

O EPPLE COMO INSTRUMENTO DE AVALIAÇÃO DE PROFICIÊNCIA PARA PROFESSORES DE LÍNGUA ESTRANGEIRA: UM PROCESSO DE VALIDAÇÃO

EPPLE AS AN ASSESSMENT TOOL FOR FOREIGN LANGUAGE TEACHERS: A VALIDATION PROCESS.

Priscila Petian Anchieta ¹
Universidade Tecnológica Federal do Paraná

RESUMO

Este artigo traz como foco de investigação a avaliação de proficiência de professores de língua estrangeira (LE) realizada por meio de testes de proficiência. O objetivo do trabalho é o de descrever e contribuir para o aprimoramento do teste escrito do EPPLE (Exame de Proficiência para Professores de Língua Estrangeira), por meio da verificação de aspectos que englobam a validade interna do instrumento. Para isso, dividimos este trabalho em dois momentos. No primeiro momento são abordados assuntos referentes à avaliação de proficiência realizada por meio de testes e, no segundo momento, apresentamos os três grandes pilares que fazem parte da validação de instrumentos de avaliação: a praticidade, a validade e a confiabilidade.

PALAVRAS-CHAVE: avaliação; validade; tarefas; testes de proficiência.

ABSTRACT

This research focus is the proficiency assessment of foreign language teachers that is developed through proficiency tests. The main goal is to describe and improve the written test of EPPLE (Proficiency Exam for Foreign Language Teachers) through analyzing some aspects that englobe the internal validity of this tool. For that, we have divided the article into two parts: In the first part we discuss about some topics related to proficiency assessment and proficiency tests. In the second one, we present the three big cornerstones that are part of assessment tools validation: the validity, the practicality and the reliability.

KEYWORDS: assessment; validity; tasks; proficiency tests.

INTRODUÇÃO

O tema deste artigo se origina da nossa inquietação a respeito da avaliação da proficiência de professores de língua estrangeira (LE). As discussões teórico-críticas aqui desenvolvidas têm como referência o EPPLE (Exame de Proficiência para Professores de Língua Estrangeira), que é um exame de proficiência específico para professores que devem utilizar a língua estrangeira em ambientes de ensino e aprendizagem, no contexto educacional brasileiro.

O EPPLE avalia as habilidades de compreensão e produção linguística em LE, nas modalidades oral e escrita. O exame é composto por dois testes distintos: um de leitura e produção escrita, e outro de compreensão auditiva e produção oral. Nossa pesquisa, por sua vez, está direcionada à análise e validação do teste escrito.

¹ Doutora em Estudos Linguísticos, com foco na área de Ensino e Aprendizagem de Línguas.
(pri_anchieta@yahoo.com.br)

Consideramos o EPPLÉ como um exame de base comunicativa, pois a proficiência do candidato é avaliada por meio do desenvolvimento de tarefas a partir de materiais autênticos e contextualizados. Os critérios de avaliação são predominantemente holísticos², e a avaliação é realizada considerando-se a coleta de amostras da proficiência do candidato com base no seu desempenho em tarefas com propósitos específicos.

Nosso objetivo neste artigo é apresentar conceitos por meio dos quais buscamos definir os diversos aspectos que dizem respeito ao teste, como o método, as tarefas, a confiabilidade, a praticidade e a validade que permeiam a implementação e o aprimoramento de um instrumento de avaliação.

1. DEFINIÇÕES PRELIMINARES

Teixeira da Silva (2004) aponta o trabalho de Hymes (1979) como o primeiro a incorporar a dimensão social e a capacidade de usar a língua ao termo competência, dando origem, assim, ao conceito de competência comunicativa. Para Hymes (1979), mais do que dominar estruturas e vocabulário, para apresentar competência em uma dada língua, é necessário saber quando ou não falar, a quem falar, onde e de que maneira falar. Em 1993, Almeida Filho estabelece algumas considerações a respeito do termo “competências”, levando em conta as diferentes tarefas que o professor desempenha em sala de aula, discutindo, assim, competências voltadas ao professor de línguas.

Almeida Filho (1993) acredita existirem cinco diferentes tipos de competências envolvidas no fazer pedagógico do professor, sendo elas: a aplicada, a teórica, a implícita, a profissional e a linguístico-comunicativa. Vale ressaltar que a competência linguístico-comunicativa que busca ser avaliada no EPPLÉ abrange o uso da língua tanto em contextos gerais, do cotidiano, como em contextos mais específicos, como é o caso das particularidades da fala do professor ao fazer uso da metalinguagem, ao discutir ou tratar de materiais didáticos, e ao fazer uso da linguagem facilitadora.

Em sua obra, Almeida Filho (1993) detalha cada uma das competências, mas, para os objetivos deste artigo, apenas destacaremos a competência linguístico-comunicativa. Essa competência diz respeito ao conhecimento implícito que o professor tem do sistema abstrato da língua e sua capacidade de uso da/na língua-alvo em que atua. Sendo assim, ao considerarmos que a competência linguístico-comunicativa irá permitir que o professor ensine o que sabe sobre a língua, acreditamos que o EPPLÉ, por meio de suas tarefas, pode avaliar tal competência, pois apresenta simulações de atividades desenvolvidas em sala de aula, levando o candidato a reproduzir práticas comuns ao seu contexto de trabalho.

