



# Avaliação de proficiência: valores culturais e consequências sociais

- **Matilde V. R. Scaramucci**  
**Universidade Estadual de Campinas**  
[matilde@unicamp.br](mailto:matilde@unicamp.br)



**V Colóquio CELU – Córdoba**  
**21 e 22 de agosto de 2009**

# Introdução



- As avaliações em geral e, especificamente, as de proficiência, que têm o CELU e o Celpe-Bras (*Certificado de Proficiência em Língua Portuguesa para Estrangeiros*) como exemplos, não são instrumentos “neutros” ou meramente “técnicos”, como muitos querem crer; são, antes de tudo, políticos e ideológicos, na medida em que têm, subjacentes, os valores culturais e crenças de seus idealizadores ou daqueles responsáveis pelas políticas implementadas a partir desses exames.
- Quando menciono valores culturais, não estou me referindo à cultura que é parte da língua e que é avaliada em muitos exames. Como exemplos de exames culturalmente marcados, podemos citar aqueles utilizados por serviços de imigração de muitos países como barreiras para concessão de cidadania.

# Introdução



- Os valores culturais subjacentes aos exames são determinantes nas escolhas feitas em relação à orientação teórica do exame (o que é considerado saber uma língua ou visão de proficiência, o que é língua(gem) ou concepção de língua(gem), e o que é avaliar uma língua, estrangeira ou segunda, ou concepção de avaliação).
- São ainda determinantes na escolha das habilidades contempladas no exame (compreensão oral, produção oral, etc), do peso dado a cada dessas habilidades, da forma de avaliá-las (separadamente ou de forma integrada), no peso dado a cada um dos critérios de avaliação, e assim por diante.
- Essas escolhas irão, portanto, determinar o perfil dos candidatos que devem “passar” (os incluídos) e os que devem ser “reprovados” (os excluídos), no caso de exames de seleção. Ou o perfil dos que devem ser classificados nos altos níveis de proficiência e os perfis a serem classificados nos níveis baixos, no caso de exames de proficiência.



# Introdução

- Essas escolhas, por sua vez, terão maiores chances de serem avaliadas como “positivas” por aqueles que compartilham desses mesmos valores, e “negativas” por aqueles que têm valores distintos.
- O que quero dizer é que a cultura de avaliar do exame pode ser compatível com a dos avaliados e dos professores que preparam esses candidatos, mas também pode estar em conflito, principalmente quando se trata de exames de LE, implementados em contextos variados, cada um com sua cultura de ensinar/aprender/avaliar.

# Introdução



- Outras vezes, os construtos efetivamente avaliados estão implícitos, fugindo do controle até mesmo dos elaboradores; nesse caso, os construtos avaliados não são os explicitados e ostensivamente mostrados, mas aqueles controlados pelos elaboradores das políticas que o exame visa implementar.
- Como exames são mecanismos potenciais de mudanças e de ação política (mais eficientes do que as propostas curriculares!), as consequências que terão na sociedade, portanto, também são avaliadas segundo esse mesmo mecanismo: o que para alguns é considerada uma consequência “positiva”, para outros é “negativa”.

# Introdução



- A consideração do que seria uma consequência positiva (ou negativa) é dependente, portanto, em grande parte, de quem faz a análise em um determinado contexto, assim como onde, quando, porque, como, etc (Cheng & Curtis, 2004) e de seus valores.
- Interpretar uma consequência como boa ou ruim envolve considerações sobre os objetivos da educação e de seus valores subjacentes. O que se considera adequado ou justo em um contexto ou ainda educacionalmente benéfico pode não o ser para alguém com um conjunto diferente de valores.

Cheng, L. & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In Cheng, L. & Watanabe, Y. *Washback in language testing*. Lawrence Earlbaum Associates, Publishers. Mahwah, New Jersey.



## Objetivo desta fala

- Mostrar em que medida o fato de os exames terem subjacentes valores culturais criam dificuldades para a coleta de evidências sobre as consequências que as avaliações têm no ensino/aprendizagem em especial e também na sociedade em geral.
- A partir de uma perspectiva mais construtiva e até mesmo “otimista”, refletir sobre o que elaboradores e usuários de exames podem fazer para coletar evidências sobre os efeitos de seus exames e minimizar suas consequências adversas, tornando-os instrumentos socialmente e educacionalmente mais benéficos.



