



Revista Brasileira de Linguística Aplicada

ISSN: 1676-0786

rblasecretaria@gmail.com

Universidade Federal de Minas Gerais  
Brasil

Soufen Tumolo, Celso Henrique; Braga Tomitch, Lêda Maria  
Avaliando a leitura em inglês: uma reflexão sobre itens de testes  
Revista Brasileira de Linguística Aplicada, vol. 7, núm. 2, julio-diciembre, 2007, pp. 67-89  
Universidade Federal de Minas Gerais  
Belo Horizonte, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=339829601004>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica  
Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal  
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

# Avaliando a leitura em inglês: uma reflexão sobre itens de testes

---

Celso Henrique Soufen Tumolo<sup>1</sup>

Lêda Maria Braga Tomitch<sup>2</sup>

Universidade Federal de Santa Catarina - UFSC

**RESUMO:** O desenvolvimento de testes de qualquer natureza requer um estágio de investigação dos itens elaborados, investigação essa que deve ser feita considerando como critério principal a validade de construto. Neste artigo, trazemos uma análise de itens de testes desenvolvidos para avaliar a habilidade de leitura em inglês como língua estrangeira. Dividimos os itens em defensáveis e não defensáveis: os defensáveis permitem a demonstração da habilidade como definida no construto, enquanto os não defensáveis não permitem coletar evidências para uma interpretação válida da habilidade relevante. Sugerimos uma divisão dos itens não defensáveis em três categorias: itens independentes do texto, itens com pistas para a resposta e itens de vocabulário, cada qual apresentando problemas para validação de construto. É possível concluir que itens usados em testes podem não ser adequados ao objetivo para o qual foram desenvolvidos. As análises e conclusões aqui feitas pretendem colaborar com o desenvolvimento de testes que possibilitem coletar evidências para uma interpretação mais válida da habilidade de leitura em inglês.

**PALAVRAS-CHAVE:** leitura, testes, validade, língua estrangeira.

**ABSTRACT:** Test development of any nature requires a stage for the investigation of the items designed, which must be carried out considering construct validity as the main criterion. In this article we bring an analysis of test items designed to assess the reading ability in English as a foreign language. We divide the items into defensible and non-defensible: defensible items allow for the demonstration of the reading ability as described in the construct, whereas non-defensible items do not allow for collecting evidence for a more valid interpretation of the relevant ability. We suggest a division

---

<sup>1</sup> Doutor em Letras/Língua Inglesa e Linguística Aplicada e pesquisador na área de leitura do Departamento de Língua e Literatura Estrangeiras da Universidade Federal de Santa Catarina.

<sup>2</sup> Doutora em Letras/Linguística Aplicada. Professora e pesquisadora na área de leitura do Departamento de Língua e Literatura Estrangeiras da Universidade Federal de Santa Catarina.

of the non-defensible items into 3 categories: text-independent items, items with unintended clues for the answer, and vocabulary items, each one presenting some problems for construct validation. It is possible to conclude that items used in tests may not be adequate for the purposes they were designed. The analyses and conclusions presented here aim at collaborating with the development of tests that allow for the collection of evidence for more valid interpretations of the ability of reading in English as a foreign language.

KEY-WORDS: reading, tests, validity, foreign language.

## Introdução

Testes têm sido usados para diversos objetivos como classificação, certificação, seleção e mesmo formação. Seu desenvolvimento tem sido, assim, uma prática comum. Bachman e Palmer (1996) descrevem os três estágios necessários para esse desenvolvimento: planejamento, operacionalização e administração.

O primeiro estágio, planejamento, envolve a definição do que se quer medir, as especificações norteadoras do desenvolvimento de um teste. Faz-se, assim, necessária a definição do construto, entendido como o conjunto de habilidades e/ou conhecimentos que pode ser plausivelmente argumentado e/ou teoricamente justificado como esperado. O segundo, operacionalização, é caracterizado pela operacionalização do construto, isto é, o desenvolvimento de itens que sejam referenciados ao construto definido no primeiro estágio. O terceiro estágio, administração, muitas vezes negligenciado, envolve fases de revisão do próprio teste, isto é, o teste é colocado à prova. O *feedback* resultante é usado para saber se o teste está permitindo coletar informações necessárias para interpretação de habilidade baseada no desempenho de cada pessoa, isto é, se de fato o teste permite levantar evidências sobre habilidades e/ou conhecimentos representativos do, e relevantes para o, construto.<sup>3</sup>

Mas como saber se os testes desenvolvidos são defensáveis? Este texto tem como objetivo trazer uma reflexão sobre o desenvolvimento de testes, mais especificamente sobre a elaboração de itens que podem ser considerados defensáveis. Para isso, apresentamos análises tanto de itens que consideramos defensáveis como também os não defensáveis. O critério

---

<sup>3</sup> Para uma discussão sobre a importância das especificações e da revisão do teste, ver Tumolo (no prelo a).

usado para a análise da defensabilidade dos itens será validade. Seguindo a definição mais recente aceita por pensadores norte-americanos influentes como Lee Cronbach e Samuel Messick,<sup>4</sup> validade não deve ser um critério a ser usado para avaliar o teste, ou seja, não existe teste válido ou inválido. Validade refere-se às interpretações e às ações decorrentes do uso de um teste e podem ser consideradas mais, ou menos, válidas. Dessa forma, um item será considerado defensável se ele permitir coletar evidências que contribuam para interpretações e ações com maior grau de validade para o objetivo do teste.

Dentre os três tipos de validade considerados atualmente importantes na literatura sobre testes, a saber, validade de critério, de conteúdo, e de construto,<sup>5</sup> a validade de construto é a mais importante (MESSICK, 1989), já que as interpretações baseadas nos testes devem ser feitas em termos de habilidades e/ou conhecimentos descritos como construto. Assim, a análise de itens de teste, neste artigo, considerará as evidências em relação ao construto de leitura em inglês como língua estrangeira.

Messick (1989) considera dois pontos importantes na investigação da validade de construto. O primeiro refere-se ao fato de testes, por vezes, não incluírem aspectos essenciais estabelecidos no construto usado, ou de incluir aspectos irrelevantes que não deveriam ter sido incluídos, ou ambos. O segundo refere-se ao fato de que a pontuação do teste deve refletir as habilidades e/ou conhecimentos que podem ser atribuíveis ao construto usado, e não a qualquer construto alternativo.

