

# CRITÉRIOS DE AVALIAÇÃO EM TESTES DE PROFICIÊNCIA A DISTÂNCIA

Evaluation criteria in distance proficiency tests

Simone Correia Tostes (CEP)

## RESUMO

Avaliar a fluência oral numa língua estrangeira é reconhecidamente uma tarefa extremamente difícil (cf. LUOMA, 2004). A inserção de descritores linguísticos no modelo de verificação e a adoção de situações que visem ao uso da língua para realizar funções comunicativas possibilitam a avaliação mais ponderada da fluência do falante de LE. Este trabalho pretende relatar uma experiência positiva de implementação desse método.

Palavras-chave: Avaliação. Proficiência. Descritores. Comunicativo. Oral.

## ABSTRACT

To evaluate oral proficiency in a foreign language is acknowledged as an extremely difficult task (cf. LUOMA, 2004). The insertion of language descriptors in the verification model and the adoption of situations that lead to the use of language for communicative purposes enable a more balanced evaluation of a FL speaker's fluency. This work aims at reporting a successful experience with this method.

Keywords: Evaluation. Proficiency. Descriptors. Communicative. Oral.

## INTRODUÇÃO

Em sua vida cotidiana, o indivíduo é constantemente testado. Desde testes médicos a testes para verificar a capacidade profissional ou para uma promoção, esses recursos são sempre critérios valiosos a serem utilizados para assessorar uma decisão sobre a vida ou a carreira das pessoas. Dentro desse esquema, é espantoso que tão pouco se saiba a respeito do processo de avaliação concebido de maneira geral (cf. Macnamara, 2000, p. 3). Normalmente, esse conhecimento é confiado a um grupo seletivo de especialistas especificamente designados para a elaboração e a aplicação de testes.

A complexidade do processo de avaliação da fluência num idioma estrangeiro advém da própria natureza da tarefa. Trata-se de estabelecer padrões de competência mínimos de acordo com tarefas e funções comunicativas pré-estabelecidas conforme um dado parâmetro considerado. Várias tentativas foram feitas para se estabelecerem parâmetros ideais que dêem conta de toda a gama de habilidades linguísticas do aprendiz de uma língua estrangeira.

Definir o que seja *dominar* um idioma estrangeiro configura-se como tarefa extremamente complexa. Por sua vez, essa concepção não pode ser desvinculada dos usos da linguagem para comunicação em situações cotidianas. O conceito de *domínio* de uma língua estrangeira varia de acordo com a natureza das interações que se vislumbra. Num restaurante ou num balcão de atendimento de passageiros, a amplitude do domínio pode ser bem mais restrita e simplificada do que a estabelecida para funções diplomáticas, por exemplo. Para esta última, habilidades como analisar, negociar e convencer vão bem além do simples fornecimento de informações de escopo restrito que as funções de atendentes de restaurante ou balcão de passageiros requerem.

No âmbito do Exército Brasileiro, existem algumas tentativas de se estabelecerem padrões aceitos internacionalmente para descrever os diferentes estágios de fluência num idioma estrangeiro. Atualmente, tem-se adotado como critérios de avaliação os descritores linguísticos. Esses permitem que, de acordo com a produção oral evidenciada em interações comunicativas, o avaliado seja classificado num nível que retrate seu perfil linguístico. Os descritores foram implementados em avaliações reconhecidas internacionalmente, como os exames de Cambridge, por exemplo, que se valem dos parâmetros de descritores linguísticos preconizados na *Common European Framework of Reference for Languages*, publicada pelo Council of Europe.

No Exército Brasileiro, exames de proficiência oral são realizados pelo Centro de Estudos de Pessoal (CEP, daqui em diante). Esses exames são, em última instância, os instrumentos que separam um militar bem sucedido na carreira de missões e cursos no exterior. Essas oportunidades de aperfeiçoamento profissional coroam uma carreira de dedicação e competência profissional. Por esse motivo, dominar um idioma estrangeiro tem sido, há quase duas décadas, a diretriz dos comandantes da Força Terrestre. Ênfase especial é dada aos idiomas inglês e espanhol: o inglês configura-se como a língua franca internacionalmente reconhecida, e o espanhol favorece a integração do Brasil com os demais países do Mercosul.