# Validade

- Fundamental quando falamos de avaliações/exames é o conceito de validade, vista hoje não mais como uma propriedade do exame ou teste (“teste válido” ou “não válido”), mas como o julgamento da medida em que as evidências empíricas e as justificativas teóricas apoiam a adequação das inferências e das ações baseadas nos resultados das avaliações.





# Validade

- Resultados de testes ou exames são usados para fazermos inferências ou interpretações sobre o desempenho daqueles que a eles se submetem.
- Essas inferências servirão de base para as nossas decisões (aprová-los ou reprová-los, classificá-los em um determinado nível, dentre outros, dependendo da natureza do teste).
- Não podemos, pois, fazer inferências sobre os resultados de um exame elaborado para um determinado fim sem antes justificá-las, ou sem fornecer explicações e evidências que justifiquem esse uso, ou seja, sem antes validá-las.

# Validade, valores culturais e consequências



- Na visão contemporânea de validade, a adequação, o significado e a utilidade das interpretações baseadas nos resultados dos testes também dependem das consequências sociais dessa avaliação.
- Para Messick (1989), valores e consequências sociais não podem ser ignorados em considerações de validade. Para esse autor, um teste que não cumpre com seu objetivo proposto (consequências intencionais) ou que apresenta consequências adversas (não intencionais) não pode ser considerado um teste válido, ou melhor dizendo, um teste em que as inferências não são válidas.

# Validade, valores culturais e consequências



- Portanto, para que se possam fazer inferências válidas sobre os resultados de um exame, este teria que levar em conta as consequências que tem na sociedade (intencionais e não intencionais).
- Entretanto, temos um problema sério, de difícil solução: Como julgamos as consequências se elas também encerram valores? Quem é responsável em examinar as consequências dos testes? Aqueles que o desenvolvem ou seus usuários, ou seja, aqueles que fazem a decisão política de usar um teste para um determinado propósito? Como conduzir esse processo, se os valores desses grupos podem não ser necessariamente os mesmos?
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13–103). New York: American Council on Education & Macmillan.

# Consequências



- Tema considerado extremamente controverso, com implicações sérias não apenas para os processos de validação mas também para a formação de professores e para as mudanças que queremos implementar a partir de testes (políticas linguísticas, no caso de testes de línguas).
- Enquanto a maioria dos autores hoje em dia parece concordar em relação à importância das consequências das avaliações, entretanto, não parece haver um **consenso** se uma análise dessas consequências deve ser conduzida como **parte** de um processo de validação.
- Embora desejável, a consideração das consequências parece ser vista como uma atividade extensa e complexa, que necessita da colaboração dos responsáveis pelo desenvolvimento de testes, usuários desses testes e toda a profissão envolvida com medidas.



Algumas questões terminológicas...



- Consequência
- Impacto
- Efeito
- Influência
- Efeito retroativo (*washback*) ou (*backwash*)



# Consequências

- Efeitos, na sociedade e na educação, das decisões baseadas nos resultados dos testes.
- Termo utilizado na literatura de educação e de avaliação/medidas. Incorpora impacto e efeito retroativo (“washback”).
- Propósito ou intenção (explícitos e implícitos) de um teste → desempenho (escores) → inferências baseadas nesses escores (interpretação) → decisões → consequências (intencionais) e (não intencionais ou efeitos colaterais)

# Consequências não intencionais



- Uma consequência não intencional ocorre quando agências decidem usar um teste para um propósito que não foi originalmente pensado pelos seus elaboradores e para o qual não há um argumento de validade (Fulcher 2003:197).
- Uma universidade quer restringir o acesso a vagas nas engenharias por falta de recursos. Ela decide usar o exame de colocação de inglês para excluir aproximadamente 20 por cento dos candidatos que não tenham inglês como L1.
- Esse tipo de consequência não intencional ('gate-keeping') foge geralmente do controle do elaborador do teste; para evitar que ela ocorra, pode-se colocar explicitamente no manual que determinados usos não devem ser autorizados (Fulcher 2003:197) [teríamos aqui um caso de uso equivocado ou de uma intenção implícita?].





# Impacto/efeito retroativo

- Impacto (social): termo geral, sinônimo de consequência, influência e efeito? Ou efeito mais amplo da avaliação em materiais de ensino, sistemas educacionais e sociedade em geral?
- Para Shohamy (2000) impacto = efeitos na sociedade, envolvendo exclusão ('gate-keeping'), ideologia, eticalidade, moralidade e justiça ('fairness'); efeito retroativo (washback) = efeitos educacionais no currículo, métodos de ensino, estratégias de aprendizagem, materiais e práticas de avaliação.
- Alderson & Wall (1996), por sua vez, distingue efeito retroativo de impacto. Efeito retroativo refere-se somente ao que acontece na sala de aula por influência dos testes, em geral testes de alta relevância (high-stakes). Os efeitos educacionais mais amplos, no currículo e materiais são considerados "impacto".