Baseados no primeiro ponto de Messick, podemos dizer que um teste que não inclui algo que deveria incluir não será representativo do construto usado; e um teste que inclui algo que não deveria ter traços de irrelevância. Assim, podemos considerar que os dois aspectos mais importantes para a investigação de validade de construto são representatividade e relevância. As perguntas a serem feitas em um processo de investigação de validade de construto devem se basear na seguinte pergunta geral: em que medida os itens do teste permitem demonstrar habilidades e/ou conhecimentos representativos e/ou relevantes em relação ao construto? Baseados no segundo ponto de Messick, podemos indagar: será que os itens não estão

---

<sup>4</sup> Para uma revisão das contribuições de Cronbach e Messick, ver Tumolo (2005).

<sup>5</sup> Para uma revisão sobre definições e redefinições do conceito 'validade' e os vários tipos sugeridos e mantidos durante o século XX, ver Tumolo (2005).

associados a um construto que não seja aquele definido durante os estágios de elaboração do teste, mas sim, a um construto alternativo?

A análise apresentada neste artigo é feita somente baseada na investigação da seguinte pergunta: os itens usados possibilitam coletar evidências para interpretação sobre habilidade em leitura em inglês como língua estrangeira? A idéia é que um teste, através de seus itens, deve permitir que a pontuação conseguida por cada pessoa reflita seu nível da habilidade, neste caso, a habilidade de leitura em inglês como língua estrangeira. Assim, um teste que permite interpretações mais válidas é um teste que possibilita a diferenciação entre um indivíduo com maior nível da habilidade em questão daquele com menor nível da habilidade.

Na próxima seção, apresentamos a definição do construto leitura a ser usado para nossa análise de itens de testes que objetivam avaliar a habilidade de leitura em inglês. Nas seções seguintes, procedemos nossa análise de itens escolhidos para ilustrar uma investigação de validade de construto. Por fim, fazemos nossas considerações finais.

### **Definição do construto**

Buscando compreender o construto leitura, fazemos uso das definições postuladas por Gagné, Yekovich e Yekovich (1993) sobre os componentes da leitura, como também das contribuições de Hutchins (1987) sobre os processos envolvidos no desenvolvimento das microestruturas e das macroestruturas textuais necessárias para a compreensão da leitura.

Gagné *et al* (1993) definem leitura como um processo que envolve conhecimento declarativo e conhecimento procedural. O conhecimento declarativo refere-se ao conhecimento dos tópicos tratados pelo texto, esquemas formais e conhecimento de vocabulário. Embora sejam necessários para o processo de compreensão da escrita (KINTSCH; VAN DIJK, 1978; HUTCHINS, 1987), esses conhecimentos são, em grande parte, de natureza individual e de difícil definição (CLAPHAM, 1996), não devendo, portanto, ser o foco de itens usados em testes desenvolvidos para medir habilidades lingüísticas.<sup>6</sup> Entendemos que a leitura de um texto envolve o acionamento e o engajamento de esquemas relevantes para que um modelo mental adequado seja construído e a compreensão alcançada (TOMITCH, 2000;

---

<sup>6</sup> Para uma discussão ampla sobre essa questão, ver Tumolo (2005).

2002; 2003). Entretanto, acreditamos que, em situações de teste, em que o objetivo for medir a habilidade lingüística em língua estrangeira, textos com assuntos muito específicos, que requeiram conhecimento especializado de uma determinada área, devam ser preteridos.

O conhecimento procedural, por sua vez, envolve processos componentes mais inferiores, como também os mais superiores. Os processos inferiores são, segundo Gagné *et al.* (1993), decodificação e compreensão literal. Decodificação refere-se ao processo de reconhecimento do vocabulário que ativa seu significado na memória de longo prazo. Esse processo estimula os processos de compreensão literal, que envolvem acesso lexical – identificação e seleção do significado apropriado das palavras, e também parseamento – atribuição de significado a unidades maiores de significados (frases, orações ou sentenças), com base nas regras sintáticas da língua. Esses processos inferiores referem-se ao conhecimento do sistema lingüístico, neste caso, da língua inglesa.

Dentre os processos superiores apontados pelos autores, os considerados relevantes para a determinação do construto usado para análise neste artigo são integração e sumarização.<sup>7</sup> Integração permite uma representação coerente das idéias presentes no texto, quando o leitor estabelece relações entre as proposições. Sumarização é o esboço mental de proposições hierarquicamente organizadas que resume as idéias principais do texto. Esses processos são considerados superiores, pois envolvem inferências, por parte do leitor, baseadas no conhecimento de mundo e do assunto tratado no texto.

Baseando-se no modelo explicativo do processo de compreensão postulado por Kintsch e van Dijk (1978) e van Dijk e Kintsch (1983), Hutchins (1987) enfatiza que a compreensão escrita envolve necessariamente dois processos: o reconhecimento das estruturas textuais e a habilidade de sumarizar. Em relação ao primeiro aspecto, há uma distinção das estruturas textuais entre microestruturas e macroestruturas.

As microestruturas são construídas a partir do estabelecimento das relações anafóricas e dos elos léxico-semânticos, da coesão semântica e da progressão temática. Relações anafóricas referem-se ao estabelecimento de referência, como no caso de pronomes, enquanto elos léxico-semânticos

---

<sup>7</sup> Gagné *et al.* falam também do processo de elaboração e dos processos de monitoramento da compreensão, considerados irrelevantes para nossa análise.

referem-se à identificação de itens lexicais recorrentes e pertencentes ao mesmo campo semântico, como *tráfego*, *automóvel* e *pára-choques*. Coesão semântica é estabelecida através de conjunções ou de itens lexicais como *situação*, *questão*, *diferença*, *comparação*, cujos significados são atribuídos por referência interna ao texto. A progressão temática é conseguida através do reconhecimento do já conhecido ou do já expresso no texto, implicando a identificação do tema e rema das sentenças e/ou parágrafos.