A experiência com testes de proficiência oral do CEP é pioneira e digna de destaque por possibilitar a verificação de aspectos da conversação, em língua estrangeira a distância em tempo real. Apesar de ser uma ferramenta conhecida na área de educação a distância, não existem registros de uso da videoconferência para a realização de exames de proficiência em língua estrangeira.

A aprovação no Exame de Proficiência Oral (EPO, daqui em diante) habilita o profissional no mais alto nível de proficiência linguística reconhecido no Exército. Isso

significa que o aprovado nesse teste possui as habilidades profissionais básicas de utilização do idioma estrangeiro, sendo capaz de responder a perguntas sobre assuntos variados e negociar significados, bem como de expor sua opinião sobre diversos temas relacionados ou não à profissão militar.

## JUSTIFICATIVA

Criado em 1965, o CEP possui uma vasta experiência na área do ensino de idiomas a distância, dentre várias outras especialidades, como os cursos de pós-graduação, realizados em convênio com várias universidades do Rio de Janeiro. O CEP é o único estabelecimento de ensino que atende aos militares brasileiros de todas as regiões do Brasil. Realizar o EPO presencialmente exigiria que o candidato se deslocasse de sua organização militar para o Rio de Janeiro a fim de realizar a prova. Devido a custos operacionais importantes com deslocamentos de pessoal, a videoconferência colocou-se como solução mais viável encontrada. Dessa forma, o candidato necessita tão somente de se inscrever no exame e se encaminhar para o Comando Militar de Área, sede do exame na sua região, na data e hora marcadas para a entrevista.

Esse sistema completou cinco anos e provou ter sido uma experiência inovadora na utilização das tecnologias de informação para avaliação de proficiência oral. Entretanto, alguns problemas foram detectados com relação ao modo de condução desses exames. Por esse motivo, no último ano foram introduzidas modificações que possibilitam condições mais favoráveis de realizar os exames, bem como a aplicação de parâmetros de avaliação mais adequados.

Esta pesquisa vem atender à necessidade de avaliação e validação do sistema adotado, indicada em Carvalho et al. (2007). Conforme aponta Macnamara (2000, p. 48), o foco da

investigação sobre validação deve ser sobre os procedimentos adotados. Se esses forem falhos, as conclusões tendem a ser infundadas. Da mesma forma, a validação de testes envolve refletir sobre a lógica, particularmente sobre seu formato e intenção, bem como incluir lidar com dados empíricos, colhidos em situações de administração do teste. Caso contrário, “sem procedimentos de validação dos testes, existe grande potencial para injustiças” (Ibid.).

Assim, foi realizada uma pesquisa com os indivíduos avaliados neste ano, a fim de verificar-se o nível de adequação desse novo formato de avaliação após serem introduzidas modificações com base em observações dos próprios examinadores sobre exames realizados anteriormente. Os comentários dos profissionais recaíam sobre dois aspectos principais: (1) insegurança em relação aos critérios de avaliação; (2) sensação de injustiça ao adotar procedimentos inadequados que refletiam negativamente no desempenho do candidato.

## NATUREZA DA AVALIAÇÃO

Para caracterizar a natureza da avaliação, será utilizada a distinção apresentada por Macnamara (2000), que oferece dois tipos básicos de testes: os do tipo papel-e-caneta e os testes de desempenho. O autor refere-se ao primeiro como um teste tradicional de perguntas e respostas, que se destina a avaliar componentes estanques da língua estrangeira, tais como aspectos de gramática e coesão textual. Por seguirem um padrão que verifica o domínio de pontos isolados da língua estrangeira, normalmente são os testes que se enquadram no formato “múltipla escolha” (p. 5). Por serem de fácil administração e controle, são, com frequência, os mais adotados e não se prestam, portanto, a verificar aspectos produtivos da competência comunicativa, como a capacidade de persuasão ou negociação, por exemplo.

