

Reflexões sobre avaliação em ensino/aprendizagem de línguas, com ênfase em educação a distância

Reflections on evaluation language teaching in with emphasis on distance learning

Prof. Me. Thereza Cristina de Souza Lima¹

¹Mestra pela Universidade Estadual Paulista (Unesp), doutoranda em Estudos da Tradução. Professora de Linguística de *Corpus* no curso de Pós-graduação da Pontifícia Universidade Católica do Paraná (PUCPR) intitulado: *Língua Inglesa: Metodologia do Ensino e Tradução*. tcslcristina@gmail.com.

RESUMO

Esta pesquisa trata da questão da avaliação no ensino de idiomas, concentrando-se principalmente no ensino/aprendizagem a distância. Apresenta algumas reflexões sobre avaliação em geral, e enfoca onze itens para avaliação a distância. Tais itens variam desde respostas totalmente restritas, as quais, considerando o potencial das novas tecnologias de informação, podem ser respostas limitadas demais, a respostas totalmente construídas, as quais constituem um desafio em relação à análise, até mesmo para os computadores atuais mais sofisticados.

Palavras-Chave: Avaliação. Ensino-aprendizagem a distância. Respostas limitadas. Respostas construídas. Análise computacional.

ABSTRACT

This paper approaches the issue of assessment in language teaching, concentrating mainly on distance learning. It presents some reflection upon assessing in general, and focuses on eleven items for distance learning assessment. Such items vary from fully “constrained” responses, which may be too limiting considering the potential of new information technologies, to fully “constructed responses”, which constitute a challenge even for today’s sophisticated computers to analyze.

Keywords: Assessment. Distance learning. Constrained responses. Constructed responses. Computer analysis.

1 Introdução

No presente artigo, pretendemos investigar questões relacionadas à avaliação, mais especificamente, à avaliação no ensino e aprendizagem de línguas. É necessário lembrar que, na língua inglesa, há dois termos que podem ser confundidos quando falamos de avaliação. São eles: *evaluation* e *assessment*. Tais termos, certamente, já foram utilizados como sinônimos, mas em relação a ensino e aprendizagem de idiomas, é importante apresentar a diferença entre eles.

Na visão de Garrison e Anderson (2003, p. 93), o termo *evaluation* refere-se à comparação de uma unidade, um curso ou um programa com algum critério de resultados, podendo relacionar-se com níveis de satisfação, preços, custo-benefício etc. Já o termo *assessment*, em educação formal, é usado para se referir ao papel crítico de avaliar alunos em relação aos objetivos educacionais propostos. Pode incluir, por exemplo, aquisição de habilidades e competências comportamentais, competência para aplicar habilidades cognitivas, incluindo a capacidade de aplicar soluções críticas e criativas a problemas complexos etc. *Assessment* geralmente fornece aos alunos e ao professor certo retorno, ou seja, certa informação, principalmente a respeito do que fora ensinado e aprendido durante um curso. Consequentemente, quando se trata da avaliação no ensino e aprendizagem, fala-se de *assessment* e não de *evaluation*.

Etimologicamente, o vocábulo *assessment* vem do latim *assidere* e significa "sentar-se ao lado", possibilitando a observação, bem como a coleta de informações. Este é um dos papéis do professor ao avaliar, que é adquirir informações, cujos propósitos, segundo Shepard (2000), dividem-se em três grandes categorias: a avaliação administrativa, que inclui certificação e promoção; a avaliação instrucional, que inclui nivelamento, avaliação de currículo; e a avaliação relacionada à pesquisa, que inclui experiências, conhecimento sobre aprendizagem e sobre o uso da língua.

2 Princípios em avaliação

O processo de conduzir uma avaliação, qualquer que seja seu propósito ou instrumento de avaliação, inclui diversas fases. A primeira fase diz respeito à

definição dos objetivos da avaliação, bem como à parte do conhecimento da língua a ser avaliado e à seleção de procedimentos avaliativos adequados para a execução dessa avaliação. Uma vez que tais passos tenham sido dados, devemos partir para a segunda fase, que é a produção do instrumento de avaliação em si e a sua aplicação, sendo importante a observação de quaisquer dificuldades que possam ocorrer antes, durante e após essa aplicação. Os passos seguintes relacionam-se com o processo de avaliar a qualidade do instrumento avaliativo, atentando-se à sua validade e confiabilidade. Na última fase, os resultados são interpretados e reportados aos interessados.

No que tange à decisão de qual instrumento de avaliação utilizar, Shohamy e Inbar (2006) afirmam que "depende do propósito da avaliação e de como o conhecimento da língua é definido"¹. É relevante ressaltar a grande discordância entre pesquisadores em relação ao que seja conhecimento da língua, mas concordamos com a maioria dos estudiosos da área, segundo os quais esse conhecimento consiste de muitas variáveis; logo, faz-se necessário o desenvolvimento de múltiplos procedimentos, ou seja, "o uso de multiplicidade" (Shohamy; Inbar, 2006). Desse modo, os referidos pesquisadores sugerem que a avaliação seja efetuada por meio de vários instrumentos, levando-se em conta as diversas opções disponíveis para cada fase do processo de avaliação.