As macroestruturas, por sua vez, são o resultado das generalizações das proposições estabelecidas a partir das microestruturas e do reconhecimento das proposições irrelevantes ou redundantes que devem ser ignoradas, e constituem, assim, a idéia principal de um texto, uma macroproposição. Elas podem ser explicitadas pelo autor do texto por sinalizadores como *em suma*, *principalmente*, etc., ou inferidas pelo leitor a partir do uso de seus conhecimentos do tópico e/ou de mundo, como também conhecimento da própria organização textual.

Em relação ao segundo processo, sumarizar, Hutchins (1987) acrescenta que é um trabalho feito cotidianamente quando produzimos condensações das partes essenciais de um discurso oral ou escrito, enfatizando que são expressões da macroestrutura dos textos, interpretada a partir do conhecimento anterior de cada indivíduo, envolvendo necessariamente quatro componentes: a compreensão da microestrutura, a identificação do esquema global, a aplicação de macro-regras para generalizar e condensar a uma representação da macroestrutura, e, por fim, a expressão dessa macroestrutura em um texto coerente. Esses processos componentes da habilidade de leitura apontados por Gagné *et al.* (1993) e por Hutchins (1987) são, neste artigo, utilizados para a análise de itens feita nas próximas seções, sempre com vistas a uma investigação da validade de construto.

### **Análise de itens**

Análise de itens ou do desempenho de um item de teste é um procedimento considerado, na literatura especializada, essencial no desenvolvimento de testes. Nessa fase, testes com itens ainda em análise podem ser aplicados em grupos em uma espécie de simulação do real, como no caso do IELTS (WEIR; MILANOVIC, 2003), ou mesmo em situações reais de teste em que o item não é considerado para efeito de contagem da pontuação, como no caso do TOEFL (PIERCE, 1992). Técnicas específicas

podem ser usadas para estimar, por exemplo, o nível de dificuldade de cada item. Contudo, apesar de sua importância, é um procedimento pouco usado, por várias razões. Para testes de pequena escala, pode ser muito trabalhoso e/ou oneroso para o professor. No caso de testes de grande escala, como o vestibular, pode haver o problema de vazamento de informações, o que comprometeria seu uso.

As análises aqui feitas não são baseadas em itens testados em situações reais ou simulações, mas, sim, em balanço de argumentos, procedimento de investigação de validade de natureza mais interpretativa, defendido por autores como Kane (1992) e Chapelle (1994; 1999).<sup>8</sup> Elas têm por objetivo propiciar uma reflexão que auxilie o processo de desenvolvimento de itens que permitem interpretações e ações mais válidas em relação ao construto. Mostramos, a seguir, itens defensáveis, como também itens não defensáveis, com suas respectivas análises.

### **Itens defensáveis**

Nesta seção, fazemos uma análise de itens que podemos considerar mais defensáveis, isto é, itens que permitem coletar evidências para a interpretação de habilidade com maior grau de validade de construto.

Testes que avaliam a habilidade de leitura em língua estrangeira têm tradicionalmente incorporado itens que focam a compreensão da idéia principal do texto e algumas vezes de parágrafos, compreensão do significado de palavras em contexto, como também na identificação de informações localizadas (datas, números, lugares, nomes, etc.) e identificação de referência (pronomes relativos, pronomes objetivos, pronomes possessivos, etc.).

Itens que focam a compreensão de idéia principal do texto têm o argumento favorável que avaliam todo o trabalho de leitura envolvendo processos inferiores, como estabelecimento da relação sintática e atribuição de significado de palavras, e processos superiores de inferências, como também a progressão temática e a sumarização. Apesar dos argumentos favoráveis, existem aqueles que consideramos desfavoráveis. A construção da idéia principal depende do processo de sumarização. Esse processo envolve o desenvolvimento de uma macroestrutura textual construída, envolvendo o conhecimento de mundo de cada indivíduo, levando a interpretações distintas de um mesmo texto (HUTCHINS, 1987). Desta forma,

---

<sup>8</sup> Para uma discussão de método de investigação de validade, ver Tumolo (2005).

esses itens podem levar à dificuldade, em situações de teste, de identificar o que pode ser considerado como resposta correta ou incorreta.

Pearson e Johnson (1978) apontam para uma distinção entre intrusão textual e intrusão de *script*. Os autores definem a primeira como uma leitura resultante da seleção aleatória de segmentos do texto, sem argumento de plausibilidade, e a segunda, como resultante da cabeça do leitor, para o qual não existe linha plausível de raciocínio. O primeiro caso pode ser de fácil identificação e a resposta considerada incorreta. O segundo sempre será mais difícil de identificar, já que leitores terão linhas distintas de raciocínio e sumarização, baseadas em seus conhecimentos prévios distintos. Assim, no balanço dos argumentos para validade de construto, esses itens parecem ser defensáveis com a ressalva que as distinções sejam consideradas no momento da interpretação (e pontuação) das respostas.<sup>9</sup>

Analisemos itens focando compreensão mais local. Consideremos o item 1:<sup>10</sup>

*Parte do texto*

Almost all of an organism's energy can be diverted to reproduction, with very little allocated to building the body. Organisms at this extreme are 'opportunists'. At the other extreme are 'competitors', almost all of whose resources are invested in building a huge body, with a bare minimum allocated to reproduction.

*Item*

According to the passage, the classification of organisms as 'opportunists' or 'competitors' is determined by

- a) how the genetic information of an organism is stored and maintained.
- b) the way in which the organism invests its energy resources.
- c) whether the climate in which the organism lives is mild or extreme.
- d) the variety of natural resources the organism consumes in its environment.

Para responder corretamente a esse item de múltipla escolha, o leitor precisa estabelecer a coesão e resolver a elipse da qual *little* faz parte, isto é, o leitor precisa compreender que *little* se refere a *energy*, palavra omitida. Reinterpretar *all of energy diverted to reproduction, with very little allocated to building the body* no texto como *the way in which the*

---

<sup>9</sup> Particularmente quando itens como perguntas abertas ou dissertativas são usados.

<sup>10</sup> Item extraído da prova de demonstração do TOEFL iBT, disponível no site do ETS: [www.ets.org](http://www.ets.org), acessado em julho de 2004.