Os testes de desempenho, dentre os quais se inclui o EPO, são bem mais complexos e vão ao encontro da definição de *competência linguística* (cf. Chomsky, 1965, p. 4), que

engloba não apenas a capacidade de dominar aspectos isolados da gramática, mas também de empregá-los competentemente de acordo com a situação comunicativa que se apresenta. Foi então que o autor fez a distinção clássica entre *competência* e *desempenho*, sendo o primeiro utilizado para designar a capacidade de aplicar regras internalizadas na gramática universal dos indivíduos, e a segunda configura-se como a habilidade de utilizar produtivamente esse conjunto finito de regras linguísticas em um número infinito de possibilidades em situações comunicativas.

Os testes de desempenho permitem, portanto, a verificação de nuances de linguagem, não verificáveis em testes de respostas fechadas. Macnamara (op. cit.) estabelece uma distinção crucial dos testes de desempenho comparados a outras modalidades de testes de língua – eles avaliam a capacidade de o indivíduo *comunicar-se* (p. 6). Por esse motivo, são mais adequados para verificar habilidades produtivas de escrever e falar. Nessa categoria de testagem, ressalta o autor, “um excerto de discurso mais ou menos longo é elicitado do examinado e julgado por um ou mais avaliadores treinados usando um procedimento de avaliação acordado” (Ibid.). Para se atingir a esse objetivo, normalmente utilizam-se tarefas que o avaliado tenha que desempenhar utilizando a linguagem como ferramenta.

Outro tipo de categorização relevante apresentada por Macnamara (op. cit.) refere-se ao propósito a que se destinam os testes, fazendo, portanto, a distinção entre *testes de nível* e *testes de proficiência*. O primeiro tipo está diretamente relacionado à atividade de instrução e normalmente é realizado em momentos pontuais do processo ensino-aprendizagem. Enquanto testes de nível relacionam atividades passadas e tentativas de mensurar um trabalho desenvolvido num período de tempo definido, os testes de proficiência tentam realizar um prognóstico sobre o potencial de uso da língua estrangeira em diversas situações comunicativas. O critério definido passa a ser, nesse caso, o uso real da língua em situações futuras (p. 7). Entretanto, esse tipo de avaliação é inevitavelmente influenciado pelo

fenômeno *Observer's Paradox* (cf. Labov, 1972), isto é, o fato de ter seu comportamento observado modifica o comportamento em si de qualquer indivíduo avaliado.

O EPO se assemelha em alguns aspectos ao *Oral Proficiency Interview* (OPI), que verifica progressivamente as habilidades de comunicação no idioma estrangeiro tomando-se por base a caracterização dos níveis de proficiência nas quatro habilidades linguísticas. O OPI, entretanto, configura-se como um exame que pode ser realizado sem a presença do examinador, que é, necessariamente, um falante nativo da língua-alvo. Essa flexibilidade diminui o custo de aplicação de testes presenciais, pois o entrevistador pode, por exemplo, utilizar-se do recurso de gravação em áudio de perguntas que ensejem uma resposta ou uma argumentação sobre um ponto de vista. Nesse caso, não se verifica uma efetiva interação entre o entrevistador e o entrevistado, uma vez que a “conversa” é elicitada através de recursos artificiais, além de não oferecer a naturalidade de uma conversa real. Por se tratar de um modelo de exame com mais de um avaliador, o EPO se utiliza do padrão de *conceituação analítica*, segundo o qual todos os examinadores avaliam variados aspectos da comunicação que são, então, enquadrados em escalas de comunicação estabelecidas para cada um dos aspectos isoladamente.

## BASE TEÓRICA

A opção por um ou outro modelo de teste de proficiência está relacionada ao tipo de verificação que se pretende. As abordagens de ensino e avaliação atualmente mais utilizadas baseiam-se no modelo de Hymes (1972) de competência comunicativa. Esse conceito foi introduzido na Linguística Aplicada em contraposição às teorias altamente focadas nas estruturas gramaticais. Assim, a noção de competência comunicativa enfoca os usuários da linguagem e os usos linguísticos adequados ao contexto.