Faz-se importante também discutirmos o termo “proficiência linguística”, que, segundo Scaramucci (2000), se caracteriza em duas modalidades: a do uso técnico e a do uso não técnico. O sentido não técnico diz respeito a julgamentos de teor impressionista, baseados em visões mais holísticas da linguagem, e conduz à interpretação de proficiência como um conceito estável, único. Geralmente, o termo proficiência, visto sob essa perspectiva, é baseado em descritores mais gerais e representa uma separação entre indivíduos proficientes e não proficientes. Em seu sentido técnico, o termo proficiência é utilizado no âmbito de controle operacional da língua, variando conforme a especificidade da situação de uso da mesma, englobando uma gradação de níveis, ou seja, oscila do menos proficiente ao mais proficiente.

² Holístico nesse contexto de avaliação faz referência à análise e à consideração de todos os intervenientes que fazem parte da produção linguístico-comunicativa.

No construto do EPPLÉ, o termo proficiência está embasado pela definição de Scaramucci (2000) dentro do seu sentido técnico, visto que a proficiência é definida a partir do contexto em que o falante se encontra, ou seja, proficiência para desenvolver determinadas tarefas, como a de interpretar um texto e escrever uma redação.

A partir do momento em que consideramos o contexto específico de uso da língua, podemos também concordar com a autora quando ela afirma existirem níveis de proficiência e não apenas uma proficiência absoluta. De acordo com Scaramucci (2000), a proficiência só deve ser definida a partir do contexto específico de ensino-aprendizagem vivenciado pelos alunos e dentro do qual ocorrem as possíveis interações. Assim, devem-se avaliar os aspectos da língua não de forma isolada e indivisível, mas de forma a abranger toda a sua complexidade, que é refletida na comunicação.

O presente artigo está organizado em dois momentos. No primeiro, apresentamos questões referentes a testes de proficiência, como é o caso de testes para fins específicos, o método e as tarefas de um teste. Em um segundo momento, apresentamos os três grandes pilares que fazem parte do processo de validação de um instrumento: a confiabilidade, a praticidade e a validade.

2. Momento I

2.1 Avaliação: testes de proficiência

Ao longo da história, as pessoas têm sido colocadas à prova para confirmar suas capacidades ou referências em vários aspectos da vida social, como é o caso do profissional da área de educação, que muitas vezes precisa se submeter a uma prova de concurso, a um exame de proficiência, a uma entrevista de trabalho.

Utilizamos a definição de teste trazida por Brown (2005, p. 3), segundo o qual o teste “é um método para medir a habilidade, o conhecimento ou o desempenho de uma pessoa em uma determinada área.” Testes para verificar como uma pessoa atua, particularmente em relação a um nível limite de ação, tornaram-se instituições sociais importantes e exercem a função de *gatekeeping*, ou seja, de porta de entrada, no sentido de que controlam o acesso a vários papéis sociais importantes (McNAMARA, 2000).

De acordo com McNamara e Roever (2006), na área de formação de professores, as consequências da avaliação de LE são particularmente óbvias, pois a avaliação em contextos educacionais continua a levantar questões sobre seus efeitos sociais e consequenciais. De acordo com Bachman (1990), testes de língua estrangeira podem ser utilizados principalmente para as funções de: a) apresentar informações para que decisões sejam tomadas dentro de um contexto educacional, a partir dos seus resultados e b) servir como indicadores de habilidades de interesse para pesquisas na área de ensino e aprendizagem de línguas. Seguindo as ideias de Bachman (1990), acreditamos que o EPPLÉ poderá atuar também como um indicativo de habilidades de interesse ao contexto de ensino e aprendizagem de línguas.

Scaramucci (1999) estabelece algumas características que compõem um teste de base comunicativa e que são identificadas também no construto do EPPLÉ. De acordo com a autora, essas características são: a) ênfase na comunicação, através de situações reais de interação, sendo que os instrumentos para avaliação buscam coletar amostras de tarefas que apresentam o uso real da língua; b) conteúdos autênticos e contextualizados, (o termo autêntico tem sido usado em oposição à linguagem artificial utilizada em livros didáticos e exames); c) os critérios de avaliação utilizados são holísticos, tomando como base o desempenho global do examinando dentro do

desempenho do objetivo da comunicação a ser atingido, que serve como parâmetro para correção da prova; d) o resultado da avaliação é expresso em faixas de proficiência.

Tendo levantado algumas características que compõem testes de proficiência de base comunicativa, apresentamos, a seguir, alguns aspectos presentes em testes para fins específicos, categoria na qual também classificamos o EPPLÉ.

2.2 Testes de proficiência para fins específicos

Segundo Douglas (2000), nos instrumentos de avaliação de proficiência para fins específicos, os conteúdos e métodos resultam da análise da situação de uso da língua-alvo para um fim específico, sendo assim, as tarefas e o conteúdo da avaliação permitem a interação entre a proficiência e o conhecimento de conteúdos específicos dos avaliados.

Essa característica se faz presente no EPPLÉ uma vez que, por meio de temas relacionados ao contexto de atuação dos professores de línguas, os candidatos devem desenvolver tarefas que cabem a esse profissional, como é o caso da leitura de um texto teórico em Linguística Aplicada, da correção de erros, da escrita de uma carta a um coordenador da área de educação, da identificação de desvios gramaticais, de elaboração de uma explicação, de dar instruções, de discutir teoria.