# Efeito retroativo

- ('washback' ou 'backwash') é usado em contextos de educação linguística, avaliação de línguas e lingüística aplicada.
- 'Backwash' é mais utilizado em educação na tradição inglesa.
- Entretanto, para Andrews (2004), não há distinção entre os dois termos: efeito retroativo é interpretado amplamente, referindo-se aos efeitos dos testes no ensino e aprendizagem, no sistema educacional, e nos vários protagonistas do processo educacional. Impacto seria o termo não técnico, sinônimo de efeito.

# A pesquisa sobre efeito retroativo



- Para Taylor (2005), efeito retroativo pode ser considerado um aspecto de impacto, sendo o primeiro uma efeito em nível micro e o segundo em nível macro (Weir & Milanovic 2003). Talvez uma melhor distinção possa ser feita em termos de efeito direto (efeito retroativo) e indireto (impacto).
- As questões relativas aos impactos ou efeitos retroativos potenciais que os exames exercem no ensino e na sociedade tornaram-se um foco de pesquisa importante a partir dos anos 90, com a publicação do artigo seminal de Alderson & Wall intitulado *Does washback exist?*
- Até então: conceito popular mas mal compreendido; falta de evidências empíricas que comprovassem sua existência e explicasse seu mecanismo de ação; visão determinista: bom exame bom efeito, mau exame mau efeito.

# Efeito retroativo e validade



- A partir dos anos 90, em inúmeras partes do mundo: estudos sobre o conceito para elaboração de uma teoria.
- Hoje: podemos dizer que temos uma teoria sobre efeito retroativo. Reconhecimento da complexidade do fenômeno; ação não determinista, na medida em que aspectos como personalidade, motivação, cultura de aprender/ensinar, status da língua avaliada e a formação dos professores do contexto de implementação interagem com o exame na determinação de seu efeito.
- Pesquisas sobre efeito retroativo ocorreram, entretanto, de forma independente das reflexões que vinham sendo feitas na área de educação e de medidas educacionais, que passaram a considerar um conceito expandido de validade, em que as consequências são um aspecto importante, como mostrei antes.

# Por que conseqüências?



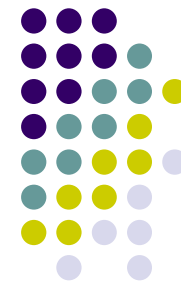
- Foi apenas no final dos anos 80s e 90s que a literatura mostrou sinais reais de preocupação com o conceito.
- Popularidade das chamadas avaliações alternativas ou autênticas, uma vez que muitos benefícios e decisões de alta relevância começaram a ser associados a essas avaliações: conduzir reformas na educação, minimizar preparação específica ('coaching'), desenvolver habilidades de alto nível (inferências, levantamento de hipóteses), permitir entrada na educação superior, oportunidades futuras de emprego, responsabilização do professor, etc (Madaus, Raczek and Thomas 1999).



# Por que consequências?

- Testes começam a extrapolar seu papel tradicional de monitoramento e controle para serem mecanismos de mudança, de reforma ou de responsabilização ou “accountability” (altos escores são recompensados e baixos escores são punidos, independentemente de refletirem aprendizagem ou não).
- Como a intenção é melhorar (ou reformar), é razoável que esses testes sejam avaliados como programas educacionais. Programas educacionais sempre incluíram uma avaliação dos resultados intencionais e dos não intencionais (Kane 2002; Kane, 2006).

# Exemplos de consequências



- O exame Celpe-Bras foi elaborado não apenas para certificar a proficiência geral de candidatos que precisam estudar ou trabalhar no Brasil ou no exterior, usando a língua portuguesa, mas também para (re) orientar ou (re) direcionar o ensino e, conseqüentemente, a aprendizagem do português tanto no Brasil como no exterior.
- Portanto, ele deverá ser avaliado para sabermos se cumpre também com esse papel.



- Essa intenção também está subjacente ao exame CELU? Se essa for uma intenção, ela deverá ser avaliada para sabermos se realmente essa “consequência intencional” está ocorrendo, assim como outras não intencionais, geralmente adversas.