Hymes (op. cit.) sugeriu quatro níveis de análise que permitem compreender regularidades nos usos da linguagem: (1) o nível *gramatical*, que dá conta de explicar o sistema de regras da língua; o nível da *possibilidade*, que inclui restrições de tempo e processamento; (3) o nível da *adequação*, que trata do que é apropriado em diversas situações de uso; (4) o nível do *uso*, que explica por que determinadas estruturas são mais usuais do que outras.

Embora tenha sido elaborado para analisar o desenvolvimento da língua materna, esse modelo tem sido cada vez mais utilizado em contextos de aprendizagem de língua estrangeira. Essa teoria vem sendo adaptada para dar conta de situações mais concretas como a avaliação, sendo o padrão de Bachman & Palmer (1996) o mais utilizado nos dias de hoje.

Bachman & Palmer (op. cit.) entendem o uso da linguagem como interação entre participantes num determinado contexto. Sua concepção de linguagem é constituída de duas partes: o conhecimento linguístico e a competência estratégica. Dos componentes do conhecimento linguístico, destacam-se duas categorias principais – o conhecimento organizacional e o pragmático. Este envolve o relacionamento entre a forma, a intenção do falante e o contexto; aquele tem a ver com o modo como as falas e textos são organizados.

A competência estratégica, por sua vez, envolve as *funções* que as pessoas realizam através da linguagem. Esse conceito passa pelas ideias de Halliday (1976), que identifica as seguintes funções de uso da linguagem: (a) ideacionais – expressam experiências sobre o mundo real; (b) manipulativas – visam a ter um efeito sobre o mundo; (c) heurísticas – estendem o conhecimento das pessoas para o mundo; (d) imaginativas – incluem o uso criativo da linguagem para fins estéticos ou humorísticos.

Esse entendimento sobre capacidade linguística é altamente favorável como pano de fundo de avaliações orais em línguas estrangeiras, especialmente para auxiliar com definições sobre o que se pretende avaliar ou não em determinado procedimento. A limitação do modelo,



entretanto, está relacionada ao seu foco nos aspectos linguísticos da situação de uso da língua, bem como de sua concepção estática da comunicação, restringindo-se a comportamentos verificáveis de uso da língua. Por esse motivo, outros aspectos importantes passam a receber menos ênfase, uma vez que ocorrem na mente do falante.

## MODIFICAÇÕES INTRODUZIDAS

Nos 4 primeiros anos de utilização do EPO, os critérios de avaliação foram aplicados de acordo com parâmetros que mediam, prioritariamente, a quantidade de erros verificados pela banca examinadora durante a execução da avaliação. De maneira geral, pode-se afirmar que os examinadores não possuíam em mente a imagem do falante competente em inglês. Assim, o foco da avaliação recaía sobre o que o candidato era incapaz de realizar, em detrimento dos aspectos que caracterizavam a capacidade de expressar e realizar funções comunicativas na língua estrangeira.

Os critérios de avaliação eram distribuídos em cinco: pronúncia/entonação, gramática, fluência, coerência, vocabulário e compreensão auditiva. Dentro desse esquema, os professores marcavam em suas planilhas a quantidade de erros que eram evidenciados durante a entrevista. O resultado desse tipo de avaliação refletia uma disparidade aparente entre a performance e o resultado final do avaliado. Por vezes, este era totalmente limitado na sua produção oral, emitindo frases curtas e desconexas, com longas pausas durante sua produção oral, com limitações de ordem pragmática para realizar o que lhe era solicitado. Entretanto, devido ao baixo índice de erros gramaticais, de pronúncia ou de coerência, era favorecido no resultado final e aprovado no exame.

Por outro lado, houve candidatos reprovados no exame que, apesar de demonstrar visível desenvoltura para se expressar na língua inglesa, exibiam controle mediano de regras

gramaticais, assim como emitiam palavras com pronúncia incorreta. Apesar da frequente incorreção gramatical, muitas vezes o desempenho desses candidatos superava a do perfil do candidato descrito anteriormente. Entretanto, como a soma total de pontos concedidos não era suficiente para atingir a média mínima para aprovação, eles eram reprovados.