A noção de multiplicidade engloba uma visão ampla de conhecimento de língua e de avaliação e se relaciona à multiplicidade em várias áreas:

propósitos múltiplos de avaliação de língua, definições múltiplas de conhecimento de língua, procedimentos múltiplos para medir esse conhecimento, critérios múltiplos para determinar o que é uma boa língua, e maneiras múltiplas de interpretar e reportar os resultados da avaliação (SHOHAMY, 1998, p. 242, apud SHOHAMY; INBAR, 2006, p. 5)².

Propósitos múltiplos de avaliação referem-se a diferentes motivos para se efetuar uma avaliação. Yorke (2005, p. 12) sugere três grandes grupos de razões ou propósitos para o uso de avaliações: promover aprendizagem, certificar realizações e fornecer dados que podem ser utilizados segundo diferentes objetivos.

¹ (...) depends on the purpose of the assessment and on how language knowledge is defined.

² (...) multiple purposes of language assessment, multiple definitions of language knowledge, multiple procedures for measuring that knowledge, multiple criteria for determining what language is, and multiple ways of interpreting and reporting assessment results.

No que tange a definições múltiplas de conhecimento da língua, já fora mencionado que se trata de um assunto polêmico devido à grande discordância entre pesquisadores. Todavia, esse conhecimento depende do propósito estipulado para a avaliação e é muitas vezes relacionado à variedade de tarefas que os alunos conseguem realizar na língua estudada, o que os leva a serem classificados como usuários iniciantes, intermediários ou avançados dessa língua.

No que concerne a procedimentos múltiplos, há uma grande diversidade de instrumentos avaliativos que podem ser usados, além dos tradicionais testes ou provas. Por exemplo, portfólios, debates orais, projetos, apresentações, autoavaliação, entre outros.

No que diz respeito a múltiplos modos de interpretar resultados, devemos considerar o propósito e a situação da avaliação. Por exemplo, para avaliar um imigrante, um dos fatores que devemos levar em conta é sua quantidade de convivência e contato com falantes do país de língua estrangeira, o tipo de programa de aprendizagem do idioma, bem como sua real vontade de aprender esse idioma. Já as múltiplas maneiras de se reportar resultados variam de acordo com o propósito e com o uso futuro que se fará deles, pois tudo depende de quem vai receber esses resultados. Se forem os pais ou os responsáveis pelo avaliando, os resultados podem até mesmo ser reportados verbalmente, de uma maneira informal. Se o receptor dos resultados for uma empresa ou instituição, como, por exemplo, a Capes, que recebe resultado de teste de proficiência para a concessão de bolsa de estudo, o resultado é reportado por escrito, de um modo mais formal.

Critérios múltiplos para determinar o que é uma “boa língua”, isto é, para determinar qualidade, dependem do propósito do teste, do tipo de conhecimento de língua e da habilidade almejados e das tarefas ou itens a serem realizados. Já os critérios múltiplos para determinar a qualidade dos procedimentos de avaliação envolvem examinar a confiabilidade e a validade dos instrumentos utilizados.

A confiabilidade refere-se à consistência do resultado da avaliação, ou seja, indica se os resultados são precisos, confiáveis e verdadeiros, e leva em conta qualquer erro que possa ocorrer no processo de avaliação.

Por sua vez, validade significa ter valor, e, em ensino e aprendizagem, um instrumento de avaliação tem validade quando reflete, ou seja, vai de encontro aos objetivos propostos pelo avaliador, os quais se coadunam com o que foi ensinado e

aprendido antes da avaliação. Shohamy e Inbar (2006, p. 8) distinguem alguns tipos de validade, que são a validade de conteúdo, de concorrente, de previsão, de face e de construto. A validade de conteúdo examina até que ponto a avaliação representa o conteúdo a ser avaliado; a validade de concorrente examina se duas avaliações que avaliem o mesmo conhecimento produzem informações semelhantes; a validade de previsão examina se um avaliando que teve sucesso em uma avaliação será realmente capaz de usar, em sua vida futura, o conhecimento sobre o que foi avaliado; a validade de face examina se há uma concordância entre o *layout* da avaliação e o que ela se propõe avaliar; e a validade de construto examina se a avaliação segue as teorias relacionadas ao que está sendo avaliado. Atualmente, as noções de validade de construto expandiram-se, englobando questões ligadas às consequências de avaliações, mais especificamente, ao impacto social e educacional que as avaliações podem causar aos alunos. Alguns pesquisadores, porém, questionam se tais questões devem ser vistas como tipos separados de validade ou como parte integral do construto de validade. Contudo, considera-se pertinente examiná-las mais detalhadamente.

3 Avaliação e impacto social e educacional

Até poucos anos atrás, ensinar significava apenas transmitir conteúdo. O professor atuava como o "dono do saber", aquele que passava informações, muitas vezes inquestionáveis, ao aluno, considerado mero receptor. Com base nesse modelo, as avaliações simplesmente examinavam até que ponto o conteúdo tinha sido assimilado, ou mesmo memorizado pelo receptor/aluno. Fidalgo (2006, p. 19) as chama de "avaliação bancária", ou seja, aquelas que consistem na

capacidade do aluno de buscar nos seus arquivos mentais os depósitos ali deixados, exatamente como foram feitos, sem interpretações, sem acréscimos, sem qualquer tipo de juros ou deduções, e devolvê-los ao depositante, mediante requisição: a prova, o teste, o exame final.