Além disso, as tarefas buscam avaliar habilidades tais como compreender e produzir textos na área de atuação, responder a questões de compreensão de texto em formato dissertativo, identificar informações explícitas e implícitas, desenvolver uma compreensão detalhada de determinado conteúdo, parafrasear textos escritos, identificar erros gramaticais, construir questões, escrever sobre teoria, utilizar a metalinguagem, dar instruções.

Com isso, trazemos a perspectiva de Alderson e Banerjee (2001), de que tanto o conteúdo quanto o método do teste estão presentes em um contexto específico de uso da língua e não de situações gerais. Assim, para que inferências possam ser feitas sobre a capacidade de um candidato usar a língua em contextos específicos, é preciso envolvê-los em tarefas que representem um contexto de forma autêntica. A tentativa de reproduzir atividades normalmente desenvolvidas em contextos de ensino e aprendizagem de línguas, ainda que não correspondam a representações fiéis de uso da língua, contribui para que as tarefas do teste sejam mais autênticas.

Assim, a tentativa de associar as tarefas do teste a tarefas realizadas fora do teste é positiva, pois leva o candidato a por em prática habilidades que supostamente fazem parte do seu dia-a-dia, o que poderá propiciar a coleta de amostras mais autênticas da proficiência. Ao discutir sobre contextos de comunicação, faz-se necessário considerar também o fato de que o professor de línguas está inserido em um contexto comunicativo em que deverá também tratar da língua em geral, ou seja, espera-se que o professor, para ensinar um dado idioma, seja capaz de utilizá-lo em contextos não só específicos, mas também gerais de uso da língua-alvo uma vez que ele, possivelmente, ensinará seus alunos por meio de tópicos comuns do dia-a-dia, como viagens, tarefas de casa, hotéis e atividades cotidianas.

Podemos assim afirmar que o EPPLÉ é um exame para fins específicos, na medida em que considera o domínio de uso da língua do professor como sendo amplo, e esse uso amplo se dá também por meio de habilidades específicas.

Para melhor compreendermos aspectos que envolvem a elaboração de um instrumento como o EPPLÉ, discutiremos também a respeito do método dos testes de proficiência, embasando-nos em Genesee e Upshur (1996).

2.3 O método dos testes de proficiência

Concordamos com Genesee e Upshur (1996) quando afirmam que um teste é um método de coleta de informações. Em termos educacionais, um teste deve apresentar um dado conteúdo a ser avaliado e deve também apresentar um conjunto de tarefas que permitem a observação do comportamento (linguístico) do avaliado. Para Genesee e Upshur (1996), a importância do conteúdo de um teste é a de que o desempenho individual do avaliado será reflexo do seu conhecimento e das suas habilidades voltadas para tal conteúdo, o que permitirá a elaboração das faixas de proficiência. Os autores afirmam que cada tarefa é como um teste que permitirá realizar inferências sobre as habilidades linguísticas do candidato. Desse modo, segundo os autores, os métodos de um teste equivalem “aos tipos de tarefas que os candidatos são convidados a desenvolver”.

Acreditamos que os métodos dos testes podem trazer efeitos nas faixas de proficiência independentemente do conteúdo que eles trazem, pois ao responder a uma questão de múltipla escolha, por exemplo, o candidato estará desempenhando também a habilidade de eliminar respostas incorretas, de fazer inferências e de escolher a resposta mais apropriada. Assim sendo, um candidato que é mais familiarizado com questões de múltipla escolha poderá obter um melhor desempenho do que aquele que não o é.

Para melhor compreendermos a natureza das questões que podem aparecer em testes de proficiência, utilizamos a definição de tarefas trazida por Willis (1986), segundo a qual tarefas são atividades em que a língua-alvo é usada pelos aprendizes com objetivos de comunicação para alcançar um resultado. Em nosso contexto de avaliação, essas atividades são questões de um teste escrito de proficiência, empregadas com propósitos comunicativos que contribuirão para que um recorte linguístico da proficiência do candidato seja obtido.

Bachman e Palmer (1996) apontam dois domínios de tarefas diferentes: as tarefas de teste, como o próprio nome define, e as tarefas de uso da língua, não necessariamente utilizadas em avaliações formais. Essa combinação permite elaborar ou selecionar tarefas de teste correspondentes a tarefas de uso da língua, trabalhadas previamente no processo de ensino e aprendizagem e, ainda, a maneira como as tarefas de teste se caracterizam auxilia a determinar a capacidade linguística do avaliado. Assim, a equivalência entre as tarefas de teste e as tarefas de uso da língua pode determinar a autenticidade e a validade da tarefa de teste por oferecer ao candidato condições para expressar sua capacidade linguística por tarefas já experimentadas.

Para o Genesee e Upshur (1996), as características das tarefas de testes de proficiência podem ser descritas como: a) *close ended*: são mais limitadas e com respostas mais previsíveis, como as questões de múltipla escolha. Segundo os autores, elas avaliam primordialmente aspectos de compreensão; b) *open ended*: são menos previsíveis, como, por exemplo, o caso das questões dissertativas; c) *limited*: são as que transitam entre as duas anteriores, consideradas medianas. Aqui, um leque de respostas estará disponível para o candidato.

No teste escrito do EPPLE, os candidatos se deparam com tarefas do tipo *limited response tasks* quando, por exemplo, lhes é perguntado sobre qual sentença apresenta um erro gramatical a ser corrigido, mas não se explicita qual é o erro e não são elaboradas alternativas. Nesse caso, além de apontar o erro, o candidato deverá também corrigi-lo. Sendo assim, seguindo as ideias de Genesee e Upshur (1996), temos um caso de *transformation task*, em que o avaliado deve transformar a sentença de acordo com as instruções do examinador, como é o caso da reestruturação gramatical feita após a correção de erros.