Preocupados com tal disparidade de resultados, os professores resolveram fazer adaptações ao formato preconizado para a prova. Assim, em vez de fazer perguntas de nível de produção altamente estruturada desde o início da avaliação, foi inserido o *icebreaker*, uma conversa bastante breve que serve para ambientar o aluno ao que será realizado durante a entrevista, assim como se presta a acalmar aqueles candidatos mais ansiosos, fazendo com que se sintam mais à vontade durante o tenso momento da avaliação.

Além do *icebreaker*, foi elaborado um conjunto de questões sobre os possíveis temas de discussão da prova. Dessa forma, mesmo antes de se iniciar a conversa, o professor entrevistador dispunha de um rol de perguntas, que evoluíam das mais simples e objetivas para as mais complexas e subjetivas. Assim, tornou-se possível observar uma progressão no nível de dificuldade da prova, permitindo aos candidatos de níveis mais incipientes interagirem com os examinadores, ainda que em nível básico. Conseqüentemente, dependendo da evolução do candidato, a entrevista poderia ser interrompida após o primeiro nível de perguntas, sem causar constrangimentos ao candidato.

Outra modificação introduzida foi o trabalho com descritores linguísticos. Ao invés de contar a quantidade de erros de cada candidato, os examinadores atribuíram pontos de acordo com o perfil linguístico demonstrado pelo examinado. Outra modificação introduzida foi a de designar apenas um entrevistador para interagir todo o tempo com o candidato. Esse examinador não se encarregaria de atribuir pontos ou avaliar o desempenho do candidato, mas limitar-se-ia a conduzir a conversa da maneira mais natural e descontraída possível. Com isso,

evitou-se o constrangimento do candidato ao constatar que o examinador toma notas sobre sua performance ao mesmo tempo em que conversa.

Em síntese, em vez de priorizar-se o tratamento quantitativo dos erros cometidos, os examinadores buscaram encontrar o perfil que caracterizasse a performance do candidato. Em vez de quantificar erros, em detrimento da verificação das habilidades linguísticas dos sujeitos examinados, o foco de atenção deslocou-se para os “acertos”, ou para a competência comunicativa demonstrada em determinadas funções linguísticas. Com isso, o formulário de avaliação também mudou, e o resultado disso foi uma interação mais próxima do que seria uma conversa natural.

#### COLETA DE DADOS: ENVIO DE QUESTIONÁRIOS

Foram enviados questionários ao total de 155 candidatos inscritos para o exame deste ano. Destes, 54 foram preenchidos e devolvidos. Vários candidatos deixaram de responder, pois, apesar de inscritos, explicaram que não puderam comparecer para a realização do exame. Outro fator que contribuiu para reduzir a quantidade de questionários respondidos se deve à inoperância de alguns endereços eletrônicos informados.

Comando Militar	Cidade	Candidatos
CML (Leste)	Rio de Janeiro-RJ	76
CMSE (Sudeste)	São Paulo-SP	14
CMS (Sul)	Porto Alegre-RS	8
CMNE (Nordeste)	Recife-PE	12
CMO (Oeste)	Campo Grande-MS	11
COTER (Brasília)	Brasília-DF	25
CMA (Amazônia)	Manaus-AM	9
Total		155

## TABULAÇÃO DOS QUESTIONÁRIOS: PRINCIPAIS RESULTADOS

O questionário compunha-se de 3 tipos básicos de perguntas: (a) as que pretendiam verificar se o candidato havia se preparado para o exame; (b) as que pediam sua opinião sobre procedimentos e (c) as que solicitavam mais comentários.

Quanto à preparação para a realização do exame oral, 6 afirmaram não ter estudado, 17 consideraram que estudaram pouco, e 31 declararam ter estudado o suficiente. Com relação ao comportamento da banca, 35 entrevistados declararam que houve efeito positivo sobre seu desempenho; 14 declararam que o comportamento da banca não surtiu efeito sobre seu desempenho; 5 entrevistados declararam que o comportamento da banca teve efeito negativo sobre seu desempenho.

Do total de entrevistados, 49 consideraram que o EPO deste ano se assemelhou a uma conversa natural. Apenas 5 entrevistados responderam negativamente a essa pergunta.