# Exemplos de consequências



- O número de alunos que fazem aulas de música, arte, educação física e outros cursos e atividades extracurriculares não avaliadas no teste diminuíram?
- Os alunos estão passando mais tempo e concentrando seus esforços em atividades de preparação específica para o teste ['coaching']?
- O ensino das habilidades orais foi incluído em um contexto em que um exame de habilidades orais foi implementado?
- O controlador de voo, por não ter passado no exame de inglês obrigatório para sua profissão, perdeu seu emprego?
- O médico, por não ter passado no exame de proficiência obrigatório, não conseguiu revalidar seu diploma e perdeu o emprego?

# Exemplo (1) (Scaramucci, 2008)



- Exame de entrada de minha universidade (Unicamp) inclui um teste de inglês, como é de praxe no Brasil. Nesse caso, trata-se de um teste de leitura, na medida em que essa habilidade, entre todas, é aquela que tem uma função social maior na universidade pela necessidade de os alunos terem que ler bibliografia em inglês.
- Em decorrência desse exame, que é de altíssima relevância (50.000 candidatos por ano para aproximadamente 2.500 vagas) a leitura passou a ser focalizada no ensino médio.
- Enquanto esse efeito é visto como positivo por aqueles que valorizam a leitura, é visto como negativo (estreitamento do currículo) por todos os outros que valorizam as habilidades orais, não avaliadas no exame.

# Exemplo (1)



- Além das restrições práticas e do custo de avaliar as habilidades orais de pelo menos 18.000 candidatos que passam para a segunda fase do exame a cada ano, quando a prova de inglês é realizada, há uma outra razão, não explicitamente colocada, para o uso de uma prova de inglês que apenas focaliza a leitura: a avaliação de habilidades orais tornaria a seleção muito mais elitista, na medida em que alunos de classes menos favorecidas não têm a chance de estudar em escolas particulares e institutos de línguas onde habilidades orais são enfatizadas.
- Essa intenção (não explicitada) é reforçada quando observamos que inglês é a única disciplina não prioritária avaliada no exame, o que faz com que o aluno não tenha que tirar uma nota mínima para passar. Em outras palavras, somente não pode tirar nota zero, o que equivale a dar à prova de inglês um peso menor.

## Exemplo (2)



- Na Austrália, as notas dos exames de língua obtidas pelos alunos no último ano do ensino médio são utilizadas como base para decisões sobre seleção de vagas altamente competitivas na universidade.
- Elder (1997) mostrou que a habilidade comunicativa (capacidade de uso) nas línguas estrangeiras é descontada pelas autoridades de admissão se essa habilidade for resultante de uma exposição naturalista à língua (por exemplo, quando os candidatos têm algum tipo de background lingüístico, os chamados falantes de herança) e é apenas considerada relevante se for o produto de trabalho árduo e aplicação entre “verdadeiros” aprendizes de segunda língua.
- O construto subjacente ao exame, em outras palavras, não é o construto ostensivamente mostrado — proficiência lingüístico-comunicativa — mas habilidade intelectual e determinação.

## Exemplo (3)



- O Celpe-Bras fundamenta-se em um construto de proficiência definida como “ser capaz de engajar-se em diferentes atividades no mundo, mostrando adequação às demandas dos vários contextos. Em outras palavras, “o uso adequado da língua para desempenhar ações no mundo” (Brasil, 2006a e 2006b, p. 4); uma proficiência, portanto, não absoluta, mas relativa, situada nos contextos de uso, co-construída na interação oral e escrita.
- Só para lembrar: proficiência significa domínio, saber uma língua. O que é saber uma língua varia de acordo com a concepção de língua(gem) adotada.
- Concepção estruturalista: saber uma língua significa ser capaz de manipular e controlar formas gramaticais e de vocabulário, geralmente avaliadas a partir de *testes de itens isolados* (*testes de lápis e papel*) de múltipla escolha ou *respostas fixas*, em que as habilidades e os itens são avaliados isoladamente através de **testes de conhecimentos**.

## Exemplo (3)



- Concepções mais contemporâneas: saber uma língua significa ser capaz de usar a língua adequadamente. A proficiência é, portanto, co-construída socialmente e localmente por seus participantes (Clark, 1996) e geralmente avaliada diretamente a partir de **testes de desempenho**.
- A visão de proficiência do exame Celpe-Bras pressupõe letramento ('literacy') em língua estrangeira, se entendermos letramento como "estado ou condição de quem *não* só sabe ler e escrever, mas exerce as práticas sociais de leitura e de escrita que circulam na sociedade (...), conjugando-as com as práticas sociais de interação oral" (Soares, 1999, p. 3).