Já no caso de uma tarefa do EPPLE em que o candidato deve fornecer um *feedback*, uma explicação para seu aluno, apontando qual é o desvio que ele cometeu e de que forma ele poderia ser corrigido, temos, de acordo com Genesee e Upshur (1996), uma *imitation task*, em que o avaliado é convidado a reproduzir, por meio da escrita, aquilo que o enunciado indicou. Nesse contexto, é necessário imitar a postura de um professor, utilizando a metalinguagem para construir uma pequena explicação.

Podemos também encontrar no teste escrito do EPPLE dois tipos de *open-ended response tasks*: as perguntas dissertativas e a composição escrita. Nesse tipo de tarefas, de acordo com Genesee e Upshur (1996), as respostas não são limitadas pelo teste, visto que os candidatos são livres para elaborar uma variedade de possíveis respostas. Concordamos com os autores quando eles afirmam que esse tipo de questão apresenta maior complexidade ao ser corrigida, sendo assim, é necessário o desenvolvimento de um protocolo de avaliação que especifique as regras utilizadas para corrigir cada item da questão, ou seja, é necessário que os critérios de avaliação sejam bem definidos e que todas as correções, de todos os exames, sigam esses critérios.

Ao prosseguir com nossa discussão, apresentamos o “Momento II” deste artigo que aborda os três grandes pilares que envolvem o processo de validação de um instrumento de avaliação: a confiabilidade a praticidade e a validade.

3. Momento II: a confiabilidade

Para um teste de proficiência ser considerado válido é preciso também que ele seja confiável. A confiabilidade diz respeito à consistência ou estabilidade dos resultados obtidos em um teste quando aplicado em ocasiões diferentes. Erros poderão aparecer durante o processo avaliativo e algumas medidas de confiabilidade podem contribuir para a elaboração de uma estimativa de quais foram os erros nos resultados obtidos, sendo que quanto mais alta a medida de confiabilidade, mais baixo será o índice de erro e mais confiável será o resultado.

De acordo com Shohamy e Inbar (2006), existem alguns critérios que podem ser utilizados para testar a confiabilidade de um teste: 1) a confiabilidade entre avaliadores, quando há um acordo entre os examinadores sobre os critérios de avaliação e os resultados de um teste; 2) confiabilidade intra-avaliador, quando o mesmo avaliador, apesar de possíveis problemas, é capaz de emitir os mesmos resultados avaliativos para um candidato; 3) confiabilidade na aplicação-reaplicação, quando os resultados de um teste são estáveis e aplicações são feitas em momentos distintos; 4) consistência interna: quando os itens do teste medem as mesmas características.

Por meio desses critérios, podemos concluir que a confiabilidade de um teste depende, em grande parte, do quanto seus examinadores foram treinados e estão preparados para avaliar os candidatos de maneira coerente, apesar das diferenças entre eles, da influência de fatores externos e do tempo entre diferentes aplicações de um mesmo teste.

Um ponto positivo dos testes ou exames eletrônicos, como é o caso do EPPLE, e que contribui para a confiabilidade, é que existe uma regularidade na forma com que as informações são transmitidas aos candidatos, ou seja, todo insumo apresentado pelo teste é regular, uma vez que não existem examinadores ou aplicadores e, sim, uma aplicação padronizada, que acontece por meio de uma máquina.

A questão do tempo também não varia, uma vez que, a partir do momento em que o candidato inicia a prova, o tempo começa a ser contado e o sistema irá ser encerrado assim que o tempo se esgotar, independente de o candidato ter ou não concluído o teste. Tais aspectos contribuem para que a avaliação mediada pelo computador seja mais confiável e sem variações,

como possivelmente aconteceria em um teste aplicado por examinadores que não respeitassem o cumprimento do tempo pré-estabelecido ou que oferecessem instruções de formas variadas aos candidatos, o que influencia diretamente no resultado dos dados.

Podemos considerar alguns fatores que podem afetar a confiabilidade de um teste, segundo Weir (2004): 1) A consistência no conteúdo das questões, no formato e na duração do exame; 2) O grau de familiaridade dos candidatos com o formato do teste; 3) Os fatores relacionados à administração do teste, como as condições do local do teste (iluminação, assentos, ambiente acústico, equipamentos eletrônicos) e o modo como o examinador administra a aplicação; 4) Os fatores afetivos tais como o nervosismo e a ansiedade; 5) O corretor ou a equipe de correção.

As questões relacionadas às condições de aplicação do teste são de grande valia quando consideramos as diferenças existentes entre os testes em formato papel e aqueles em formato eletrônico. Ao mesmo tempo em que as condições do local podem ser igualmente afetadas em ambos os contextos, nas avaliações em formato eletrônico temos a questão da manutenção e qualidade dos equipamentos eletrônicos, ou seja, o teste apenas será aplicado de forma confiável se os equipamentos funcionarem corretamente. Dessa forma, encontramos aqui um pré-requisito para que esse tipo de avaliação aconteça, uma vez que, caso não existam equipamentos adequados, a aplicação da prova se torna inviável.

