

O PROCESSO AVALIATIVO EM EDUCAÇÃO A DISTÂNCIA

Thereza Cristina de Souza Lima (UNINTER)
tclslicristina@gmail.com

RESUMO

A presente pesquisa tem por objetivo investigar questões relacionadas ao processo avaliativos, mais especificamente, em relação à avaliação na educação a distância. Para tanto, vários fatores foram considerados, entre os quais se destacam a multiplicidade, a confiabilidade, a validade e o impacto social e educacional que pode ser ocasionado pelas avaliações. Kathleen Scalise & Bernard Gifford (2006) sugerem 28 tipos de questões que podem ser úteis em avaliações efetuadas por meio do computador, entre os quais foram selecionados onze tipos, por serem considerados mais adequados para o ensino/aprendizagem de idiomas a distância. Tais itens variam entre, em um extremo, questões cujas respostas são totalmente limitadas, em outro extremo, questões que possibilitam que o avaliando 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 computador. Concluiu-se que certamente, o ensino/aprendizagem do século XXI e XXII progredirá na direção da educação a distância e acredita-se, as oportunidades proporcionadas pelo uso de novas e, cada vez mais efetivas tecnologias venham contribuir, cada vez mais significativamente para o processo do ensino/aprendizagem de idiomas, principalmente no que diz respeito à educação a distância.

Palavras-chave: Avaliação. Ensino/aprendizagem de idiomas. Educação a distância.

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 vocábulos que podem ser confundidos quando falamos de avaliação. São eles *evaluation* e *assessment*. Tais vocábulos, certamente, já foram utilizados como sinônimos, mas em relação a ensino e aprendizagem de idiomas, é importante apresentar a diferença entre os dois termos.

Na visão dos pesquisadores Donn Randy Garrison e Terry Anderson (2003, p. 93), o termo *evaluation* refere-se à comparação de uma unidade, curso ou 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 referir-se 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, inclu-

indo 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 foi ensinado e aprendido durante um curso. Consequentemente, quando nos referirmos à avaliação no ensino e aprendizagem, estamos falando 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, que é um dos papéis do professor ao avaliar: adquirir informações cujos propósitos, segundo Elana Shohamy e Ofra Inbar (2006, *apud* 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.

Contudo, 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, deve-se partir para a segunda fase, que é a produção do instrumento de avaliação em si e 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 a 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, Elana Shohamy e Ofra Inbar (2006) afirmam que "depende do propósito da avaliação e de como o conhecimento da língua é definido"¹⁵⁶. É relevante ressaltar a grande discordância entre pesquisadores em relação ao que seja conhecimento da língua, mas, na visão da maioria dos estudiosos da área, 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 com o uso

¹⁵⁶ ... depends on the purpose of the assessment and on how language knowledge is defined. (tradução nossa)

de várias abordagens, 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 relaciona-se 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)¹⁵⁷.

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 possam ser utilizados segundo diferentes objetivos.

No que tange a definições múltiplas de conhecimento da língua, já se mencionou 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 etc.

No que diz respeito a múltiplos modos de interpretar resultados, devem-se considerar o propósito e a situação da avaliação. Por exemplo, para avaliar um imigrante, um dos fatores que se deve 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 o uso futuro que se fará desses resultados, os quais dependem de quem vai receber esses

¹⁵⁷ (...) *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.*

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 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 ao encontro dos objetivos propostos pelo avaliador, os quais se coadunam com o que foi ensinado antes da avaliação. Elana Shohamy e Ofra 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 atuais relacionadas ao que está sendo avaliado. Atualmente, as noções de validade de construto expandiram-se, englobando questões ligadas às consequências das avaliações, mais especificamente, ao impacto social e educacional que as avaliações podem causar aos alunos. Alguns pesquisadores, porém, questionam se tais questões devem se vistas como tipos separados de validade ou como parte integral do construto de validade. Contudo, consideramos pertinente examiná-las mais detalhadamente.

2. *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, mero receptor dessas informações. Com base nesse modelo, as avaliações simplesmente examinavam até que ponto o conteúdo havia sido assimilado, ou mesmo memorizado pelo receptor/aluno. Sueli Salles 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 juro ou deduções, e devolvê-los ao depositante, mediante requisição: a prova, o teste, o exame final". (FIDALGO, 2006, p. 19)

Nesse contexto, certamente, o professor/depositante, "dono do saber" era também o dono do poder, o que o tornava apto a exercer tal poder com atitudes que podiam causar graves danos sociais e educacionais nos alunos. Por exemplo, o aluno que, por algum motivo que nunca era levado em consideração, 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 amparam.