## Exemplo (3)



- Enquanto esse exame tem sido valorizado por muitos por avaliar letramento e distanciar-se da avaliação da forma, como muitos exames de proficiência, tem sido criticado por outros exatamente por essa característica, sob a acusação de que coloca, para os alunos estrangeiros mais exigências do que para alunos brasileiros. O letramento, nesse caso, deveria ser desenvolvido em língua materna e transferido para a língua estrangeira.
- A escolha em relação à orientação teórica do exame foi deliberada, uma ação política pois não queríamos a manutenção do status quo no ensino de português L2/LE, em que as práticas têm sido reconhecidas como tradicionais, com materiais geralmente centrados em listas de palavras, regras gramaticais e exercícios de preenchimento de lacunas, e com limitadas atividades de leitura e escrita.
- Não quisemos por exemplo, aceitar um exame elaborado por Portugal porque nos julgamos no direito de tentar determinar o rumo das mudanças que almejávamos implementar nos contextos de ensino de Português tanto no Brasil como no exterior.

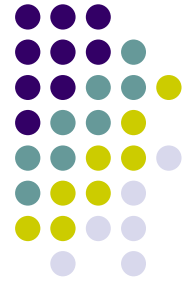
# Afinal, o que fazer?



- Por tudo isso que acabamos de constatar, é que muitos autores não considerem possível controlar os resultados ou as interpretações que as pessoas fazem dos exames na implementação de seu ensino ou os impactos que os exames têm na sociedade em geral como evidências para o estabelecimento da validade de um teste.
- Como vimos, bons testes podem ter efeitos ruins e testes ruins podem ter bons efeitos por outras coisas feitas ou não no sistema educacional e, portanto, para muitos autores (Alderson & Wall, por exemplo, e também Messick) não podem ser consideradas efeito retroativo mas resultado de práticas educacionais boas ou ruins. Nesse caso, eles não representam problemas de medida, mas de ensino e de política educacional e social.



# Afinal, o que fazer?



- Efeito retroativo é uma consequência que pode ser relacionada à validade apenas se houver evidência de ser um efeito do teste e não de outras forças operantes [qualificação do professor, por exemplo] no cenário educacional. Essas forças interagem com o teste na determinação de seu efeito (Messick 1996).
- Para esses autores, as consequências sociais adversas invalidariam o uso de um teste principalmente se forem devidas a defeitos no teste.
- São ameaças à validade: **sub-representação do construto** [não avaliação de habilidades orais em um exame eu se diz de uso, por exemplo]; ou **dificuldade irrelevante ao construto** [influência do formato da tarefa, influência da falta de acordo dos avaliadores], pois podem precipitar práticas educacionais ruins [por exemplo, coaching]; minimizar essas ameaças à validade deve facilitar boas práticas educacionais e efeito retroativo positivo ('wasback by test design').



# Limites do poder do teste

- Messick (como Alderson & Wall 1993), portanto, reconhecem os limites do poder do teste em exercer 'washback' positivo: eles podem mudar o que é ensinado (conteúdos), mas não como é ensinado (abordagem, metodologia). Embora sejam instrumentos eficientes de política linguística, ainda assim têm seus limites.
- Por isso mesmo, reconhecem os limites da responsabilidade social do elaborador, que seria apenas produzir evidência de escores válidos.
- Embora concorde com esses autores que a questão é complexa, acho que o problema dessa visão é que para a produção de escores válidos, não são consideradas as consequências, num primeiro momento.
- O fato de considerar as consequências sociais uma ameaça à validade apenas quando são devidas a fontes de invalidade é dar a elas um papel secundário no processo de validação.



# Outros autores

- Para Cronbach [influenciado por seu trabalho em avaliação de programas] as decisões e ações baseadas nos escores dos testes além das interpretações descritivas, devem ser parte da validade.
- O papel das consequências são mais diretos. Consequências negativas devem invalidar o uso do teste mesmo se as consequências não puderem ser relacionadas a algum defeito no teste, porque “testes que afetam os direitos e causam mudanças nas vidas das pessoas são inerentemente questionáveis” (Cronbach, 1988, pg 6).
- Os responsáveis pela validação têm a obrigação de rever se a prática tem consequências adequadas para os indivíduos e instituições, e especialmente protegê-los contra consequências adversas (pg 6).