*organism invests its energy resources* no item demanda reconhecimento do vocabulário no contexto e processo de inferência, o que Nuttall (1996) chamou de inferência elementar usada para reinterpretação de informação. Esse item parece ter argumentos favoráveis mais relevantes, já que permite coletar evidências para uma interpretação mais válida em relação ao construto. Analisemos outros itens. Consideremos o item 2:<sup>11</sup>

*Parte do texto*

She was the daughter of Bronson Alcott, a well-known teacher, intellectual, and free thinker who advocated abolitionism, women's rights, and vegetarianism long before they were popular. He was called a man of unparalleled intellect by his friend Ralph Waldo Emerson. Bronson Alcott instilled in his daughter his **lofty** and spiritual values and in return was idolized by his daughter. (cont.)

*Item*

Na linha 5, a palavra "lofty" assemelha-se mais em termos de significado a

- a) comum.
- b) generoso.
- c) egoísta.
- d) simpático.

No balanço dos argumentos, esse item de múltipla escolha parece ter mais argumentos favoráveis. A palavra *lofty* é bem pouco freqüente,<sup>12</sup> o que possibilita que esse não seja um item de vocabulário ou item independente do texto, respondido sem referência ao texto.<sup>13</sup> O leitor, assim, precisa fazer uso do contexto para inferência do significado da palavra, isto é, das informações espalhadas pelo texto de forma a integrá-las. Embora possa ser considerado fácil, visto que há as palavras cognatas *spirituale* e *idolized* próximas, acertar esse item demonstra um trabalho de leitura pelo reconhecimento de vocabulário em contexto, estabelecimento de relações (integração), em particular de coesão léxico-semântica. Esse item permite coletar evidências para interpretação mais válida em relação ao construto. Consideremos o item 3:<sup>14</sup>

<sup>11</sup> Item de teste usado por professor de Inglês Instrumental, lecionando na UFSC.

<sup>12</sup> Esta palavra está na categoria de palavras com mais baixo índice de freqüência de uso no *Collins Cobuild Dictionary*.

<sup>13</sup> Tanto item de vocabulário como item independente do texto serão tratados em seções específicas abaixo.

<sup>14</sup> Item extraído da prova de demonstração do TOEFL iBT, disponível no site do ETS: [www.ets.org](http://www.ets.org), acessado em julho de 2004.

*Parte do texto*

Similarly, a plant or animal cannot squander all its energy on growing a big body if none would be left over for reproduction, for this is the surest way to extinction.

*Item*

The word squander in the passage is closest in meaning to

- a) extend.
- b) transform.
- c) activate.
- d) waste.

Esse item de múltipla escolha tem como palavra alvo *squander*, não cognata e também não freqüente,<sup>15</sup> sendo provavelmente desconhecida, fazendo do item um item que requer referência ao texto. A informação *expenditure of energy* será usada pelo leitor para inferência de *squander*. As palavras dadas como alternativas não impõem problemas. A palavra chave (resposta correta), *waste*, é freqüente.<sup>16</sup> Caso não seja conhecido, o item pode ser respondido pela eliminação dos distratores, já que todos são cognatos. O item permite coletar evidências sobre a habilidade de identificar e selecionar o significado apropriado da palavra baseada na compreensão local das palavras próximas, assim com foco na habilidade de construção de microestrutura textual.

Analise agora itens focando referência, que permitem coletar evidências sobre a habilidade de integrar elementos textuais. Consideremos o item 4,<sup>17</sup> usado no mesmo teste que o item anterior:

*Parte do texto*

Similarly, a plant or animal cannot squander all its energy on growing a big body if none would be left over for reproduction, for this is the surest way to extinction.

---

<sup>15</sup> Esta palavra está na categoria de palavras com mais baixo índice de freqüência de uso no *Collins Cobuild Dictionary*.

<sup>16</sup> Esta palavra está na categoria de palavras freqüentemente usadas no *Collins Cobuild Dictionary*.

<sup>17</sup> Item extraído da prova de demonstração do TOEFL iBT, disponível no site do ETS: [www.ets.org](http://www.ets.org), acessado em julho de 2004.

*Item*

The word none in the passage refers to

- a) food.
- b) plant or animal.
- c) energy.
- d) big body.

Esse item permite ter evidências sobre a habilidade de estabelecer referência entre *none* e *energy*. Estabelecer essa relação é essencial para integrar informações e para a compreensão. Hutchins (1987) descreve esse processo como o estabelecimento das relações léxico-semânticas pela coesão textual, nesse caso através de itens lexicais, cujos significados são atribuídos por referência interna ao texto.<sup>18</sup> Assim, itens como esse fornecem evidências sobre a habilidade do leitor de estabelecer relações no texto de forma a integrá-lo. Consideremos o item 5:<sup>19</sup>

*Item*

Select the proposition(s) which contains (contain) correct references to the following words, underlined in the text

- 01. which (paragraph 1): the films
- 02. their (paragraph 1): Mary Pickford, Douglas Fairbanks, D.W. Griffith, and Chaplin
- 03. its (paragraph 2): the circus clown
- 08. it (paragraph 2): the new technology
- 16. them (paragraph 2): several tributes
- 32. the actor (paragraph 2): Chaplin
- 64. that (paragraph 2): the actor

Esse item possibilita ter evidências sobre a habilidade do leitor de identificar referência dos vários tipos, o que permite demonstrar sua habilidade de integrar informações e estabelecer relações entre proposições, ou o desenvolvimento das micro e macroestruturas, como descritos por Hutchins (1987). Nesse caso específico, trata-se de um item que envolve um conjunto de 07 subitens binários, em que cada um pode ser ou verdadeiro ou falso.

---

<sup>18</sup> Nuttall (1996) refere-se a esse processo como coesão lexical, essencial para estabelecer as relações determinadas pelo escritor, e compara com um grupo de palavras dentre as quais estão: questão, método, eventos, explicação, visões, opiniões, fenômenos e abordagem, todas a terem seus significados preenchidos dentro do contexto, por referência interna ao texto.

<sup>19</sup> Item de teste usado pelo vestibular da UFSC, ano 2003.

Os subitens usados não trazem necessariamente a resposta correta, como no caso de múltiplas escolhas. Dessa forma, o item diminui a possibilidade do acerto por 'chute', aumentando sua confiabilidade, que é condição para a validade. Esse item permite coletar evidências para interpretação de habilidade com alto grau de validade de construto. Esses são apenas alguns exemplos de itens que consideramos defensáveis. Na próxima seção, examinamos itens não defensáveis.