Dos entrevistados, 44 realizaram o exame pela primeira vez; 9 realizaram pela segunda vez e 1 realizou pela terceira vez. Nesse universo dos que já haviam realizado o exame, 8 examinados declararam que os procedimentos adotados na condução do exame foram mais adequados se comparados aos de anos anteriores. Dentre os motivos apresentados para essa melhoria, foram apontados os seguintes:

- (1) A utilização do *icebreaker*, conversar inicial com a finalidade de ambientar o candidato a respeito das fases do exame, bem como para tentar tranquilizá-lo. Esse aquecimento foi apontado como um fator que faz “brotar” a fluência do candidato para que o exame transcorra com mais tranquilidade.
- (2) O desenvolvimento de um diálogo mais natural em comparação com os procedimentos adotados em anos anteriores, em que o entrevistador fez perguntas “fechadas” e

demonstrou uma atitude mais “seca”, além de não ter buscado relacionar os tópicos da conversa.

- (3) O uso de situações reais para uso da língua dentro de um contexto. Ao se discutir o tema, foram propostas atividades de *role-play*, em que o candidato deveria interagir com o entrevistador de acordo com um papel que lhe foi atribuído na conversa.
- (4) A atitude positiva do entrevistador, que deixou o candidato à vontade e possibilitou mais desenvoltura na interação.

Apenas 1 dos entrevistados atribuiu evolução positiva do processo à melhora das condições técnicas dos equipamentos. O nono entrevistado não observou diferenças nas rotinas adotadas e considerou os procedimentos semelhantes à oportunidade anterior.

## CONSIDERAÇÕES FINAIS

O EPO representa um esforço considerável com o intuito de reduzirem-se distâncias tão grandes num país de dimensões continentais como o Brasil. Consiste numa iniciativa sem precedentes de valorização do conhecimento e domínio de um idioma estrangeiro pelos profissionais militares do Exército Brasileiro, com repercussão internacional.

A valorização do conhecimento de idiomas estrangeiros iniciou-se há mais de uma década, e o reflexo dessas medidas é visível ao constatarmos a melhor preparação e domínio dos idiomas por parte do jovem oficialato, indo ao encontro das diretrizes emanadas pelos escalões superiores que consideram o domínio de uma língua estrangeira uma ferramenta essencial para o exercício profissional militar. A importância que é atribuída às línguas estrangeiras na carreira militar deve encontrar respaldo nos procedimentos e rotinas de avaliação.

Os resultados desta pesquisa contribuem para a validação de processos e para a melhoria continuada que os procedimentos de avaliação ensejam. Ao conduzir exames mais próximos de uma conversa natural e munidos de critérios claros e precisos, os avaliadores encontram situações mais favoráveis para o processo de avaliação. Em consequência, podem ser mais justos e obter resultados mais confiáveis.

## REFERÊNCIAS BIBLIOGRÁFICAS

CARVALHO, K. R. D., FARIAS, G. O., TOSTES, S. C. Exames de Proficiência Oral a Distância. In: GOMES, J. C. & SCHAFFEL, Sarita Léa. (Orgs.). *Formação docente: diferentes percursos*. Rio de Janeiro: Centro de Estudos de Pessoal, 2007, p. 119-131.

CHOMSKY, N. *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press, 1965.

\_\_\_\_\_. *Rules and Representations*. Oxford, Basil Blackwell, 1980.

COUNCIL OF EUROPE. Common European Framework of Reference for Languages.

HALLIDAY, M. A. K. The form of a functional grammar. In G. Kress (Ed.), Halliday: *System and function in language*. Oxford, OUP, 1976.

HYMES, D. On communicative competence. In J. B. Pride and J. Holmes (Eds.) *Sociolinguistics*. Harmondsworth, Penguin, 1972, pp. 269-293.

LABOV, W. *Language in the Inner City. Studies in the Black English Vernacular*. New York, 1972.

LUOMA, S. *Assessing Speaking*. New York, Cambridge, 2004.

MACNAMARA, T. *Language Testing*. New York, Oxford, 2000.