

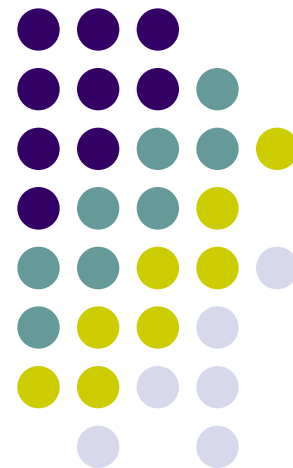
Exames de proficiência em LE em contextos profissionais



Matilde V. R. Scaramucci
Departamento de Linguística Aplicada
Unicamp

matilde@unicamp.br

Seminário do GEIA – Grupo de Estudos em Inglês Aeronáutico
26 de maio de 2015 - ICEA (Instituto de Controle do Espaço Aéreo) -
Campus do DCTA - São José dos Campos, SP



objetivo



- a partir da revisão de alguns conceitos/noções da área de avaliação, examinar questões importantes para o desenvolvimento de exames/testes de proficiência em LE em contextos profissionais (propósitos específicos)

por que testes com propósitos específicos?

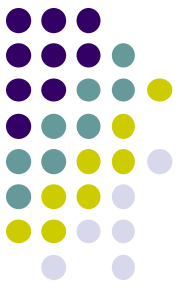


- desempenhos na língua variam de acordo com o contexto (*TLU situation* – situação de uso da língua alvo)
- língua com propósitos específicos é precisa – o que normalmente chamamos de jargão (críticos dizem que ESP é apenas jargão) significa precisão – e tem uma função comunicativa
- não podemos falar de ensino de línguas com propósitos específicos se não pensarmos em avaliação de línguas com propósitos específicos, dada a relação simbiótica existente entre ensino e avaliação; não poderíamos ensinar e avaliar conteúdos distintos

por que testes com propósitos específicos?

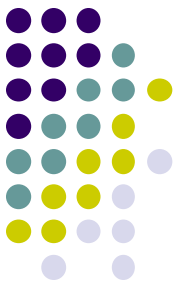


- como salienta Douglas (2000), testes são desenvolvidos sempre com algum propósito, e portanto, podemos pensar em um continuum de especificidade:
- geral específico
altamente específico
- avaliação de língua com propósitos específicos é, por natureza, comunicativa, uma vez que é baseada na capacidade de usar a língua de forma comunicativa e contextualizada; o que Bachman denomina de *communicative language ability*



teste de desempenho

- testes/exames de língua com propósitos específicos já nasceram comunicativos: pularam a etapa dos itens isolados, na medida em que envolvem o contexto de uso específico que influencia a situação de comunicação, ou seja, as características do contexto físico e temporal, o papel ou papéis dos examinandos, os propósitos da comunicação, o tópico e o conteúdo da mensagem, entre outros
- foram esses contextos de uso específicos que motivaram a elaboração de testes de desempenho



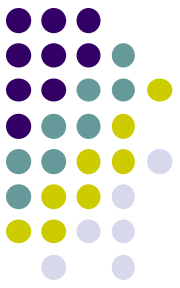
teste de desempenho

- avalia a capacidade dos candidatos de desempenhar determinadas tarefas, geralmente associadas com requisitos de trabalho ou de estudo (Davies et al., 1999)
- tarefas são elaboradas de forma a permitir que os candidatos demonstrem os tipos de habilidades/competências linguísticas que são necessárias nos contextos reais. Em outras palavras, que possam mobilizar, nos candidatos, aqueles conhecimentos/habilidades/competências em situações reais de comunicação

exame e construto



- não podemos falar em exames sem antes deixar claro que todo exame avalia um **construto**, que deve estar muito bem definido para os elaboradores e bem explicitado para todos os *stake-holders* envolvidos
- o caso dos **exames de proficiência para contextos profissionais** não é uma exceção. O que ocorre, nessa situação, é que o construto é ainda mais complexo, na medida em que os conhecimentos específicos do contexto profissional interagem com as capacidades linguísticas



critério e construto

- avaliação é um processo inferencial (McNamara, 2008); os exames e as avaliações em geral são baseados em inferências sobre um determinado **critério** (*TLU situation*): conjunto de comportamentos que se deseja avaliar. Esses comportamentos não são observáveis. A única maneira de torná-los observáveis é caracterizá-los para que possam ser simulados ou **representados**, sempre de forma amostral, na elaboração do instrumento (**construto**)