### Itens não defensáveis

Diferentemente dos defensáveis, os itens não defensáveis apresentam problemas no momento da coleta de evidências para interpretação da habilidade sendo medida, o que pode ser identificado em uma investigação de validade de construto. Estaremos focando três tipos de problemas que podem ser recorrentes no processo de desenvolvimento e uso de testes: a) itens independentes do texto, b) itens que fornecem pistas para respostas, e c) itens de vocabulário.

### Itens independentes do texto

Itens independentes do texto são aqueles que podem ser respondidos sem referência ao texto. Como o construto usado refere-se à habilidade de leitura em inglês como língua estrangeira, texto é definido, para efeito de nossas análises, como texto escrito, envolvendo necessariamente o sistema lingüístico da língua inglesa. Alguns itens podem ser respondidos sem referência ao texto escrito. Consideremos o item 6:<sup>20</sup>

#### *Parte do texto*

Thus, a minimum investment has been made in the body that becomes a platform for seed dispersal. These very short-lived plants reproduce prolifically; that is to say they provide a constant rain of seed in the neighborhood of parent plants.

#### *Item*

The word dispersal in the passage is closest in meaning to

- a) development.
- b) growth.
- c) distribution.
- d) protection.

---

<sup>20</sup> Item extraído da prova de demonstração do TOEFL iBT, disponível no site do ETS: [www.ets.org](http://www.ets.org), acessado em julho de 2004.

Esse item de múltipla escolha pode ser respondido apenas pelo reconhecimento das palavras cognatas *dispersal* (palavra alvo) e *distribution* (resposta correta), com base somente no conhecimento da língua mãe. Assim, pode ser considerado um item independente de seu texto e não permite coletar evidências sobre a habilidade de leitura em inglês como língua estrangeira. Consideremos o item 7:<sup>21</sup>

*Item*

A que se refere ‘*Shadow Point*’? Por que recebeu esse nome?

O item tem por base um texto que descreve as conseqüências da construção de um prédio alto em uma cidade pequena onde existem somente casas, texto esse ilustrado por uma figura (ANEXO A). Para responder, o leitor pode fazer uso da figura e de processo de inferência, baseado em seu conhecimento de mundo. Dessa forma, estaria respondendo ao item sem referência ao texto escrito, podendo, assim, fornecer apenas evidências fracas para a interpretação da habilidade em leitura em inglês como língua estrangeira.

Itens independentes do texto podem assumir várias formas e não são raros. Alderson *et al.* (1995) alertam que esse tipo de item é comum, e enfatizam que isso tem medido mais conhecimento de mundo do que competência ou habilidade em leitura em língua estrangeira. Nuttall (1996) faz ressalva semelhante, alertando que itens desenvolvidos para medir a habilidade em leitura em língua estrangeira não podem ser respondidos sem referência ao texto escrito para serem considerados confiáveis. E como confiabilidade é condição para validade, eles não permitem coletar evidências para uma interpretação mais válida. Não são, portanto, defensáveis.

### **Itens que fornecem pistas para respostas**

Itens com pistas para respostas, segundo tipo de item não defensável analisado aqui, são aqueles que fornecem elementos para a resposta correta. Popham (1981) se refere a eles como aqueles que fornecem pistas não intencionadas (*unintended clues*), e Messick (1989), como aqueles que possibilitam facilidade (para a resposta) irrelevante ao construto sendo medido; nas palavras do autor, quando “pistas externas no item ou formato do teste permitem que alguns indivíduos respondam corretamente de formas

---

<sup>21</sup> Item de teste usado pelo vestibular da UNICAMP, ano 1998.

irrelevantes ao construto sendo investigado” (p. 35, tradução nossa). Consideremos o item 8:<sup>22</sup>

*Item*

O texto menciona mudanças. Que mudanças são essas?

Esse item fornece, na língua mãe, informações para sua resposta. Essas informações sobre mudanças, juntamente com a figura fornecida com o texto (ANEXO A), já fornecem subsídios para a resposta, permitindo, assim, que o item seja respondido, em grande parte, com base nas informações do próprio item dadas na língua mãe, não na língua-alvo. Consideremos outro item semelhante, o item 9<sup>23</sup>:

*Item*

De que maneira a violência urbana pode estar afetando a saúde de pessoas idosas?

Podemos fazer, novamente, a mesma análise. A pergunta *De que maneira a violência urbana pode estar afetando a saúde de pessoas idosas?* contém a proposição de que a violência urbana pode estar afetando a saúde das pessoas idosas. Nesse caso, além das pistas fornecidas na língua-mãe para a resposta, por um processo de inferência baseado no conhecimento de mundo, pode-se chegar à resposta satisfatória. Isso pode explicar a razão pela qual o item foi considerado fácil nas estatísticas baseadas nas pontuações, e mais fácil ainda para os futuros estudantes de medicina.<sup>24</sup> Esses itens podem ser considerados um problema em uma investigação de validade de construto. Como reforça Messick (1989), itens como esses são uma ameaça à validade de construto, pois fornecem evidências irrelevantes ao construto, evidências (ou pistas) na língua mãe para uma interpretação da habilidade de leitura em língua estrangeira.

### **Itens de vocabulário**

Itens de vocabulário, terceiro tipo de item não defensável analisado, são aqueles que medem conhecimento de vocabulário, independentemente do contexto em que está inserido e de possíveis processos de inferência

---

<sup>22</sup> Item de teste usado pelo vestibular da UNICAMP, ano 1998.

<sup>23</sup> Item de teste usado pelo vestibular da UNICAMP, ano 1998.

<sup>24</sup> Os dados estatísticos são apresentados em publicação feita pela comissão responsável pelo vestibular (Comvest, Comissão Permanente para os Vestibulares).

de seu significado. Conhecimento de vocabulário é essencial e tem uma correlação positiva com a habilidade de leitura. Deve, porém, ser considerado distinto do construto de leitura, por ter características próprias, podendo, até mesmo, ser considerado um construto pré-requisito,<sup>25</sup> argumento que tem nosso suporte.