No item sobre a familiaridade do candidato com o teste, acredita-se que a prova deve ser divulgada juntamente com informações sobre seu construto, com a finalidade de possibilitar melhor preparo dos seus candidatos. No caso do EPPLE, temos um site de divulgação (www.epplebrasil.org) que contém, além de um simulado da prova, vários artigos publicados pelo grupo de pesquisa que têm contribuído para a elaboração e o aprimoramento do exame.

Já sobre a consistência no conteúdo das questões, no formato e na duração do exame, acreditamos que, por ser aplicado em formato eletrônico, o EPPLE pode ser um instrumento confiável, pois o tempo é controlado pelo computador e o insumo é padronizado. Vale lembrar que um teste eletrônico exige também que o candidato possua familiaridade com o computador.

Entretanto, é certo dizer que a avaliação em formato papel também apresenta algumas exigências, como é o caso de escrever de forma legível, o que pode exigir mais tempo do candidato. Ainda que existam os equipamentos necessários, antes da execução de um teste eletrônico é preciso que tudo esteja devidamente testado e instalado, pois problemas na máquina durante a aplicação da prova poderão influenciar em sua confiabilidade, afetando questões de tempo e ocasionando maior ansiedade e nervosismo dos candidatos.

Em relação ao corretor ou a equipe de correção, concluímos que, no formato eletrônico, não é necessária a figura de um examinador que aplique o teste oral ou que coloque um áudio para que os candidatos possam realizar dadas tarefas, uma vez que todo o teste já contém os dispositivos necessários de forma padronizada.

É importante esclarecer que um teste pode ser confiável e, ao mesmo tempo, pode não ser válido. Isso acontece quando, apesar de os resultados serem confiáveis, o conteúdo e as tarefas do teste não refletem os objetivos do seu elaborador, ou seja, não fazem referência ao seu construto. Dessa forma, é preciso explicitar ao candidato em quais aspectos ele estará sendo avaliado.

Como consequência, a cada mudança que ocorre no exame, seu construto também deve sofrer alterações. Isso justifica o fato de o construto ser elaborado antes mesmo da construção do teste, uma vez que, depois de concluído, sem que tenha havido mudanças naquilo que se deseja

avaliar, o teste deve fornecer resultados similares caso aplicado em diferentes momentos, aos mesmos candidatos.

3.1 A praticidade

Para que um teste seja considerado prático é preciso que ele seja desenvolvido, aplicado e corrigido utilizando tempo e recursos que, de fato, possam ser disponibilizados. Sendo assim, deve-se averiguar se a quantidade de dados e de resultados que o teste fornece sobre o perfil linguístico-comunicativo de seus candidatos é condizente ao tempo utilizado para o seu desenvolvimento, administração, correção e processamento de resultados (WEIR, 2004).

No caso de testes em meios eletrônicos, de acordo com Rosa e Maltempi (2006), eles trazem algumas vantagens em relação àqueles utilizados presencialmente, uma vez que, nos primeiros, o professor pode acessar o progresso ou o nível de proficiência dos alunos com maior frequência; os alunos podem monitorar seu próprio progresso, como acontece com os portfólios; o feedback pode ser instantâneo, como é o caso de alguns testes de proficiência, além de possibilitar o uso de artefatos tecnológicos. Também no que diz respeito à praticidade do processamento de dados, as avaliações eletrônicas podem apresentar maior praticidade e rapidez na interpretação e correção de dados, caso haja programas adequados que facilitem essas operações.

Apesar de toda a praticidade que um teste em formato eletrônico pode trazer, ele se torna inviável e não confiável caso os equipamentos não sejam corretamente utilizados, além de existir um custo para a utilização de tais equipamentos. Um teste extremamente caro para ser aplicado não é prático, assim como não o é um teste que seja complexo de ser corrigido e interpretado.

Concordamos com a ideia de Morrow (2004) de que pode existir uma certa tensão entre validade, confiabilidade e praticidade, uma vez que um teste pode ser válido e confiável e, ao mesmo tempo, pode não ser prático, por ser de alto custo. Aos elaboradores cabe a responsabilidade de lidar com essa situação, buscando um equilíbrio que respeite os objetivos do teste.

3.2 A validade

Instrumentos de avaliação, de forma geral, são baseados em inferências sobre um determinado critério, visto como o conjunto de comportamentos que se deseja avaliar. Esses comportamentos, quando aparecem em situações reais de comunicação, são dificilmente observáveis. Sendo assim, a forma que encontramos para caracterizá-los acontece por meio da simulação e da representação, de forma amostral, como feito nos testes de proficiência.

Alguns pesquisadores apresentam a definição de validade como sendo em que proporções um teste mede aquilo que se deve medir (FULCHER, 2003); (WEIR, 2004); (FULCHER e DAVIDSON, 2007). Sendo assim, a validade pode ser compreendida como o traço de um teste que define se ele avalia adequadamente o que pretende avaliar. Essa visão deixa claro que, ao se elaborar um teste de proficiência linguístico-comunicativo, existe a intenção de testar recortes da proficiência do candidato obtida por meio da realização de tarefas.

Vale dizer que um teste pode ser válido (quando suas tarefas avaliam realmente aquilo que se almeja), sem ser confiável, como acontece quando os critérios não são claros e quando as pessoas envolvidas na elaboração e na aplicação do teste não são corretamente preparadas para tais atividades.