Nesse contexto, certamente, o professor/depositante, "dono do saber", era também o dono do poder, o que o tornava apto a exercê-lo com atitudes que podiam causar graves danos sociais e educacionais nos alunos. Por exemplo, o

aluno que, por algum motivo, nunca era levado em consideração e não conseguia memorizar as supostas respostas esperadas e escrevê-las em "avaliações", era tido como fracassado, rótulo que, muitas vezes, o acompanhava ao longo da vida, ocasionando-lhe profundo dano social e educacional. Tal rótulo também poderia causar ao aluno parcial ou total exclusão, uma vez que fracassados não são aceitos na sociedade, vivem à margem dela, ou seja, tornam-se marginais, excluídos, para os quais o único destino é a convivência com outros excluídos, que os aceitam e os amparam.

Com a inclusão do aluno no processo pedagógico, não falamos mais de ensino, mas de ensino/aprendizagem, em que o aluno não é mais um mero receptor, mas um agente ativo. Como agente ativo desse processo, o aluno passou a ter o direito de ser avaliado mais democraticamente. Surgiram, então, outros termos relacionados à avaliação, tais como "avaliação qualitativa", que segundo os Parâmetros Curriculares Nacionais (BRASIL, 1998, p. 81), é entendida "como um conjunto de atuações que têm a função de alimentar, sustentar e orientar a intervenção pedagógica" e não de julgar sucessos ou insucessos. Por outro lado, a avaliação formativa tem por objetivo a regulação e se interessa pelos procedimentos que levam o aluno a aprender. Opõe-se, assim, à avaliação somativa, que trata da questão do controle e se opõe também à avaliação normativa, comprometida com a hierarquização de alunos por seus resultados e muitas vezes por seu nível social.

Atualmente, o avanço da tecnologia, o crescimento da internet e o acesso cada vez maior aos recursos computacionais têm possibilitado uma nova perspectiva em relação à educação, que hoje pode ser realizada de maneira presencial ou a distância. Tais avanços têm causado grande revolução e impacto em todas as áreas do saber, especialmente no ensino/aprendizagem de idiomas, inclusive em relação à avaliação realizada no ensino/aprendizagem de idiomas a distância.

4 Avaliação no ensino/aprendizagem de línguas a distância

Conforme mencionado anteriormente, o avanço da tecnologia tem trazido considerável enriquecimento educacional em relação à avaliação, pois possibilita o

uso de estímulos visuais, sonoros e interativos, entre outros, expandindo as possibilidades avaliatórias para muito além da limitação imposta pela utilização tradicional de “papel e lápis”. Assim, por meio do computador, em uma abordagem presencial ou mais frequentemente a distância, torna-se possível elaborar avaliações de alta qualidade, que unam as atividades instrucionais e os objetivos propostos em um espaço de tempo relativamente curto. Para tanto, acreditamos que seja necessária a preparação de questões e tarefas, com as quais os usuários dos computadores possam efetivamente interagir, visando à obtenção de evidências significativas concernentes à medida do conhecimento, bem como ao fornecimento de resultados fidedignos.

Segundo Scalise e Gifford (2006), o tipo de questão mais frequente em avaliações a distância são as de múltipla escolha, geralmente compostas por uma pergunta, direta ou indireta, seguida por um pequeno conjunto de respostas, das quais se espera que o aluno selecione a melhor opção de resposta. Certamente, tal questão pode ser facilmente avaliada, porém, indubitavelmente, muito do potencial proporcionado pela utilização do computador “parece ser sacrificado” (SCALISE; GIFFORD, 2006, p. 5).

Os mencionados pesquisadores sugerem, então, a introdução de vinte e oito tipos de questões que podem ser úteis em avaliações efetuadas por meio do computador, entre os quais selecionamos onze tipos, por serem considerados mais adequados para o ensino/aprendizagem de línguas a distância. É importante ressaltar que, de acordo com os autores, tais tipos de questões encontram-se entre dois extremos: de um lado, questões cujas respostas são totalmente limitadas, e, de outro lado, aquelas cujas respostas são inteiramente “construídas”³, e que, por isso, apresentam um grande desafio para análise computadorizada, apesar de todas as ferramentas computacionais sofisticadas atualmente disponíveis. Um exemplo ilustrativo de questões cujas respostas podem ser consideradas totalmente limitadas são as questões de múltipla escolha, enquanto que as questões que envolvem respostas em forma de redação podem ser consideradas mais “construídas”.

³ (...) *constrained responses, which may be far too limiting (i.e., the conventional multiple-choice questions) (...) and constructed responses (i. e., the traditional essay).*

Baseando-se nessas duas grandes divisões, Scalise e Gifford (2006, p. 8) desenvolveram questões e tarefas para avaliação a distância com limitação de nível intermediário e as classificam como: múltipla escolha, seleção/identificação, reordenação/reorganização, substituição/correção, preenchimento, construção e apresentação/portfólio. É de grande valia observar que todos constituem tipos de itens de avaliação⁴ e relacionam-se a um estímulo e a uma ou a várias respostas provenientes desse estímulo.