Com a inclusão do aluno no processo pedagógico, muda-se a terminologia da palavra ensino, para ensino/aprendizagem, em que o aluno não é mais um mero receptor, mas um agente ativo. Como agente ativo no processo ensino/aprendizagem, o aluno passou a ter o direito de ser avaliado mais democraticamente e surgiram outros termos relacionados à avaliação, tais como, avaliação qualitativa que, segundo os *Parâmetros Curriculares Nacionais* (PCN, 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 interessa-se 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.

3. Avaliação no ensino/aprendizagem de idiomas 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 avaliativas 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, acredita-se que seja necessária a preparação de questões e tarefas, com as quais possam-se obter evidências significativas concernentes à medida do conhecimento, bem como ao fornecimento de resultados fidedignos.

Segundo Kathleen Scalise e Bernard 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 28 tipos de questões que podem ser úteis em avaliações efetuadas por meio do computador, entre os quais foram selecionados onze tipos, por serem considerados mais adequados para o ensino/aprendizagem de idiomas a distância. É importante ressaltar que, de acordo com Kathleen Scalise e Bernard Gifford (*Idem*), 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”¹⁵⁸, e que, por isso, apresentam um grande desafio para análise computadorizada, apesar de todas as ferramentas computacionais sofisticadas de hoje. 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, ou seja, questões discursivas podem ser consideradas mais “construídas”.

Baseando-se nessas duas grandes divisões, Kathleen Scalise e Bernard 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 valia observar que todos constituem tipos de itens de avaliação¹⁵⁹ e relacionam-se a um estímulo e a uma, ou a várias respostas provenientes desse estímulo.

3.1. Múltipla escolha

Podem ser constituídos de 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:

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

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

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

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.

¹⁵⁸ ... constrained responses, which may be far too limiting (i.e., the conventional multiple-choice questions)... and constructed responses (i. e., the traditional essay)

¹⁵⁹ 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”?

Item 1 B: Escolha Alternada (Alternate Choice)

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

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*

Item 1C: Múltipla escolha convencional (Conventional Multiple Choice)

Conforme mencionado 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, dentre os quais ressalta-se a limitação imposta à abordagem computadorizada, o que impede 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.

3.2. Seleção/identificação

Nas palavras de Kathleen Scalise e Bernard 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”.

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?

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

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

No exemplo acima, a probabilidade de adivinhações é reduzida, uma vez que para se escolher as alternativas corretas, é necessário que o

avaliado realmente entendida 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.

Read the excerpt below and, based on it, select the correct answer for 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.

Item 2 B)

Seleção/identificação com justificativa ou explicação (Yes/No with explanation)

Kathleen Scalise e Bernard Gifford (2006, p. 16) concordam com Mary E. 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.

3.3. Reordenação/reorganização

Nos tipos de itens que envolvem reordenação/reorganização solicita-se que o avaliador relacione, ou categorize as opções em uma ordem ou sequência previamente estipulada. Thomas M. Haladyna (1994c) citado por Kathleen Scalise e Bernard Gifford (2006, p.19), afirma que 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 para avaliar associações, definições e exemplos; e são eficientes em relação a espaço e a tempo (*Idem*)".

Questions beginning with "How" are very common in English. The sentences on the left are answers given to some "How" questions. Match the questions to these answers:

() About a mile.

1. How deep is the swimming pool?

() He is 25.

2. How long have you been waiting?

() For 10 or 15 minutes.

3. How far is the school from here?

() I just love it

4. How do you like the new house?

5. How old is your brother?

Item 3 A) Relação entre itens (Matching)

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

3.4. Substituição/correção

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

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

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

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.

3.5. Preenchimento

Diferentemente das categorias acima, 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 Kathleen Scalise e Bernard 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.

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

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

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

Esse tipo de item, que nas palavras de Steven J. Osterlind é “certamente um dos mais pesquisados” (OSTERLIND 1998, *apud* SCALISE & GIFFORD 2006, p. 28), tem por objetivo avaliar 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 deva 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 abaixo:

In July, just before Obamas's election, 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 disagreed 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.