Ao discutir sobre o conceito de validade, Scaramucci (2009, p. 31) afirma que

as discussões motivadas pela complexidade dos fatores envolvidos nesse conceito foram tão intensas a ponto de determinarem reformulações não apenas das visões de linguagem e de proficiência subjacentes às avaliações tradicionais, mas, principalmente, dos conceitos utilizados como parâmetros para avaliar a qualidade/aceitabilidade desses instrumentos.

Nosso trabalho está inserido nas definições de Kane (2001, 2006) que, ao discutir sobre validação, afirma existirem duas acepções diferentes do termo: a) no sentido de pesquisa de evidências que sirvam de suporte ao uso da prova e às interpretações dos resultados observados – o que ela denomina como argumento de interpretação; b) no sentido de avaliação do uso e das interpretações dos resultados observados nas provas, segundo critérios propostos – que, segundo a autora, constitui o argumento de validade.

Assim sendo, nossa investigação engloba tanto os argumentos de interpretação quanto os argumentos de validade, ao passo que não só analisamos os resultados do teste escrito, como também avaliamos as possíveis interpretações desses resultados, de forma a elaborar críticas e sugestões. O cuidado em “pilotar” o exame apenas em candidatos que preencham os requisitos exigidos pelo EPPLLE, ou seja, aqueles que são professores ou futuros professores de língua inglesa, vem ao encontro das ideias de Cohen (1986), que afirma que um determinado teste ainda que tenha características admiráveis, pode ser considerado inválido se utilizado para propósitos inapropriados.

Assim, as aplicações-piloto do EPPLLE acontecem apenas com alunos do curso de Letras/Licenciatura, professores de língua inglesa em formação, ou com professores que já estão no mercado de trabalho, com o intuito de associar os propósitos do exame aos propósitos do público ao qual é destinado.

Citamos o trabalho de Baffi-Bovino (2010) que realiza uma discussão, com base em Scaramucci (2009), Hughes (1989), Alderson et al. (1995), sobre os diferentes tipos de validade, sendo eles: a validade de face, de conteúdo, de critério, preditiva e de construto. Segundo Baffi-Bovino (2010), a validade de face é aquela que compõe a impressão positiva que os envolvidos têm a respeito do teste. No processo de aplicações-piloto do EPPLLE, a validade de face está sendo verificada por meio de um questionário no qual uma das perguntas diz respeito à opinião dos candidatos sobre o exame como, por exemplo, sobre os enunciados das tarefas e o nível de dificuldade que elas apresentam.

A validade de conteúdo representa o conjunto maior de tarefas das quais o teste deve ser uma amostra, dessa forma, investigamos se o teste escrito do EPPLLE é uma amostra de tarefas que possivelmente o professor de língua inglesa poderá realizar no seu contexto de trabalho. A validade preditiva, segundo Baffi-Bovino (2010), faz referência a quanto o teste antecipa o desempenho em algumas situações subsequentes e, segundo Hughes (1989), faz parte da validade de critério, que se refere a quanto os resultados do teste são semelhantes àqueles obtidos por outra avaliação feita simultaneamente.

Por fim, a validade de construto, segundo Alderson (1995), é o tipo de validade mais complexo de se explicar. Geralmente, a validade de construto é definida como uma descrição adequada de uma teoria subjacente ao que deve ser medido, ou seja, refere-se ao quanto um teste corresponde à implementação de um modelo teórico (BAFFI-BOVINO, 2010); (ALDERSON, 1995); (McNAMARA, 2000); (FULCHER e DAVIDSON, 2007).

Segundo Schlatter et al. (2005), com o surgimento da abordagem comunicativa, na década de 1970, o conceito de validade passou a ser definido de maneira mais abrangente e, sem desconsiderar o lado mais tradicional, apresenta um conceito de validade no qual as dimensões sociais, culturais e os efeitos retroativos são partes integrantes.

Por essa razão, validade e efeito retroativo estão interligados, uma vez que os resultados de um teste poderão ocasionar efeitos não apenas linguísticos, mas também extralinguísticos, como é o caso de uma possível aprovação de um professor de LE em um processo seletivo no qual o teste seja adotado como instrumento de seleção de candidatos.

Concordamos com Martins (2005) quando ela afirma que um exame como o EPPLLE poderia apresentar os seguintes benefícios para o professor e para as escolas ou instituições de ensino:

- a) uma certificação profissional - uma maneira de comprovar, para fins educacionais e profissionais, a proficiência do professor em inglês; b) uma contribuição para a carreira do professor, fornecendo um parâmetro para seu crescimento profissional e pessoal a partir de uma análise de seu desempenho, fornecendo, assim, elementos de uma boa prática de ensino; c) uma representação realista da competência do profissional, tornando seu trabalho mais visível para coordenadores e diretores, fornecendo um quadro real para seleção desses profissionais, plano de carreira e uma revisão de seu desempenho; d) uma forma de promover a imagem pública do ensino de inglês como língua estrangeira nas escolas da rede pública, particulares e centros de idiomas. (p. 94-95)

Dessa forma, acreditamos que um teste válido pode ocasionar mudanças sociais, pessoais e profissionais a seus candidatos. Anchieta (2010) realizou um estudo em que questionários foram aplicados a professores de línguas do curso de Letras de universidades públicas no Brasil, com o intuito de verificar a opinião de cada um deles sobre os efeitos esperados que um exame de proficiência para professores de línguas pode ocasionar. De acordo com os dados da pesquisa, 90% dos participantes acreditavam que um instrumento de avaliação para o professor de língua inglesa poderia contribuir para uma melhor qualificação do profissional.