4.1 Múltipla escolha

Podem ser constituídos por vários tipos de itens, tais como itens que apresentam apenas duas opções de respostas - Verdadeiro ou Falso - sendo, por isso, uma das mais simples e mais limitadas formas de múltipla escolha:

Item 1A: Verdadeiro/Falso (*True/False*)

The United Kingdom of Great Britain is formed of England, Scotland, Wales and Northern Ireland.

True

False

Ainda, apresentando apenas duas opções de resposta, semelhantemente ao verdadeiro ou falso, estão os itens com escolha alternada, em que o avaliado escolhe a melhor opção.

Item 1B: Escolha Alternada (*Alternate Choice*)

Betty says: "When I need to go shopping for clothes, I usually go to the nearest mall".

Student A says that this sentence was probably spoken by an American girl.

Student B says that this sentence was probably spoken by an English girl.

⁴ Um item de avaliação pode se constituir de qualquer interação com o avaliado, cujos dados são coletados com a intenção de se realizar alguma inferência sobre esse avaliado (SCALISE; GIFFORD, 2006, p. 7).

With which student do you most agree, taking into account the use of the word “mall”?

Alguns autores criticam esses dois tipos de itens citados por possibilitarem alta incidência de adivinhações. Entretanto, é possível lidar com tal problema levando em conta ajustes estatísticos.

Item 1C: *Múltipla Escolha Convencional (Conventional Multiple Choice)*

Which countries form the United Kingdom of Great Britain?

- A) *England and Scotland*
- B) *England, Scotland and Wales*
- C) *Scotland and Wales*
- D) *England, Scotland, Wales and Southern Ireland*
- E) *England, Scotland Wales and Northern Ireland*

Conforme mencionamos anteriormente, embora as questões de múltipla escolha convencionais sejam facilmente corrigidas pelo computador e sejam também menos suscetíveis à adivinhação, diferentemente das questões cujas respostas limitem-se a duas opções, há críticas a seu uso exclusivo por vários motivos, entre os quais ressalta-se a limitação imposta à abordagem computadorizada, impedindo uma possível otimização do potencial avaliativo existente em tecnologia. Inovações em itens de múltipla escolha em contextos virtuais podem incluir ações em resposta a questões incomuns em contextos não virtuais, como, por exemplo, clicar em uma área adequada para gerar determinada imagem.

4.2 Seleção/identificação

Nas palavras de Scalise e Gifford (2006, p. 15), “quando as opções de múltipla escolha se tornam suficientemente ricas de modo a limitar as chances de adivinhação, podemos classificá-las como abordagens de seleção ou identificação”.

Item 2A: Múltiplos Verdadeiros/Falsos (*Multiple True/False*)

Which of the following sentences would be more likely to appear in a newspaper due to the fact that it has a verb in the passive voice?

	Yes
No	
A) <i>After the accident, the injured child was taken to hospital.</i>	()
()	
B) <i>The presidents' meeting will be held in New York.</i>	()
()	
C) <i>The hurricane has destroyed a great part of the town.</i>	()
()	
D) <i>The chairman of Microsoft, Bill Gates, has been kidnapped.</i>	() ()
()	
E) <i>The book "The Kite Hunter" has been in the best seller list for 2 years.</i>	()
()	

No exemplo acima, a probabilidade de adivinhações é reduzida, uma vez que, para se escolher as alternativas corretas, é necessário que o avaliado realmente entenda como é formada a voz passiva, relacionando-a com a linguagem jornalística. Logo, as opções A, B e D deveriam ser selecionadas como as mais adequadas.

Item 2B: Seleção/identificação com justificativa ou explicação (*Yes/No with explanation*)

Read the excerpt and answer the following question: "Is the paragraph below about economy"?

Like a Hollywood monster that is impervious to bullets, the credit crisis refuses to lie down and die. The authorities have bombarded it with interest-rate reductions, tax cuts, special liquidity schemes and bank bail-outs, but the creature lumbers forward, threatening new victims with every step.

A) () No, because the paragraph is about a Hollywood monster.

B) () Yes, because the paragraph is about reduction in inflation.

- C) () *No, because the paragraph is about how the authorities are dealing with the problems caused by a film in which there is a Hollywood monster.*
- D) () *Yes, because the paragraph is composed of a comparison between a monster and the present crisis in economy.*

Scalise e Gifford (2006, p. 16) concordam com McDonald (2002), segundo o qual tal tipo de item, quando bem formulado, pode constituir uma forma de avaliação mais efetiva do que questões convencionais de múltipla escolha.

4.3 Reordenação/reorganização

Nos tipos de itens que envolvem reordenação/reorganização, solicita-se que o avaliado relacione ou categorize as opções em uma ordem ou sequência previamente estipulada. Segundo Haladyna (1994c), apud Scalise e Gifford (2006, p. 19), esses tipos de itens são frequentemente usados em classe, mas raramente usados em avaliações, o que pode ser uma lástima, pois “são fáceis de construir e administrar; são úteis em avaliar associações, definições e exemplos; e são eficientes em relação a espaço e a tempo (Idem)”.

Item 3A: Relação entre itens (*Matching*)

Questions beginning with “*How*” are very common in English. The sentences on the left are answers given to some “*How*” questions.