Há estudos que revelam distinções de construtos, mesmo quando altamente correlacionados. Para ilustrar, Messick (1989) cita o exemplo do conhecimento matemático e do raciocínio quantitativo, e argumenta que, mesmo sendo altamente correlacionados, devem ser considerados como construtos separados, já que podem haver evidências empíricas para tal distinção (p. 51). O mesmo pode ser dito na relação entre leitura e vocabulário, entendidos como construtos distintos aqui, apesar de altamente correlacionados.

Consideremos, para nossa análise de itens de vocabulário enquanto itens não defensáveis, dois itens de perguntas abertas,<sup>26</sup> usados em um mesmo teste.<sup>27</sup> O primeiro é o item 10:

*Parte do texto*

He ruled over the other Titans until his son Zeus dethroned him and seized power for himself.

*Item*

Dê um significado para *seized*

Tomemos a palavra *seized*. É uma palavra pouco usada,<sup>28</sup> provavelmente desconhecida da maioria dos brasileiros, sendo uma boa escolha para entender o processo de leitura, já que deve envolver o uso do contexto para a inferência de seu significado. O conhecimento gramatical

---

<sup>25</sup> Messick (1989) menciona construto pré-requisito no argumento que sua ausência constitui grande limite de desempenho (p. 65).

<sup>26</sup> Bachman e Palmer (1996) fazem uma distinção de três tipos de respostas esperadas: resposta selecionada, resposta de produção limitada, e resposta de produção extensa. Resposta selecionada refere-se a itens como múltipla escolha; resposta de produção limitada a respostas que não se estendem para além de duas linhas; e respostas de produção extensa a respostas que vão muito além de duas linhas. Os itens 10 e 11, nessa distinção, requerem respostas do tipo de produção limitada.

<sup>27</sup> Itens de teste usados pelo vestibular da UNICAMP, ano 1999.

<sup>28</sup> Esta palavra está na categoria de palavras de frequência média de uso no *Collins Cobuild Dictionary*.

(verbo no tempo passado) auxilia na produção da inferência. Assim, o leitor chega a uma compreensão mais geral da sentença, o que ativa os esquemas conceituais associados a *Titãs*, *Zeus* e *poder*, possibilitando a inferência da palavra exigida na questão. Além disso, *seize* e *power* são palavras com alto grau de colocação, isto é, *seize* ocorre freqüentemente com *power*, o que auxilia o processo de inferência de seu significado a partir de *power* e, como consequência, o provimento da resposta satisfatória. Consideremos, agora, o segundo item, item 11:

*Parte do texto*

Fear of fraud was one reason the European Commission wanted the coins to look the same in every country – the greater the number of different coins, the harder it is to recognize a phony.

*Item*

Dê um significado para *phony*

*Phony* é bem pouco freqüente<sup>29</sup> e, assim, uma outra boa escolha. Para responder à questão, é necessário, para a maioria dos leitores, fazer uso do contexto para inferência de seu significado. Embora os esquemas conceituais associados à fraude sejam provavelmente ativados a partir da palavra cognata *fraud*, as restrições semânticas são menores e menos definidas para a inferência do significado da palavra *phony* nesse contexto se comparadas às do contexto da palavra *seized*, discutida anteriormente. Ademais, *phony* está inserida em uma expressão que, por sua vez, também é pouco conhecida, dificultando o trabalho de compreensão local e, por conseguinte, dificultando a inferência de seu significado.

A dificuldade de responder corretamente à pergunta é reconhecida em publicação feita pela comissão<sup>30</sup> responsável pelo teste. Enquanto a questão sobre *seized* foi respondida corretamente por muitos, e assim considerada fácil, a questão sobre *phony* foi a mais difícil da prova, com um total de 46,9% de respostas com pontuação de zero em uma escala de 0 a 5. Embora testes contenham tradicionalmente perguntas fáceis e difíceis, nessa comparação, fica a pergunta para uma investigação sobre a validade do construto: o que pode ser interpretado sobre a habilidade de leitura

---

<sup>29</sup> Esta palavra está na categoria de palavras com mais baixo índice de freqüência de uso no *Collins Cobuild Dictionary*.

<sup>30</sup> COMVEST (Comissão Permanente para os Vestibulares), da UNICAMP.

daqueles que conseguiram responder corretamente o significado da palavra do primeiro item (item 10), mas não o significado da palavra do segundo (item 11)? Como o contexto não auxilia a inferência do significado da palavra solicitada, uma resposta plausível a essa pergunta pode ser o conhecimento de vocabulário específico. Vocabulário específico passa, então, a ser o fator relevante para a resposta esperada, isto é, conhecer a palavra possibilita o acerto, enquanto não conhecer a palavra o impossibilita. Ademais, a compreensão do texto pode ter sido satisfatória, o que o item (item 11) não permite demonstrar. No balanço dos argumentos, o item 11 parece ter mais argumentos desfavoráveis ao seu uso para a interpretação da habilidade de leitura, assim, não é defensável.

Consideremos o item 12:<sup>31</sup>

*Parte do texto*

all aboard may have hastily abandoned the ship to avoid what they imagined might be its imminent destruction from an explosion of the combustible cargo.

*Item*

The word imminent in paragraph 3 is closest in meaning to

- a) impending.
- b) atrocious.
- c) inevitable.
- d) prospective.

Embora a palavra *imminent* seja relativamente pouco freqüente,<sup>32</sup> ela é uma palavra cognata e leitores brasileiros podem reconhecê-la e ter uma compreensão local sem dificuldades. Mesmo assim, esses leitores podem não conseguir demonstrá-la através do item. Seguindo as definições do *Collins Cobuild Dictionary*, a alternativa correta parece ser *impending*. *Impending* é uma palavra bem pouco freqüente,<sup>33</sup> não cognata, sendo provavelmente desconhecida dos leitores brasileiros. Como é dada como alternativa, ela não está inserida em contexto, não podendo, assim, ter seu

---

<sup>31</sup> Item do livro *Cambridge Preparation for the TOEFL Test, Practice Test 2*, de Jolene Gear e Robert Gear, publicado em 2002.

<sup>32</sup> Esta palavra está na categoria de palavras com baixo índice de freqüência de uso no *Collins Cobuild Dictionary*.