critério e construto



- os dados de desempenho observados a partir da aplicação do teste serão usados para fazermos inferências sobre o **critério**, permitindo observar o que antes não era observável (McNamara, 1996)
- é importante, portanto, uma distinção clara entre o **critério**, ou seja, o comportamento comunicativo na situação alvo que se quer avaliar (linguagem para comunicação entre pilotos e controladores, por exemplo) e o teste ou instrumento para avaliá-lo (**construto**)

teste e representação

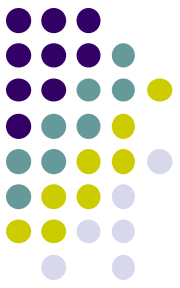


- testes, portanto, envolvem **representações**, que operam em vários níveis. No caso de exames de língua com propósitos específicos:
- **critério** - linguagem de cada contexto específico (TLU), que é caracterizada de acordo com a visão de língua(gem), valores, crenças dos envolvidos na elaboração do exame; ponto importante e complexo; mesma situação vista de maneiras distintas
- **construto** (representa essa linguagem no exame)
- **tarefas** (representam esse construto)
- **grades de avaliação** (representam esse construto)
- **resultados** (representam a capacidade avaliada, permitem inferências válidas sobre essa capacidade)

inferências



- a capacidade de fazer inferências (nunca certas) é constitutiva da avaliação. É crucial nos testes de propósitos específicos: fazemos julgamentos e inferências sobre o conhecimento/capacidades dos avaliandos com base no seu desempenho nos testes e chegamos a conclusões sobre o que sabem ou conseguem fazer para tomar decisões
- em exames de propósitos específicos, por mais que tentemos **representar** nas tarefas a complexidade da situação da língua alvo (critério), essa representação sempre será incompleta, porque a linguagem utilizada é complexa, imprevisível e a cobertura e a amostragem sempre serão inadequadas e incompletas (Douglas (1991), o que dificulta nossas inferências sobre os desempenhos



inferências

- que tipos de inferências queremos fazer? Sobre a capacidade linguística do avaliando? Ou também sobre seus conhecimentos de conteúdo da área específica de atuação (*background knowledge*)?
- necessário saber se o examinando não desempenhou bem a tarefa porque não tem as capacidades linguísticas necessárias ou porque lhe falta conhecimentos específicos da área de atuação? Os dois tipos de conhecimentos estão entrelaçados e interagem entre si

tipos de inferência

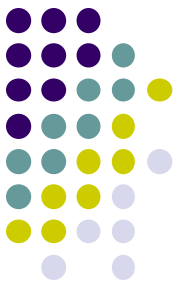


- McNamara (1996) salienta dois tipos de inferência: 1. sobre a capacidade de desempenhar tarefas futuras no trabalho (hipótese forte); e 2. sobre a capacidade de usar a língua em tarefas futuras no trabalho (hipótese fraca). A distinção é sutil, mas muito importante. McNamara alerta para dificuldades com a primeira, uma vez que temos um número grande de variáveis que podem intervir (de personalidade, por exemplo) e que são independentes das capacidades linguísticas. O autor prefere a segunda.
- Douglas (2000) reconhece a impossibilidade de se ter uma amostra adequada de tarefas ou listar todos os eventos possíveis com os quais os usuários tem que lidar na situação alvo, e concorda com McNamara que as inferências que devemos fazer devem ser sobre as capacidades de uso da língua e não sobre o desempenho no trabalho

tipos de inferência



- por outro lado, entretanto, Douglas (2000) reconhece a importância de se fazer inferências sobre a capacidade de usar a língua com propósitos específicos (construto definido com base na interação entre conhecimentos/capacidades linguísticas e *background knowledge*)
- nesse aspecto, portanto, discorda de McNamara, para quem esses conhecimentos não estariam relacionados às capacidades linguísticas. Para Douglas, se o contexto de uso restringe as possibilidades das escolhas linguísticas, o *background knowledge* (que é parte desse contexto) também deveria ser parte desse construto



duas visões de construto

- na versão fraca, o construto é definido como capacidade linguística. Na versão forte, inclui tudo que pode contribuir para realizar bem a tarefa, o que envolve um conjunto de fatores que podem confundir, tais como as características das tarefas e as interações do examinando com essas características, afetando, assim, a interpretação e uso dos do scores

competência estratégica



- capacidade linguística em exames de propósitos específicos resulta da interação entre conhecimentos de propósito específico (*background content knowledge*) e capacidades linguísticas
- a competência estratégica serve como um mediador entre o contexto situacional externo e o conhecimento background e linguístico interno exigido para dar conta da situação comunicativa, e seu engajamento nas tarefas é central para a avaliação nesse contexto
- dependendo dos propósitos da avaliação, pode ser que não seja necessário “medir” a competência estratégica separadamente, podemos assumi-la como parte do desempenho e, nesse caso, não vai receber um score separado