Match these answers with the corresponding questions:

- | | |
|----------------------------------|--|
| () <i>About a mile.</i> | 1. <i>How deep is the swimming pool?</i> |
| () <i>He is 25.</i> | 2. <i>How long have you been waiting?</i> |
| () <i>For 10 or 15 minutes.</i> | 3. <i>How far is the school from here?</i> |
| () <i>I just love it</i> | 4. <i>How do you like the new house?</i> |
| | 5. <i>How old is your brother?</i> |

Como podemos observar no exercício, o número de opções do lado direito pode ser maior que as opções do lado esquerdo, para que a chance de adivinhações seja reduzida.

4.4 Substituição/correção

Questões que envolvem substituição e correção requerem do avaliado um grau adicional de construção, pois a resposta correta precisa, ou não, ser modificada em cada local identificado, como no exemplo a seguir.

Item 4A: Dedo Machucado (*Sore Finger*)

In the sentence below, there are several underlined words. Mark the INCORRECT one:

(A)

(B)

There was a terrific rock concert at the club yesterday and Mary loved it, but

(C)

(D)

(E)

Jane, Mary's best friend didn't go. If she went, she would have enjoyed herself very much.

Apesar da existência de questionamentos sobre o foco da resposta envolver uma questão incorreta, tal tipo de item ainda é bastante utilizado em inúmeras avaliações relacionadas ao ensino/aprendizagem de idiomas, como, por exemplo, nos exames internacionais oferecidos pela Universidade de Cambridge.

4.5 Preenchimento

Diferentemente das anteriormente mencionadas, nessa categoria nem todos os componentes são fornecidos antecipadamente. Aqui, muitas vezes, o avaliado deve terminar um estímulo incompleto, como, por exemplo, fornecer respostas por meio de um ou de poucos vocábulos, como em *fill in the blanks format*, embora saibamos que essa terminologia possa ser evitada por alguns autores. Há também certa crítica em relação a preenchimento, pois alguns pesquisadores acreditam que, nesse caso, apenas a capacidade de memória esteja sendo testada, uma vez que o avaliado pode completar apenas com um ou poucos vocábulos. Entretanto, segundo Scalise e Gifford (2006, p. 28), o computador pode permitir uma variedade

maior de respostas devido a sua extensa capacidade de armazenamento de dados, o que, conseqüentemente, enriqueceria esse tipo de item.

Item 5A: Respostas curtas & Preenchimento de sentenças (*Short-Answer & Sentence Completion*)

A place where university students usually borrow books from is called a:

Osterlind (1998), apud Scalise e Gifford (2006, p. 28), acredita que os itens que envolvem preenchimento também possibilitam certa redução em relação à tentativa de adivinhação, embora afirme também que são poucas as pesquisas sobre tal assunto.

Item 5B: Preenchimentos de lacunas com múltipla escolha (Cloze Procedures)

Esse tipo de item, que nas palavras de Osterlind é “certamente um dos mais pesquisados” (1998, apud SCALISE; GIFFORD, 2006, p. 28), avalia frequentemente a habilidade de leitura por meio da retirada de vocábulos de um texto, de modo que o avaliando o complete escolhendo uma das opções apresentadas em múltipla escolha. Ou seja, vocábulos são extraídos de um texto e a tarefa do avaliando é recolocá-los, escolhendo-os entre alternativas fornecidas. Embora não existam regras estabelecidas sobre qual vocábulo deve ser retirado, uma abordagem típica é a retirada do quinto vocábulo de uma sentença, quando se avalia habilidades de compreensão de texto. Nesse caso, o vocábulo retirado constitui-se de um vocábulo comum, cujo significado pode ser facilmente identificado. Quando não se trata de avaliação de compreensão de texto, palavras-chave são retiradas. Em ambos os casos, porém, apenas uma das alternativas pode completar o texto adequadamente, como no exemplo a seguir:

Last July, John McCain addressed the National Association for the Advancement of Coloured People (NAACP). The audience was nearly all black and mostly left-of-centre. Few of them would vote for a Republican in any circumstances, let alone a Republican running against the man who could be America's first black president.

“Let me begin, if I may, with a few words about my (1) _____,” Mr McCain said. “Don’t tell Obama I said this, but his success has made America very proud of him”. Then, he (2) _____ explained why he disagrees with Obama on certain (3) _____, such as school choice.

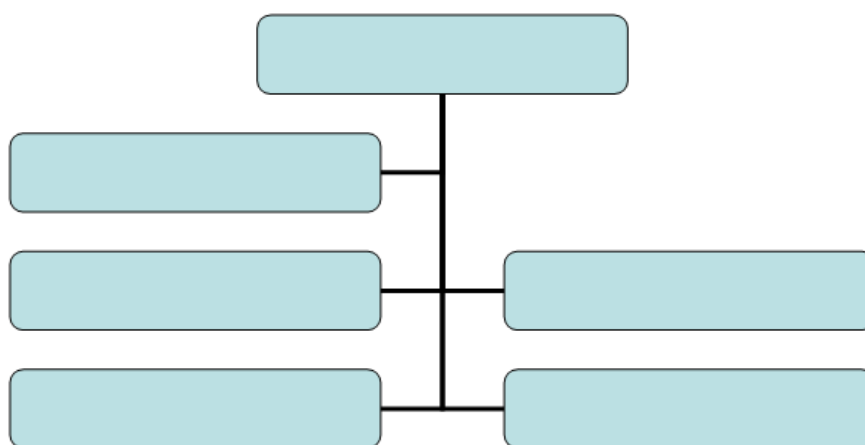
1. A) colleague; B) friend; C) enemy; D) opponent; E) relative;
2. A) obediently; B) disobediently; C) respectfully; D) truly; E) faithfully;
3. A) problems; B) matter; C) trouble; D) issues; E) causes;