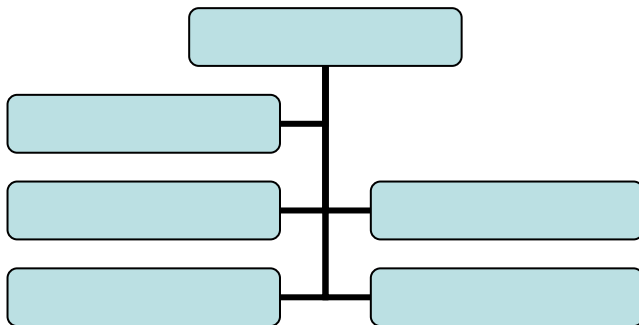
Item 5 B) Preenchimentos de lacunas com múltipla escolha (Cloze Procedures)

No exemplo acima, 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 valia lembrar que todas as opções estão gramaticalmente corretas, de modo que o avaliando tenha sempre que levar em conta o contexto.

Segundo Steven J. Osterlind, o *College Entrance Examination Board Handbook* 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)

3.6. Construção

À medida que os itens dependem mais de construção, a atribuição de resultados torna-se, cada vez mais, um maior desafio. Kathleen Scalise e Bernard 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. Foi selecionado, então, como exemplo ilustrativo de construção o *Concept Map*.¹⁶⁰ Em tal tipo de item, o avaliando deve construir um mapa conceitual 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 abaixo:

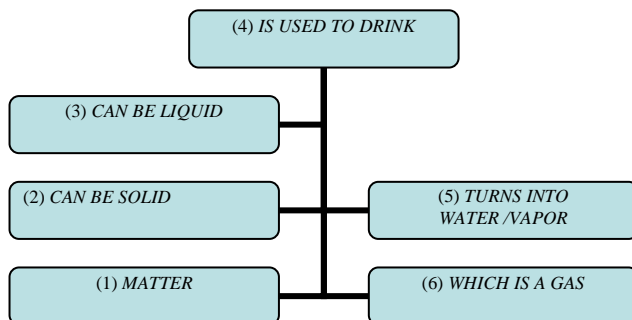


Tal tipo de mapa pode ser preenchido pelo avaliado, a partir dos vocábulos que deverão ser inseridos no mapa, como por exemplo:

Item 6 A) 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 – Gas – Solid – Water Vapor – Matter – Used to drink –

¹⁶⁰ O exemplo de *concept map* acima foi retirado de Kathleen Scalise e Bernard Gifford (2006, p. 35), mas o enunciado do item é de nossa criação.



Nas palavras de Kathleen Scalise e Bernard 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 acima, o avaliando mostra *matter*/matéria como algo que pode ter dois estados, *liquid/liquido* e *solid/sólido*. Embora o avaliando 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 e os avaliandos podem construir diferentes mapas com diversos tipos de configuração.

No exemplo de Kathleen Scalise e Bernard Gifford (*Idem*), 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 or the situation in which company A would be a better deal than company B*

Item 6 B) 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)

As abordagens relacionadas à produção de *essays on line* podem apresentar muitas variedades. A maioria das inovações relacionadas a *essays on line* envolvem questões de avaliação, análise de critério, confiabilidade, modelos estatísticos entre outras. No presente artigo, optou-se pelas questões relacionadas à avaliação, com base, principalmente, em

Cassandra Scharber, Sara Dexter e Eric Riedel (2008) e Jijun Wang e M. S. Brown (2007):

Em artigo de Cassandra Scharber, Sara Dexter e Eric Riedel (2008) intitulado *Students' Experiences with an Automated Essay Scorer* menciona-se o uso de um software denominado ETIPS (*Educational Theory into Practice Software*)¹⁶¹, que fornece ferramentas para a correção de *essays* avaliados *on line*, cujos resultados são eletronicamente disponíveis para o professor e para o avaliando, juntamente com feedback formativo detalhado sobre o processo de avaliação e comentários específicos para cada caso. É interessante notar que o referido software proporciona ao avaliando a possibilidade de correção e recepção de feedback também de rascunhos elaborados pelo avaliando a partir de determinado tópico. Esse material oferece a possibilidade de prever um resultado e reescrever uma resposta final adequando-a para que tenha melhor conteúdo e qualidade. Além disso, oferece a ambos, avaliador e avaliando, acesso a todas as produções submetidas, ou seja, aos rascunhos e à produção final, bem como ao conceito e comentários finais.