Alderson, Clapham e Wall (1995) utilizam dois termos para separar os vários conceitos de validade: interna e externa. A validade interna seria aquela relacionada a estudos de conteúdo inferidos a partir do teste e seus efeitos (validade de face, validade de conteúdo, validade de resposta e validade de construto). Já a validade externa estaria relacionada aos estudos que comparam o desempenho dos candidatos em um teste com os resultados de sua capacidade obtidos fora do teste (validade paralela e validade preditiva).

Vale ressaltar que, durante o processo de validação do teste escrito do EPPLLE, não abordamos a validade paralela, que envolve uma comparação de resultados de um teste de um candidato com outros resultados de outros testes formais ou informais e/ou situações na vida real ou em sala de aula dos mesmos candidatos.

3.3 A validade de construto

De acordo com Fulcher e Davidson (2007), a validade de construto é a base de evidência para a interpretação ou uso de um teste. Com base nessa afirmação, concluímos que nosso trabalho

é também o de desenvolver uma discussão teórica que permita favorecer a interpretação e o uso do EPPLÉ, de forma a explicitar seus objetivos, suas tarefas, seu público-alvo, e demais aspectos subjacentes a um exame dessa natureza.

Ao descrever o processo de validação, Bachman (1990, p. 236) considera a validade do construto como um termo guarda-chuva para as outras formas de validação, ou seja, validação de conteúdo e referenciada em critério. Segundo Bachman e Palmer (1996), existem dois propósitos básicos para a descrição de construtos: a) servir de base para o uso do escore do teste; b) servir de guia para a elaboração do próprio teste; c) permitir que tanto elaboradores quanto usuários do teste demonstrem a validade do construto e das interpretações feitas sobre os resultados do teste. Ainda segundo os autores, é preciso decidir quais competências linguísticas serão incluídos na definição do construto.

Por meio das nossas reflexões a respeito do EPPLÉ, almejamos contribuir para o estabelecimento das competências linguísticas que podem ser, de fato, incluídas no construto do exame.

Brown (2000) propõe um experimento para se analisar a validade do construto, segundo o qual são comparadas as atuações de dois grupos, sendo que um grupo tem o construto e o outro não. Se o grupo que possui o construto sair-se melhor do que o outro que não o possui, pode-se dizer, então, que o resultado comprova a validade de construto do teste. Entretanto, o autor acrescenta que não há um único método que possa ser considerado o melhor para investigar validade de construto. Na opinião do autor, a validade de construto de um teste deveria ser demonstrada com um acúmulo de evidências originárias de diferentes perspectivas e estratégias.

Dessa forma, se entendermos que um teste de proficiência escrito em língua estrangeira avalia a capacidade de produzir textos adequados ao contexto e ao interlocutor, de forma a atingir satisfatoriamente seu propósito comunicativo, por exemplo, esse construto deve ser revelado por meio das tarefas do teste.

Testes diferentes são, normalmente, operacionalizações de construtos diferentes (SCHLATTER, et al., 2005). Uma avaliação de produção oral, por exemplo, realizada por meio de testes de pronúncia pode revelar um construto de proficiência desvinculado com o propósito comunicativo, uma vez que, para se contar uma piada na língua-alvo, por exemplo, a pronúncia é um dos conhecimentos exigidos e talvez não seja o mais importante.

4. CONSIDERAÇÕES FINAIS

Tomando como ponto de partida o intuito de contribuir para o processo de validação EPPLÉ, este artigo teve como objetivo elaborar uma discussão crítico-descritiva para investigar aspectos como o método do teste, as tarefas, a confiabilidade, a praticidade e a validade.

Com base nas discussões desenvolvidas, foi possível enfatizar que a elaboração de um teste de proficiência é um processo detalhado e criterioso, que se vale de uma série de pressupostos e pré-requisitos para atingir seu objetivo final. Além disso, vale ressaltar que o processo de validação de um instrumento de avaliação como o EPPLÉ é um processo contínuo e, por este motivo, acreditamos que muitas das considerações que estabelecemos nesta pesquisa poderão contribuir para o aprimoramento do exame.

Espera-se, ainda, que este artigo tenha conseguido esclarecer elementos teórico-críticos capazes de suscitar pesquisas e discussões, além de promover a verificação de aspectos que englobam a validade interna do instrumento, servindo como base para futuras pesquisas análogas.

Objetivamos continuar investigando aspectos que fazem parte da validação de instrumentos de avaliação, visto ser esta uma área de grande importância para o sistema de ensino e aprendizagem de línguas, principalmente para a formação de professores de LE, nosso foco de análise e de pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

ALDERSON, J. C.; BANERJEE, J. Language testing and assessment (part I) (State-of-the-art review). *Language Teaching*, vol. 34, p. 213-236, 2001.

ALDERSON, J. C.; CLAPHAM, C.; WALL, D. *Language test construction and evaluation*. Cambridge Language Teaching Library. Cambridge: CUP, 1995.

ALMEIDA FILHO, J.C.P. *Dimensões comunicativas no ensino de línguas*. Campinas: Pontes Editores, 1993.

ANCHIETA, P. P. *Análise de testes de proficiência em língua inglesa: subsídios à elaboração de um exame para professores de inglês no Brasil*. 2010. 210 f. Dissertação (Mestrado em Estudos Linguísticos), Universidade Estadual Paulista Júlio de Mesquita Filho, São José do Rio Preto, 2010.