No exemplo, o tema está bastante claro no primeiro parágrafo, o que possibilita ao avaliado identificar detalhes que o ajudarão a escolher a opção mais adequada, como, por exemplo, em relação ao registro do texto, formal ou informal. É de grande valia lembrar que todas as opções estão gramaticalmente corretas, de modo que o avaliando tenha sempre que levar em conta o contexto. Segundo Osterlind, o *College Entrance Examination Board Handbook* (1986) reconhece a validade desse tipo de item, pois “a escolha de respostas incorretas mostra falha na compreensão” (OSTERLIND 1998, apud SCALISE; GIFFORD, 2006, p. 29).

4.6 Construção

À medida que os itens dependem mais de construção, a atribuição de resultados torna-se, cada vez mais, um maior desafio. Scalise e Gifford (2006) apresentam vários exemplos de itens que podem ser baseados em construção. Muitos deles, porém, parecem ser mais apropriados para outras disciplinas, como, por exemplo, aquelas relacionadas às ciências exatas. Seleccionamos, então, como exemplo ilustrativo de construção, o *Concept Map*.⁵ Em tal tipo de item, o avaliando deve construir um mapa de modo que os termos fornecidos sejam organizados em determinada configuração, isto é, a construção situa-se na criação de relações entre conceitos, como na figura que segue:

⁵ O exemplo de *concept map* acima foi retirado de Scalise e Gifford (2006, p. 35), mas o enunciado do item é de nossa criação.

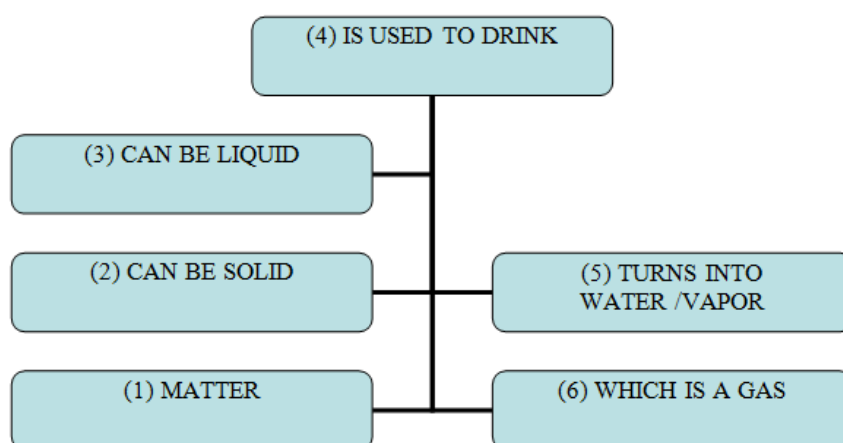
Figura 1 - Construção do *Concept Map*

Tal tipo de mapa pode ser preenchido pelo avaliado a partir dos vocábulos que deverão ser inseridos no mapa, como, por exemplo:

Item 6A: Insert the following words, or group of words in the map above, so that it becomes a true sentence. Add any two or three words which you think might be necessary:

Liquid - Gás - Solid - Water Vapor - Matter - Used to drink -

Figura 2 - Exemplo de preenchimento do mapa



Nas palavras de Scalise e Gifford (2006, p. 34), “o desafio aqui algumas vezes está na interpretação do significado das diferenças nos mapas dos alunos e em como essa informação pode ser usada para perceber as diferenças entre os

alunos e suas compreensões sobre um tópico”. O exemplo dado envolve a concepção do avaliado sobre *matter*/matéria. De acordo com a resposta, o avaliado mostra *matter*/matéria como algo que pode ter dois estados, *liquid*/líquido e *solid*/sólido. Embora o avaliado não reconheça a terceira fase, *gas*/gás, como relacionada ao conceito direto de *matter*/matéria, ele/ela reconhece que *liquid*/líquido pode se tornar *gas*/gás, que pode ser bebível e pode se tornar vapor.

É de relevância observar que tal tipo de item pode ser facilmente corrigido pelo computador, embora os avaliados possam construir diferentes mapas com diversos tipos de configuração.

Item 6B: Outro item de construção é o *essay*, ou seja, a composição, que é, certamente, “um dos tipos de respostas mais prevalentes, usadas em avaliação” (SCALISE; GIFFORD, 2006, p. 35).