Em um outro estudo intitulado *Automated Essay Scoring Versus Human Scoring*, Jijun Wang e M. S. Brown (2007) pesquisam a diferença entre a avaliação de *essays* efetuada por professores e a avaliação efetuada *on line* por meio de um software denominado *IntelliMetric*¹⁶². 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 Jijun Wang e M. S. 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.

¹⁶¹ Software Educacional Teoria na Prática

¹⁶² Assim como o ETIPS utilizado por Cassandra Scharber, Sara Dexter e Eric Riedel, o *IntelliMetric* oferece ao avaliado as mesmas possibilidades em relação a resultados.

Pode-se observar, contudo, que com 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.

3.7. Observação

Dentro da categoria de itens que apresentam menos limitações estão os projetos, portfólios, apresentações, demonstrações, atividades de grupo, discussões e entrevistas. Nas palavras de Kathleen Scalise e Bernard Gifford, “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 frequentemente avaliação humana e/ou comparação humana fazem-se desejáveis” (SCALISE & GIFFORD, 2206, 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 colegas¹⁶³.

É de valia ressaltar que Kathleen Scalise e Bernard 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 avaliando, que precisa entender as abordagens utilizadas para se chegar a determinado resultado, evitando, assim possíveis impactos que podem ser ocasionados por resultados avaliatórios.

¹⁶³ Para aprofundamento em avaliação entre colegas (*Peer Assessment*) sugerimos a leitura de artigo de Ariwa intitulado *Peer Assessment and Tutoring*, apresentado na Conferência Anual da Associação de Pesquisa Educacional Escocesa realizada em Perth, 27-29 de novembro de 2003.

4. Comentários finais

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

No presente artigo, apresentou-se uma breve introdução sobre avaliação, presencial e a distância e foram analisados 11 itens referentes ao ensino/aprendizagem de idiomas. Tais itens variam entre, em um extremo, questões cujas respostas são totalmente limitadas, em outro extremo, questões que possibilitam que o avaliando 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 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.

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

Certamente, o ensino/aprendizagem do século XXI e XXII progredirá na direção da educação a distância; acredita-se, então, como educadora 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, é inegável que as oportunidades proporcionadas pelo uso de novas e, cada vez mais efetivas tecnologias venham contribuir, cada vez mais significativamente para o processo de ensino/aprendizagem de idiomas, principalmente no que diz respeito à educação a distância.

REFERÊNCIAS BIBLIOGRÁFICAS

ARIWA, Ezendu Ifeanyi. *Peer Assessment and Tutoring (Comparative analysis of e-Learning with respect to Business Information Technology)* Paper presented at the Scottish Educational Research Association Annual Conference, Perth, 27-29 November 2003.

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

GARRISON, Donn Randy; ANDERSON, Terry. *E-Learning in the 21st Century: A Framework for Research and Practice*. Oxon & New York: Routledge Falmer, 2003.

HALADYNA, Thomas M. *Multiple-Choice Formats*. In: _____. *Developing and Validating Multiple-Choice Test Items*. New York: Routledge, 1994.

MCDONALD, Mary E. *Developing Multiple-Choice Items*. In: _____. *Systematic Assessment of Learning Outcomes*. London: Prentice Hall, 2002.

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

OSTERLIND, Steven J. *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and Other Formats*. London: Kluwer Academic Publisher, 1998

SCALISE, Kathleen; GIFFORD, Bernard. Computer-Based Assessment in E-Learning. *Journal of Tecnology, Learning and Assessment*. Disponível em: <<https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1653/1495>>. Acesso em: 28-04-2017

SCHARBER, Cassandra; DEXTER, Sara; RIEDEL, Eric. Students' Experiences with an Automated Essay Scorer. *Journal of Technology, Learning, and Assessment*. Disponível em: <<http://files.eric.ed.gov/fulltext/EJ838623.pdf>>. Acesso em: 27-04-2017.

SHEPARD, Lorrie. *Interview on assessment issues with Lorrie Shepard*. Educational Researcher: London, 1991.

SHOHAMY, Elana; INBAR, Ofra. *The Language Assessment Process: A "Multiplism" Perspective*. London: Calper, 2006.

WANG, Jijun; BROWN, M. S. Essay Scoring Versus Human Scoring: A Comparative Study. *Journal of Technology, Learning, and Assessment*. Disponível em: <<http://www.academypublication.com/issues/past/tpls/vol02/04/10.pdf>>. Acesso em: 27-04-2017.