<sup>33</sup> Esta palavra está na categoria de palavras com o mais baixo índice de freqüência de uso no *Collins Cobuild Dictionary*.

significado inferido. O fator relevante nesse item parece, assim, ser o conhecimento de vocabulário, nesse caso, de pouca frequência, já que conhecer a palavra *impending* permite a resposta correta enquanto não conhecê-la impede a resposta correta. No balanço dos argumentos, esse item parece ter mais argumentos desfavoráveis. Consideremos, agora, o item 13:<sup>34</sup>

*Parte do texto*

As a result, Luisa had to begin helping to support her family at a young age, by taking a variety of low-paying jobs as a seamstress, a maid, and a tutor.

*Item*

Qual das seguintes atividades NÃO foi exercida por Luisa para ganhar dinheiro em sua juventude?

- a) trabalhou como costureira;
- b) trabalhou como faxineira;
- c) trabalhou como professora; e
- d) trabalhou em uma loja.

Além de ser um item com pergunta envolvendo uma negativa ('NÃO foi exercida'), o que o coloca na lista de itens problemáticos de Popham (1981), a única informação contextual para a resposta será *trabalhos com baixa remuneração* (low-paying jobs), que se refere a qualquer uma das atividades apresentadas como alternativas na questão de múltipla escolha. Responder corretamente a essa questão implica, assim, conhecer o significado das palavras *maid*, *tutor* e *seamstress*, e identificá-las nas alternativas *a*, *b* e *c*, o que permite, por eliminação, a escolha da resposta correta, letra *d*. *Tutor* pode ser considerada cognata, não impondo problemas; *maid* é uma palavra relativamente pouco freqüente<sup>35</sup> e de contexto específico; e *seamstress* é absolutamente pouco freqüente<sup>36</sup> e de contexto muito específico. Nesse caso, o fator relevante que permite responder corretamente à pergunta passa a ser conhecimento dos significados das palavras *maid* e *seamstress*, sendo

<sup>34</sup> Item de teste usado por professor de Inglês Instrumental, lecionando na UFSC.

<sup>35</sup> Esta palavra está na categoria de palavras com baixo índice de frequência de uso no *Collins Cobuild Dictionary*.

<sup>36</sup> O *Collins Cobuild Dictionary* não fornece informações sobre a frequência dessa palavra. Entendemos que seu uso é bem pouco freqüente e em situações muito específicas, não justificando ser seu conhecimento um fator relevante para qualquer interpretação sobre a habilidade de leitura.

evidência inadequada para qualquer interpretação de habilidade de leitura. Além disso, esse item também tem o problema técnico de confiabilidade, já que é possível argumentar que tanto uma empregada (*maid*) como uma costureira (*seamstress*) podem trabalhar em uma loja, ficando o item sem resposta correta. Como os anteriores, esse item parece ter mais argumentos desfavoráveis para seu uso na interpretação da habilidade de leitura.

É possível concluir, assim, que há itens em teste de leitura que estão, na verdade, medindo conhecimento de vocabulário. Pode haver o argumento que esse conhecimento é necessário para a leitura em inglês e que, portanto, deveria ser parte do construto. Concordamos plenamente em relação à sua contribuição para a compreensão em leitura.<sup>37</sup> Medir conhecimento de vocabulário para a interpretação da habilidade de leitura não é, porém, um procedimento adequado, já que conhecer determinadas palavras não garante a leitura, como também não conhecer determinadas palavras não impede a leitura; neste caso, o processo de inferência de significado de palavras desconhecidas pode ter uma contribuição positiva fundamental e palavras desconhecidas podem ter seu significado inferido em contexto com auxílio de outras ao redor.

Em suma, itens como esses podem não ser representativos do construto, já que parece haver argumentos que estão focando o construto alternativo de conhecimento de vocabulário, e não o relevante, ou seja, a habilidade de leitura em inglês. Eles podem, assim, representar uma ameaça à validade de construto, já que não permitem a demonstração da habilidade relevante.

## Considerações Finais

A elaboração de testes exige um processo de desenvolvimento bastante cuidadoso, que deve caracterizar seus três estágios de planejamento, operacionalização e administração. Durante essa elaboração, o escrutínio do teste deve ser feito com a investigação da validade de construto, isto é, colocar o teste à prova para analisar se seus itens permitem coletar evidências para uma interpretação sobre a habilidade relevante, que deve estar bem definida como o construto a ser usado.

---

<sup>37</sup> Para uma discussão sobre a relação de causa recíproca entre vocabulário e leitura, ver Tumolo (1999; no prelo b).

Por concordarmos com Vianna (2002) quando afirma que atividades na área de avaliação no Brasil são escassas, buscamos, com este texto, contribuir no sentido de trazer uma reflexão sobre o desenvolvimento de testes de tal forma a possibilitar o incremento das atividades de avaliação. Assim, este texto trouxe uma análise de itens que podem ser considerados defensáveis como também não defensáveis, considerando o uso para o qual foram desenvolvidos. Os itens não defensáveis foram aqui classificados nas três categorias de itens independentes do texto, itens com pistas para resposta, e itens de vocabulário, todas representando ameaças à validade de construto. A natureza da investigação mais interpretativa, com argumentação caracterizada por considerações de pontos favoráveis e/ou desfavoráveis de cada item, permite a modificação da própria análise e das conclusões sobre cada item, fazendo da investigação de validade um processo dinâmico. Essa modificação que deve ser parte integrante da análise. Cada item passa, então, a ter um histórico de investigação que informa todo o processo, possibilitando que seus elaboradores reflitam sobre o desenvolvimento de testes que permitem interpretações mais válidas sobre a habilidade de leitura.

## Referências

- ALDERSON, J. C.; CLAPHAM, C.; WALL, D. *Language test construction and evaluation*. Cambridge, UK: Cambridge University Press, 1995.
- BACHMAN, L.; PALMER, A. *Language testing in practice*. Oxford, UK: Oxford University Press, 1996.
- CHAPELLE, C. Are C-tests valid measures for L2 vocabulary research? *Second Language Research*, v. 10, n. 2, p. 157-187, 1994.
- CHAPELLE, C. Validity in language assessment. *Annual Review of Applied linguistics*, v. 19, p. 254-272, 1999.
- CLAPHAM, C. *The development of IELTS: A study of the effect of background knowledge on reading comprehension*. Cambridge, UK: Cambridge University Press, 1996.
- COLLINS Cobuild English Dictionary*. Great Britain: HarperCollins Publishers, 1995.
- GAGNÉ, E. D.; YEKOVIČ, C. W.; YEKOVIČ, F. R. *The cognitive psychology of school learning*. New York: Harper Collins College Publishers, 1993.