No exemplo de Scalise e Gifford (2006), o avaliado deve escrever um *essay* a partir da leitura do seguinte enunciado:

*Phone Company A charges a \$5 monthly fee and 7 cents per minute.
Phone Company B charges 10 cents per minute with no monthly fee.
Describe when Company A would be a better deal than Company B*

As abordagens relacionadas à produção de *essays online* podem apresentar muitas variedades. A maioria das inovações relacionadas à *essays online* envolvem questões de avaliação, análise de critério, confiabilidade, modelos estatísticos, entre outras. No presente artigo, optamos pelas questões relacionadas à avaliação, baseando-nos, principalmente, em dois artigos, Scharber, Dexter e Riedel (2008) e Wang e Brown (2007).

O artigo de Scharber, Dexter e Riedel (2008) intitulado *Students' experiences with an automated essay scorer* menciona o uso de um *software* denominado ETIPS (*Educational Theory into Practice Software*)⁶, que fornece ferramentas para a correção de *essays* avaliados *online*, cujos resultados são eletronicamente disponíveis para o professor e para o avaliado juntamente com *feedback* formativo detalhado sobre o processo de avaliação e comentários

⁶ *Software* Educacional Teoria na Prática.

específicos para cada caso. É interessante notar que o referido *software* proporciona ao avaliado a possibilidade de correção e recepção de *feedback* também de rascunhos elaborados por ele a partir de determinado tópico. Esse material lhe oferece a possibilidade de prever seu resultado e re-escrever sua resposta final, adequando-a para que tenha melhor conteúdo e qualidade. Além disso, oferece a ambos, avaliador e avaliado, acesso a todas as produções submetidas, ou seja, aos rascunhos e à produção final, bem como ao conceito e aos comentários finais.

Em um outro estudo intitulado *Automated essay scoring versus human scoring*, Wang e Brown (2007) pesquisam a diferença entre a avaliação de *essays* efetuada por professores e a avaliação efetuada *online* por meio de um *software* denominado IntelliMetric⁷. Os resultados desse estudo parecem apontar para algumas diferenças entre a avaliação humana e a avaliação computacional, uma vez que o conceito final fornecido pelo *software* foi mais alto que o conceito final dado pelos professores. Esse resultado é diferente da investigação sobre o mesmo tema realizada por Nivens-Bower's (2002), segundo a qual não parece haver diferença significativa entre a avaliação realizada por humanos e a avaliação realizada por computador. Entretanto, segundo Wang e Brown (2007), tais diferenças podem existir devido ao fato de os participantes dos referidos estudos apresentarem diferenças linguísticas e culturais significativas. Ambos os grupos de investigadores recomendam, então, mais pesquisas sobre o tema.

Pode-se observar, contudo, que com o avanço da tecnologia e, conseqüentemente, do ensino/aprendizagem a distância, uma das grandes preocupações dos estudiosos da atualidade diz respeito à questão da avaliação de textos por meio de ferramentas computacionais.

4.7 Categorias menos limitadas

Dentro da categoria de itens que apresentam menos limitações estão os projetos, os portfólios, as apresentações, as demonstrações, as atividades de grupo, as discussões e as entrevistas. Nas palavras de Scalise e Gifford,

⁷ Assim como o ETIPS utilizado por Scharber, Dexter e Riedel (2008), o IntelliMetric oferece ao avaliado as mesmas possibilidades em relação a resultados.

é possível gerar resultados computadorizados para alguns aspectos de tarefas na categoria de apresentação; porém, há muitos desafios em relação à validade desses resultados e freqüentemente avaliação humana e/ou comparação humana fazem-se desejáveis (2006, p. 37).

Todavia, esses pesquisadores reconhecem duas possibilidades de inovações fornecidas pelo computador. A primeira refere-se à facilidade proporcionada pela tecnologia em relação ao preparo, à interatividade e ao enriquecimento que podem ser alcançados com a utilização de ferramentas computacionais adequadas.

A segunda refere-se ao fato de os computadores poderem criar novas oportunidades de atividades em grupo em diferentes locais, possibilitando, por exemplo, a avaliação entre colegas⁸.

É de grande valia ressaltar que Scalise e Gifford (2006, p. 37) sugerem que, caso a avaliação de itens de observação seja realizada por meio de ferramentas computacionais, todas as evidências referentes à validade e confiabilidade estejam sempre muito claras para o avaliado, que precisa entender as abordagens utilizadas para se chegar a determinado resultado, evitando, assim possíveis impactos que podem ser ocasionados por resultados avaliatórios.

5 Considerações finais

A educação tem passado por grandes mudanças, principalmente no que concerne a educação a distância, que, devido aos avanços tecnológicos, torna-se, cada vez mais, uma constante no mundo moderno. Conseqüentemente, essas mudanças têm sido frequentes também na avaliação, presencial e a distância.

No presente artigo, apresentou-se uma breve introdução sobre avaliação presencial e a distância e concentrou-se em onze itens que são considerados mais pertinentes para o ensino/aprendizagem de idiomas. Tais itens variam entre, em um extremo, questões cujas respostas são totalmente limitadas, em outro, questões que possibilitam que o avaliado construa suas respostas de uma maneira menos limitada, ou seja, mais livremente. Porém, independentemente do grau de limitação, todos os itens possibilitam avaliação por meio da utilização do

⁸ Para aprofundamento em avaliação entre colegas (*Peer Assessment*), sugere-se a leitura do artigo de Ariwa (2003) intitulado *Peer Assessment and Tutoring*, apresentado na Conferência Anual da Associação de Pesquisa Educacional Escocesa, realizada em Perth, entre 27 e 29 de novembro de 2003.

computador, sendo que os itens que envolvem observação, como as apresentações e os portfólios, por exemplo, são os que apresentam menor limitação em relação à resposta, mas, por outro lado, apresentam mais dificuldade em relação à avaliação com o uso de ferramentas computacionais.