características importantes dos testes de propósitos específicos



- autenticidade das tarefas e interação entre conhecimentos linguísticos e conhecimento do conteúdo que tem propósitos específicos - tarefas nesses testes devem compartilhar características críticas das tarefas da situação de uso da língua de interesse dos examinandos – *background knowledge* é uma parte integral do conceito de capacidade linguística de propósito específico

características importantes dos testes de propósitos específicos



- a avaliação deve ser autêntica no sentido de ser capaz de engagar o examinando em tarefas que sejam autenticamente representativas desse contexto, que sejam capaz de mobilizar capacidades que se deseja avaliar da situação alvo, que pressupõe interação entre a proficiência na língua e conhecimento da área
- daí a importância de uma análise da situação de uso da língua alvo, de onde as características do conteúdo do teste e do método de avaliar são derivadas, assim como a interação entre conhecimento da língua e conhecimentos do conteúdo com propósitos específicos

características importantes dos testes de propósitos específicos



- autenticidade é um conceito chave nesse tipo de testes, mas definido não como qualidade dos exemplos de língua, mas na interação entre o usuário e o texto
- posso usar um conjunto de instruções elaborados para a condução de um exercício de laboratório de química (texto autêntico) como veículo para avaliar vocabulário ou uso de imperativos, o que não é um uso autêntico ou intencionado pelo autor do manual (Douglas, 2000)
- daí o termo “genuíno” proposto por Widdowson, que poderia ser usado para o conjunto de instruções, sendo “autêntico” usado apenas quando o uso é aquele pretendido pelo autor

características importantes dos testes de propósitos específicos



- Bachman (1991) explicita dois aspectos da autenticidade: situacional e interacional, sendo que o primeiro envolve a análise das tarefas em uma situação de uso e suas características, e o segundo é relacionado ao que Widdowson propõe, e tem a ver com a interação da capacidade de língua específica do examinando com a tarefa do teste
- segundo Douglas (2000), os dois tipos devem estar presentes em testes com propósitos específicos. Não adianta a tarefa ter todas as características da situação alvo se não engajar significativamente o examinando no uso comunicativo da língua

características importantes dos testes de propósitos específicos



- portanto, a construção do teste deve ser um esforço conjunto com especialistas da área, para descrever as situações em que as pessoas conduzem seu trabalho, e quais as características da língua que usam e as tarefas que desempenham na língua alvo

características importantes dos testes de propósitos específicos



- daí poderemos falar em um contínuo de especificidade – dependendo do grau em que o teste requer o engajamento de conhecimento específico na resposta às tarefas e o grau em que permite generalizações sobre uso da língua em situações específicas. Enquanto o IELTS ficaria em um extremo do contínuo, poderíamos dizer que o EPLIS e outros exames no contexto da aviação ficariam em outro extremo – altamente específico

situações de alta especificidade



- quando a situação de uso da língua alvo (*TLU situation* ou critério) é altamente específica, há maior chance de os examinandos se saírem melhor nos testes se estes forem de sua área específica. Isso significa que quanto mais se aumenta a especificidade, maior é o efeito do *background knowledge* nos scores, mostrando que apenas a proficiência na língua é insuficiente
- entretanto, parece haver uma relação não linear entre a especificidade do insumo e a capacidade comunicativa dos examinandos: uma espécie de limiar de proficiência, abaixo do qual os examinandos são incapazes de fazer uso efetivo do conteúdo específico (*background knowledge*) e acima do qual a falta desse conhecimento pode ser compensada para que os examinandos façam completo uso de seus recursos linguísticos (Douglas, 2000)

características importantes dos testes de propósitos específicos



- não é fácil chegar a uma conclusão sobre o que é que representa a TLU situação. Em primeiro lugar porque teste é sempre uma amostra, e temos que decidir o que é representativo da situação alvo, o que pesa mais, em que proporção
- o produto são as especificações do teste, definidas como “a blueprint of the test we want to develop, a statement of purpose” o que envolve uma descrição do que o teste pretende avaliar, uma descrição dos contextos e tarefas a serem incluídas, detalhes de correção e indicação de como os scores são interpretados

características importantes dos testes de propósitos específicos



- capacidade de generalizar de uma tarefa em um teste para tarefas da vida real – temos que pensar nas características das tarefas de vida real para que elas sejam contempladas nas tarefas dos testes de forma que ao avaliarmos não estamos apenas avaliando tarefas em um teste ou produzindo um teste que é interpretável apenas como teste. Temos que ser capazes de fazer generalizações a partir desse teste – anotar as características da situação e compará-las com aquelas dos testes