# Considerações finais

- Apesar dessas dificuldades, considero desejável, assim como Alderson (2004), que os elaboradores de testes considerem os prováveis impactos de seu teste no ensino/aprendizagem e na vida das pessoas em geral, e que procurem provocar um efeito “benéfico” na maneira como o teste é elaborado, mesmo que seja em relação ao que denominam ‘washback by test design’ (Messick, 1996).
- Entretanto, também admito, seguindo Alderson (2004) que há limites em relação ao que o elaborador pode alcançar em termos das evidências para o processo de validação.

# Considerações finais



- Por outro lado, essas dificuldades não devem ser paralisantes ou desmobilizadoras. Ainda há o que fazer.
- Como elaboradores e também como usuários dos resultados dos exames (agências, por exemplo, Conselho Federal de Medicina, no caso do Celpe-Bras, universidades que aceitam os candidatos aprovados) temos que estar conscientes dos valores culturais que estão implícitos nos construtos dos testes para ter certeza de que são realmente esses valores que queremos ou se estão nos sendo impostos externamente. Mesmo que para alguns autores, esse controle seja impossível (vide McNamara & Roeber, 2006).
- Temos que teorizar mais sobre as condições do contexto político e social em que os testes estão inseridos e implementados vis a vis as características do exame e as consequências pretendidas, e explorar, com mais profundidade, a ligação entre impacto pretendido e consequências reais (dados de 'baseline' são importantes).



# Considerações finais

- Nesse contexto, muito mais atenção deve-se dar às razões pelas quais os professores ensinam da maneira que ensinam, suas crenças sobre ensino e aprendizagem, seu grau de profissionalismo, a adequação da sua formação e da sua compreensão em relação à natureza e fundamentação do exame para que possamos saber realmente aquilo que fariam independentemente do teste e o que estão fazendo por influência do teste (Alderson, 2004).
- Temos que procurar estar cientes dos valores culturais dos candidatos para melhor entender seu comportamento assim como suas expectativas, a fim de encontrar maneiras de potencializar as chances dos efeitos desejados (por exemplo, observar o comportamento em relação às instruções, como são seguidas e interpretadas, como as tarefas são respondidas pelos candidatos, entre outros).



# Considerações finais

- Essas visões podem ser conflitantes, como disse antes, e o exame pode ter consequências danosas. Isso não significa, por outro lado, que temos que aceitar os valores culturais do contexto de implementação, mas estar preparados para lidar com eles.
- Estar cientes de que as mudanças (se ocorrerem) podem demorar anos (efeitos de longo prazo). Há, por outro lado, evidências que mostram que depois de muitos anos (10, por exemplo) os efeitos potenciais dos testes são enfraquecidos, uma vez que as pessoas envolvidas já internalizaram a interpretação que fizeram do teste e de seus fundamentos e parecem não se importarem mais com ele. Daí a importância de se rever as estruturas dos exames de tempos em tempos.

# Considerações finais



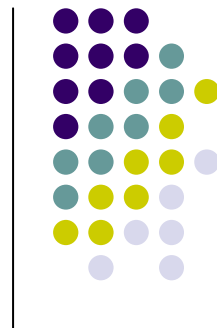
- Temos que nos atentar para o fato de que se exames inovadores são inseridos concomitantemente a mudanças nas diretrizes curriculares do estado, por exemplo. Quanto dessa mudança pode ser atribuída a cada variável? (Mehrens, 2002). Isso ocorre muitas vezes em relação a materiais didáticos, muitas vezes implementados juntamente com novos exames.
- Estar cientes de que todo exame leva a práticas simplificadoras e tentar minimizá-las na forma como elaboramos os instrumentos. Essas práticas são decorrentes da interpretação que professores e alunos fazem dos exames, segundo seus valores e crenças (“coaching” ou treinamento específico a partir das próprias tarefas do exame, por exemplo). Nesse caso, o que observamos é que esse treinamento ocorre independentemente das características do exame, ou seja, faz parte das crenças. Explicitar essa crença e seus efeitos maléficos seria uma forma de conscientizá-los sobre a necessidade de usar mecanismos mais produtivos.





## Considerações finais

- A coleta de dados sobre consequências deveria envolver uma triangulação de métodos qualitativos e quantitativos de fontes e perspectivas dos múltiplos protagonistas ('stakeholders'): alunos, pais, professores, aplicadores, responsáveis pelas políticas, e a comunidade em geral, juntamente com a observação de aulas para que as várias visões sobre as consequências sejam contempladas.



Obrigada!