25, 1983.

NATION, I. S. *Teaching and learning vocabulary*. Boston: Heinle and Heinle, 1990.

POLINSKY, M. *Heritage languages and their speakers*. Cambridge: Cambridge University Press, 2018.

POTOWSKI, K. *et al.* Developing an online placement exam for Spanish heritage speakers and 12 students. *Heritage language journal*, Cambridge, MA, v. 9, n. 1, p. 51-76, 2011. DOI: <https://doi.org/10.46538/hlj.9.1.4>.

RASO, T.; MELLO, H. (org.). *C-oral-Brasil I: corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG, 2012.

READ, J. *Assessing vocabulary*. Cambridge: Cambridge University Press, 2000.

SARDINHA, T. B. *et al.* O corpus brasileiro. In: ENCONTRO DE LINGÜÍSTICA DE COPUS, 7., 2008, São José do Rio Preto. *Anais [...]*. São José do Rio Preto: UNESP, 2008.

SCHMITT, N. Size and depth of vocabulary knowledge: what the research shows. *Language learning*, Ann Arbor, v. 64, n. 4, p. 913-951, Dec. 2014 DOI: <https://doi.org/10.1111/lang.12077>.

SCHMITT, N. *Researching vocabulary: a vocabulary research manual*. New York: Palgrave Macmillan, 2010. DOI: <http://dx.doi.org/10.1057/9780230293977>

SCHMITT, N.; SCHMITT, D.; CLAPHAM, C. Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language testing*, [S. l.], v. 18, n. 1, p. 55-88, 2001. DOI: 0265-5322(01)LT1970A.

SOUZA, A. Is Brazilian Portuguese being taught as a community of heritage language? *Language issues: the ESOL journal*, Birmingham, v. 27, n. 1, p. 21-28, Summer 2016.

SOUZA, R. A. *et al.* Estudo sobre um parâmetro de tarefa e um parâmetro amostral para experimentos com julgamentos de aceitabilidade temporalizados. *Revista de estudos da linguagem*, [S. l.], v. 23, n. 1, p. 215-248, 2015. DOI: <http://dx.doi.org/10.17851/2237-2083.23.1.211-244>

SOUZA, R. A.; SILVA, J. S. Exploring the measurement of vocabulary size to differentiate Brazilian Portuguese-English bilinguals' access to grammatical knowledge in the L2. *Revista Linguística*, Rio de Janeiro, v. 11, n. 1, p. 187-214, jun. 2015. DOI: 10.17074/2238-975X.2015v11n1p187.

SOUZA, R.; SÁ, T. M. M.; AMARAL, L. Movendo a avaliação do português como L2 para além do mal-estar. *Revista da Abralin*, [S. l.], v. 19, n. 3, p. 763-774, 2020.

TRAXLER, M. *Introduction to Psycholinguistics: understanding language science*. Hoboken: Wiley-Blackwell, 2011.

WARREN, P. *Introducing psycholinguistics*. Cambridge: Cambridge University Press, 2013.

Recebido em: 20 de abril de 2021.

Aprovado em: 8 de junho de 2021.