características importantes dos testes de propósitos específicos



- a possibilidade de generalização é um aspecto muito importante, porque essa capacidade específica deve ser vista na perspectiva dos usuários ou examinandos ou dos participantes nas atividades no contexto específico, daí a importância do que temos chamado de *indigenous assessment criteria* (critérios autóctones), definidos como “*those used by the subject specialists in assessing the communicative performances of apprentices in academic and vocational fields*” (Douglas, 2000:68), baseados em análises dos discursos em que são empregados – análise das interações e do discurso das situações de avaliação
- esse conceito (Jacoby, 1998) varia de uma situação para outra. Há, entretanto, problemas em sua aplicação: quanto mais os critérios

características importantes dos testes de propósitos específicos



- são definidos localmente, com base nas tarefas específicas das situações de propósitos específicos, mais dificuldade temos de generalizá-los ou de serem úteis para além daquelas situações
- portanto, temos que julgar esses critérios com base na definição do construto que queremos avaliar, para chegarmos a um conjunto de critérios um pouco mais generalizáveis, que reflitam a característica da situação de propósito específico na língua alvo e as restrições colocadas nos aspectos da capacidade que serão avaliados

características importantes dos testes de propósitos específicos



- não apenas os critérios devem ser baseados na situação específica, mas devem estar refletidos nas instruções para os avaliadores/corretores: não apenas características do construto refletidas nessas diretrizes (validade das instruções), mas a garantia de que esses critérios serão aplicados com consistência (confiabilidade na aplicação das instruções)

características importantes dos testes de propósitos específicos



- avaliadores/corretores podem não ter familiaridade com a situação específica e conhecimento técnico e podem ter dificuldade em julgar o desempenho. Pode ser que a linguagem produzida seja tão técnica que os avaliadores não consigam capturar a essência ou seguir a organização do discurso e encontrem dificuldades em fazer o julgamento das características que tem que avaliar: vocabulário, etc. Daí a ajuda do especialista também assistindo os avaliadores no momento de fazer julgamentos

características importantes dos testes de propósitos específicos



- os testes de propósitos específicos são referenciados em critérios (CR tests), na medida em que temos que definir precisamente o nível de capacidade ou o domínio do conteúdo que servirá de critério para o desempenho. Não apenas a definição clara das características da situação de uso da língua alvo, mas os critérios para avaliar o desempenho

dilemas e debates



- embora para Douglas (2000) autenticidade seja crucial na avaliação de propósitos específicos -- o que realmente a diferencia de outras situações mais gerais de avaliação
- a questão da autenticidade não é simples. A seleção de tarefas em um teste como representativas de tarefas autênticas no mundo real tem sido um grande debate (quem deve decidir que tarefas selecionar e baseados em que critérios?), uma vez que impactam a relevância do conteúdo e a cobertura do teste



dilemas e debates

- Bachman (2002) é cético quanto ao uso de tarefas para fazer previsões sobre desempenho na vida real por serem extremamente complexas e diversas, envolvendo grande variação e sujeitas a um grande número de fatores
- Wu and Stansfield (2001) argumentam que a autenticidade das tarefas precisam ser verificadas, e para isso são necessários comentários e críticas daqueles que realizam as tarefas no mundo real

dilemas e debates



- autenticidade se relaciona à possibilidade de generalizar o resultado. Na versão fraca, o critério de realizar a tarefa não é um problema, porque a preocupação é com as capacidades linguísticas. Na versão forte, entretanto, a realização da tarefa é fundamental: quem deve julgar se o examinando cumpriu adequadamente a tarefa? Os especialistas da linguagem ou os especialistas da área?
- alguns estudos examinaram a questão (por exemplo, Brown, 1995; Elder, 1993; Elder and Brown, 1997), que envolve um equilíbrio entre autenticidade e possibilidade de generalização
- temos aqui um dilema: enquanto a visão fraca nos permite avaliar capacidades linguísticas que são relativamente generalizáveis, a versão forte produz julgamentos que são mais autênticos e relevantes para as situações de vida real, mas que não são replicáveis em outros contextos

concluindo...



- a **specific purpose language test** is one in which test content and methods are derived from an analysis of a specific purpose target language use situation, so that test tasks and content are authentically representative of tasks in the target situation, allowing for an interaction between the test taker's language ability and specific content knowledge, on the one hand, and the test tasks on the other. Such a test allows us to make **inferences** about a test taker's capacity to use language in the **specific purpose domain**



Obrigada!