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

BACHMAN, L. F.; PALMER, A. S. *Language Testing in Practice*. Oxford: OUP, 1996.

BAFFI-BONVINO, M. A. *Avaliação da proficiência oral em inglês como língua estrangeira: foco na competência lexical e uma proposta para o processo de validação do descritor "vocabulário" de um teste de proficiência para professores de língua inglesa*. 2010. 366 f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista Júlio de Mesquita Filho, São José do Rio Preto, 2010.

BROWN, J. D. *Testing in language programs: A comprehensive guide to English language assessment* (New edition). New York: McGraw-Hill, 2005.

COHEN, S. Knowledge and Context. *Journal of Philosophy*, New York n. 83, p.574-583, 1986.

CONSOLO, D. A.; GUERREIRO, G. M. S.; HATUGAI, M. R. Mapeamento da competência linguístico-comunicativa em línguas estrangeiras de alunos de Licenciatura em Letras. Relatório de Estágio de Iniciação Científica. São José do Rio Preto: UNESP, 2001.

CONSOLO, D. A.; TEIXEIRA DA SILVA; V. L. The TEPOLI test: Construct, updated tasks and new parameters to assess EFL teachers' oral proficiency. Anais do I Congresso Internacional da ABRAPUI. Belo Horizonte: UFMG, 2007. (CD-ROM).

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

FULCHER, G.; DAVIDSON, F. Language Testing and Assessment. *An Advanced Resource Book* (Routledge Applied Linguistics Series). London & New York: Routledge. 2007.

GENESE, F.; UPSHUR, A. J. *Classroom-based evaluation in Second language education*. NY: Cambridge University Press, 1996.

HUGHES, A. *Testing for language teachers*. Cambridge: CUP, 1989.

HYMES, D. H. On communicative competence. In: Brumfit, C.; Johnson, K. (Eds.), *Communicative Approach to Language Teaching*. Oxford: Oxford University Press, 1979.

KANE, M. Validation. In: BRENNAN, R. (Ed.). *Educational Measurement, Fourth Edition*. Westport, CT: American Council on Education and Praeger, 2006. p. 17-64.

_____. Current concerns in validity theory. *Journal of Educational Measurement*, Madison n. 38, p. 319-342, 2001.

MARTINS, T. H. B. *Subsídios para a elaboração de um exame de proficiência para professores de inglês*. 2005, 112f. Dissertação (Mestrado em Linguística Aplicada) - Instituto de Estudos da Linguagem, UNICAMP, Campinas, 2005.

McNAMARA, T. *Language Testing*. Oxford: OUP, 2000.

McNAMARA, T. F.; ROEVER, C. *Language Testing: The Social Dimension*. Oxford: Blackwell. 2006.

MORROW, J. Conservatism, Authority and Tradition. In: GREGORY, C. (Ed.). *Encyclopedia of Nineteenth Century Thought*. London: Routledge, 2004. p. 91-98.

ROSA, M.; MALTEMPI, M. V. *A avaliação vista sob o aspecto da educação à distância*. Rio de Janeiro, 2006.

SCARAMUCCI, M. V. R. Validade e impactos sociais da avaliação em contextos de línguas diversos (Port LM/L2 e Inglês LE). Projeto de pesquisa em andamento na Universidade Estadual de Campinas. Campinas, 2009.

_____. Proficiência em LE: considerações terminológicas e conceituais. *Trabalhos de Linguística Aplicada*, n. 36, p. 11 - 22, 2000.

_____. Avaliação: mecanismo propulsor de mudanças no ensino/aprendizagem de língua estrangeira. *Contexturas*, São Paulo. n. 4, p. 115-124, 1999/2000.

SCHLATTER, M.; GARCEZ, P. Línguas Adicionais (Espanhol e Inglês). In: Rio Grande do Sul, Secretaria de Estado da Educação, Departamento Pedagógico. (Org.). *Referenciais curriculares do Estado do Rio Grande do Sul: linguagens, códigos e suas tecnologias*. Porto Alegre: Secretaria de Estado da Educação, Departamento Pedagógico, v. 1, 2005, p. 127-172.

SHOHAMY, E.; INBAR-LOURIE, O. Assessment of advanced language proficiency: Why performance-based tasks? CALPER Professional Development Document (CPDD) 0603. 2006. Disponível em: <<http://calper.la.psu.edu/publications.php>>. Acesso em: 15 de janeiro de 2014.

TEIXEIRA da SILVA, T. Resenha do livro de Maria Helena Vieira-Abrão *Prática de Ensino de Língua Estrangeira: experiências e reflexões*. *Revista do ISAT*, n. 3, p.155-157, 2004.

THORNE, S. L. Pedagogical and praxiological lessons from Internet-mediated intercultural foreign language education research. In: BELZ, J.; THORNE, S. L. (Eds.). *Internet-mediated intercultural foreign language education*. Boston: Thomson Heinle Publishers, 2006. p. 2-30

WEIR, S. *Analysis of school attendance data at primary and post-primary levels for 2003/2004*. Report to the National Educational Welfare Board. Dublin: Educational Research Centre, 2004.

WILLIS, J. *A Framework for Task-based Learning*. Harlow: Longman, 1986.

Recebido em 10/01/2017

Aceito em 24/05/2017

Publicado em 05/06/2017