- HUTCHINS, J. Meaning: the frontier of informatics. *Informatics 9*. Proceedings of a conference jointly sponsored by Aslib, the Aslibi Informatics Group, and the Information Retrieval Specialist Group of the British Computer Society, King's College Cambridge, p. 151-173, 1987.
- KANE, M. T. An argument-based approach to validity. *Psychological Bulletin*, v. 112, n. 3, p. 527-535, 1992.
- KINTSCH, W.; VAN DIJK, T. A. Toward a model of text comprehension and production. *Psychology Review*, v. 85, n. 5, p. 363-394, 1978.
- MESSICK, S. Validity. In: LINN, R. L. (Ed.). *Educational measurement*. Washington, DC: The American Council on Education and the National Council on Measurement in Education, 1989.
- NUTTALL, C. *Teaching reading skills in a foreign language* (new edition). Oxford, UK: Macmillan Education, 1996.
- PEARSON, P. D.; JOHNSON, D. D. *Teaching reading comprehension*. New York: Holt, Rinehart and Winston, 1978.
- PIERCE, B. N. Demystifying the TOEFL reading test. *TESOL Quarterly*, v. 26, n. 4, p. 665-689, 1992.
- POPHAM, W. J. *Modern educational measurement*. Englewood Cliffs, NJ: Prentice Hall, 1981.
- TOMITCH, L. M. B. Designing reading tasks to foster critical thinking. *Revista Ilha do Desterro*, n. 38, p. 83-90, 2000.
- TOMITCH, L. M. B. Por que o aprendiz de leitura em língua estrangeira precisa de professor? O papel do professor no ensino de leitura em LE. In: COSTA, M. J. D.; ZANATTA, M. E.; ZISPER, M. E.; MENDES, A. *Línguas: ensino e ações*. Florianópolis, SC: NUSPPLE, 2002.
- TOMITCH, L. M. B. *Reading: text organization perception and working memory capacity*. Florianópolis, SC: PGI/UFSC, 2003. (Livro da série ARES – Advanced Research in English Series)
- TUMOLO, C. H. S. *Vocabulary Instruction: the text as a source in the classroom*. Dissertação (Mestrado em Letras/Inglês) – CCE – UFSC, Florianópolis, SC, 1999.
- TUMOLO, C. H. S. *Assessment of reading in English as a foreign language: investigating the defensibility of test items*. Tese (Doutorado em Letras/Inglês) – CCE – UFSC, Florianópolis, SC, 2005.
- TUMOLO, C. H. S. Especificações e pré-testagem: etapas essenciais na elaboração de testes de leitura em língua estrangeira. In: TOMITCH, L. M. (Org.). *Aspectos cognitivos e instrucionais da leitura em língua estrangeira*. No prelo a.

TUMOLO, C. H. S. Vocabulary and reading: teaching procedures in the ESP classroom. *Revista Linguagem & Ensino*. No prelo b.

VAN DIJK, T. A.; KINTSCH, W. *Strategies of discourse comprehension*. New York: Academic Press, 1983.

VIANNA, H. M. Questões de avaliação educacional. In: FREITAS, L. C. (Org.). *Avaliação: construindo o campo e a crítica*. Florianópolis, SC: Editora Insular, 2002.

WEIR, C., HUIZHONG, Y.; YAN, J. *An empirical investigation of the componentiality of L2 reading in English for academic purposes*. Cambridge, UK: Cambridge University Press, 2000.

WEIR, C.; MILANOVIC, M. *Continuity and innovation: Revising the Cambridge Proficiency Examination 1913-2002*. Cambridge, UK: Cambridge University Press, 2003.

## ANEXO

### Anexo A: Prova de Vestibular usada pela UNICAMP, ano 1998.

Responda a todas as perguntas EM PORTUGUÊS.

Leia o trecho abaixo e responda às questões **01**, **02** e **03**.

Day by day the Point got taller and taller. And day by day the shadow got longer and longer.

All around flowers died, grass turned brown and rooms became dark and cold. Old people had to turn on heaters, even in the middle of summer.

'It's just so ugly,' said Doll to Harold as they ate dinner one night. 'Once I used to look out of the window and see trees and flowers, hear singing birds. Now all I see is that ugly grey thing. There're no flowers, no trees, no light, no grass, no birds, nothing.'

'Oh, it's not that bad,' said Harold.

'Don't give me that,' snapped Doll. 'You don't have to watch it. Day in and day out. Watch it getting bigger and bigger and bigger.'

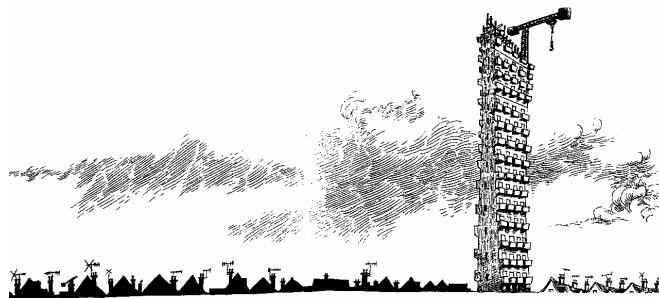
Rosie sat at the table and ate her dinner. She thought her mum was being stupid, although she didn't say so. Instead, she just filled her mouth with a forkful of mashed potato and stared at her plate.

Later, though, while Doll was washing up, Rosie couldn't help saying, 'I don't think it's ugly.' 'Well, you're as foolish as your father, then.' 'I just think it's . . . it's a gigantic finger pointing up to the sky. Or a tall flower. Or a wonderful steeple –'

'Listen, young lady,' interrupted Doll. 'It's not a finger and it's not a flower and it's not a steeple. It's just a shadow. Nothing else. It's just a point of shadow.'

And that was how the Point became known as Shadow Point.

(Philip Ridley. *Mercedes Ice*. London, Puffin Books. 1996, p. 18-19)



01. Quem é quem nessa história?

02. A que se refere "*Shadow Point*"? Por que recebeu esse nome?

03. O texto menciona mudanças. Que mudanças são essas?