No que tange à questão da validade dos onze itens aqui apresentados, observa-se que as opiniões de renomados pesquisadores da área variam, principalmente, a medida que os itens também variam. Por exemplo, em questões de múltipla escolha, em que há apenas uma única resposta correta, tal resposta frequentemente, segundo Bennett (1993, p. 24), “reside na cabeça do professor ou de quem elaborou o teste e a tarefa do aluno é descobrir a resposta por meio de adivinhação”, o que, certamente, causaria um questionamento sobre a validade do referido item. Todavia, é de relevância lembrar que a validade de construto é apenas um dos objetivos de uma avaliação. Outros objetivos a serem alcançados poderiam estar relacionados, por exemplo, a estímulos mais autênticos que, conseqüentemente, causariam respostas mais criativas e mais bem elaboradas pelos alunos.

Acredita-se que, com maiores avanços tecnológicos, sejam criados cada vez mais diferentes e adequados *softwares*, capazes de possibilitar maior confiabilidade e validade, especialmente em relação à avaliação a distância.

Como aluna, acredito que o ensino/aprendizagem do século XXI e XXII progredirá na direção da educação a distância; como educadora, acredito que a tendência será fazer uso cada vez mais efetivo da *web* e de seu potencial para a criação de melhores ambientes que favoreçam a educação a distância; como pesquisadora, considero promissoras as oportunidades proporcionadas pelo uso de novas e cada vez mais efetivas tecnologias.

Referências

ARIWA, E. I. Peer Assessment and Tutoring (Comparative analysis of e-Learning with respect to Business Information Technology). **Scottish Educational Research Association Annual Conference**. Perth, nov. 2003. p. 27-29.

BENNETT, R. E. On the Meaning of Constructed Response. In: BENNETT, R. E.; WARD, W. C. (Ed.). **Construction versus Choice in Cognitive Measurement: Issues in Constructed Response, Performance Testing, and Portfolio Assessment**. Hillsdale: Lawrence Erlbaum Associates, 1993.

BRASIL. **Secretaria de Ensino Fundamental. Parâmetros Curriculares Nacionais: terceiro e quarto ciclos do ensino fundamental, língua estrangeira**, 1998. 126 p.

CASSADY, J. C.; GRIDLEY, B. E. The effects of online formative and summative assessment on test anxiety and performance. **Journal of Technology, Learning and Assessment**, Boston, v. 4, n. 5, 2005.

CONSOLO, D. A. On a (re)definition of oral language proficiency for EFL teachers: Perspectives and contributions from current research. **Melbourne Papers in Language Testing**, n. 1, p. 1-28, 2006.

FIDALGO, S. S. **A avaliação na escola: um histórico de exclusão social-escolar ou uma proposta sociocultural para a inclusão?** São Paulo: Cortez, 2006.

GARRISON, D. R.; ANDERSON, T. **E-Learning in the 21st Century: A Framework for Research and Practice**. Oxon & New York: Routledge Falmer, 2003. p. 92-104.

MCDONALD, M. E. **Developing Multiple-Choice Items**. In *Systematic Assessment of Learning Outcomes*. Sadbury: Jones and Barlett Publishers, 2002

MANTZ, Y. **Issues in the assessment of practice-based professional learning**. [s.l.]: Open University, 2005.

NIVENS-BOWER, C. (2002). Faculty-Write Placer *Plus* score comparisons. In: **Vantage Learning, Establishing WritePlacer Validity: A summary of studies**. Upper Saddle River: Prentice Hall, 2004.

PAIVA, V. L. M. O.; SADE, L. **Avaliação, cognição e poder**. São Paulo: Cortez, 2006.

SCALISE, K. & GIFFORD, L. Computer-Based Assessment in E-Learning. **Journal of Technology, Learning and Assessment**. Boston, v. 4 n. 6

SHEPARD, L. A. The Role of Assessment in a Learning Culture. **Educational Researcher**, London, 2000.

SHOHAMY, E.; INBAR, O. **The Language Assessment Process: A “Multiplism” Perspective**. University Park, PA: The Pennsylvania State University, Center for Advanced Language Proficiency Education and Research, 2006.

SCHARBER, C., DEXTER, S.; RIEDEL E. Students’ Experiences with an Automated Essay Scorer. **Journal of Technology, Learning, and Assessment**, Boston, v. 7, n. 1, 2008.

WANG, J.; BROWN, M. S. Essay Scoring Versus Human Scoring: A Comparative Study. **Journal of Technology, Learning, and Assessment**, Boston, v. 6, n. 2, 2007.

YORKE, M. **Issues in the assessment of practice-based professional learning**. Disponível em: <http://www.open.ac.uk/cetl/pbpl/links/MY> Último acesso em 05 de novembro de 2008.