EC330
Tratamento da Informação

As quatro faces de Messick
ou
como a validação pode se tornar complexa!

Prof. Dr. Maurício U. Kleinke

PECIM – PPG em Ensino de Ciências e Matemática

Grupo de Ensino e Avaliação

DFA/IFGW/Unicamp

Textos para discussão
Escola Nacional de Ciências Estatísticas
Número 25

Da Teoria Clássica de Testes Para a Teoria da Resposta ao Item

Philip Ralph Fletcher

UNICAMP

IFGW
Instituto de Física Gleb Wataghin

Generalização da validade

- A questão fundamental, que fatalmente surge no caminho da validação, é se o significado da medida é relevante apenas num determinado contexto específico ou se, de alguma forma ou outra, ela pode ser generalizada para outros contextos.
- Embora existam diversas maneiras de acumular evidências para sustentar inferências neste sentido, seus meios são essencialmente os procedimentos da pesquisa científica.
- Procura-se estabelecer se a medida demonstra as mesmas propriedades invariantes e configurações de relacionamento observadas em diferentes grupos populacionais, sob uma grande diversidade de condições objetivas

Messick e a validação

- *Como é que a pessoa se desempenha num universo de situações?* Examina-se o conteúdo da prova em relação ao conteúdo do domínio de referência.
- *Até que ponto a prova mede uma consistência observada?* Verifica-se como as pessoas respondem aos itens ou tarefas, examinando as relações entre respostas para evidência de sua consistência *interna*.
- *Como é que a pessoa se posiciona em relação a alguma outra variável diferente da prova?* Examina-se a relação entre os resultados de prova e outras medidas de características ou de experiência para evidência de estrutura *externa*.

Messick e a validação

- *Quão generalizáveis, estáveis ou sensíveis são os resultados da prova?* Procuram-se diferenças na estrutura dos resultados ao longo do tempo, entre grupos e situações e como resposta a intervenções experimentais que manipulam o currículo, os procedimentos didáticos ou terapêuticos.

Messick e a validação

- *Será que o uso da prova provoca conseqüências sociais ou levanta questões tratadas em políticas públicas?* Examinam-se as conseqüências sociais da interpretação e uso dos resultados de prova para determinados fins, considerando não apenas seus efeitos previsíveis mas também seus efeitos colaterais não intencionados, especialmente seu viés, abuso, impacto adverso ou os efeitos sociais mais sutis associados ao seu uso habitual e recorrente.

As quatro faces de validade de Messick

Avaliações de instrução, escolares

As Quatro Faces de Messick

- Large Scale Assessment Programas for All Students – Validity, Technical Adequacy and Implementation
  - Gerald Tindal and Thomas Haladyna
- The Relevance of Messick's Four Faces to Understanding the Validity of High-Stakes Assessments
  - Russel Gersten and Scott Baker

A medida depende do contexto?

- A estimativa é relevante apenas num determinado contexto específico ou ela pode ser generalizada para outros contextos?

- Essa medida apresenta propriedades invariantes quando observada em grupos populacionais diversos?

Como olhar a prova

- *O desempenho pessoal é função das situações escolhidas para a avaliação?* Examinam-se as relações entre o conteúdo das provas em relação às matrizes de referência.
- *Até que ponto a prova mede uma consistência observada?* Verifica-se como as pessoas respondem aos itens ou tarefas, examinando as relações entre respostas para evidência de sua consistência *interna*.
- *As variáveis externas impactam na prova?* Examina-se a relação entre os resultados da prova e outras medidas de características ou de experiência que possam sugerir ou indicar evidências de estrutura *externa*.

Como olhar a prova

- *Os resultados da prova são generalizáveis, estáveis ou sensíveis?* Procuram-se diferenças na estrutura dos resultados ao longo do tempo, entre grupos e situações e como resposta a intervenções experimentais que manipulam o currículo, os procedimentos didáticos ou terapêuticos.

Como olhar a prova

- *O uso da prova provoca consequências sociais ou levanta questões tratadas em políticas públicas?* Examinam-se as consequências sociais da interpretação e uso dos resultados de prova para determinados fins, considerando não apenas seus efeitos previsíveis mas também seus efeitos colaterais não intencionados, especialmente seu viés, abuso, impacto adverso ou os efeitos sociais mais sutis associados ao seu uso habitual e recorrente.

Bases evidenciais para a validade

- A validade para Messick é um conceito único, agregador de diferentes outros conceitos, como as clássicas validades de face, de conteúdo e de constructo.
- A validade estaria muito mais associada à finalidade última do teste do que à sua estrutura analítica.
- Atualmente, os testes costumam apresentar muitos objetivos e interesses, servindo bem a alguns e nem tanto a outros.

Bases evidenciais para a validade

- Os testes são muito mais associados a marcas esperadas de avaliações anuais do que com o objetivos de fornecer informações aos professores para que possam intervir em suas atividades na sala de aula.
- Esclarecer quais são as diferenças entre os testes é um dos aspectos fundamentais para atividade dos professores.

Variáveis	Interpretação da estimativa de qualidade	Uso dos dados da estimativa de qualidade (assessment)
Bases Evidenciais para garantir validade	• Correlações com outras medidas de avaliação • Uso dos indicadores convencionais de confiabilidade	• Dados sobre como os professores usam as estimativas da avaliação para modificar sua prática docente; dados sobre como municípios, estados ou escolas utilizam as estimativas para decisões curriculares; • Dados sobre o impacto final dos resultados no desempenho dos estudantes.

Variáveis	Interpretação da estimativa de qualidade	Uso dos dados da estimativa de qualidade (assessment)
Bases Consequenciais	• Os resultados conduzem a um refinamento das teorias de ensino e aprendizagem?	• Essas estimativas são utilizadas de fato no refinamento dos processos de ensino e aprendizagem?

Variáveis	Interpretação da estimativa de qualidade	Uso dos dados da estimativa de qualidade (assessment)
Bases Conseq.	• Os professores percebem que as informações sobre a avaliação são úteis em termos de sugerir direções para o aperfeiçoamento da instrução escolar?	• Ela permite encontrar alunos que necessitam de atenção especial, ajudam os professores a reorganizar sua visão curricular ou ainda indicar as posições sobre tópicos considerados difíceis por seus alunos?

Variáveis	Interpretação da estimativa de qualidade	Uso dos dados da estimativa de qualidade (assessment)
Bases Consequenciais	• Os professores consideram os dados dos testes úteis para compreender quando os progresso dos estudantes aumenta ou falha?	• Os efeitos colaterais não desejados do uso de um novo teste são considerados?

1ª Face
Validade como conceito unificador

- Os conceitos de validade tem sido vilipendiados pelo excesso de funções atribuídos aos testes.
- As análises dos especialistas, em geral, apenas validam a suposta relação entre os testes e as propostas curriculares das diferentes áreas.

1ª Face
Validade como conceito unificador

- Os principais problemas estão associados ao fato de que a formação dos alunos é complexa, você pode receber na escola estudantes com distintos graus de conhecimento e de habilidades. O esforço e o empenho do professor é variado, mas, em última instância, a escola é avaliada nos testes por sua clientela que a frequenta.

2ª Face - Valores e significados das estimativas de avaliação

- Não basta a perfeição técnica dos testes, seus bons índices ou todos os cuidados tomados em sua elaboração;
- É necessário compreender o que a medida final do teste significa para o estudante que realiza o teste, para o seu professor e para o diretor da escola onde ele estuda.

2ª Face - Valores e significados das estimativas de avaliação

- Um estudante com baixa avaliação em alguma disciplina em um teste externo vai apresentar que tipo de atitude?
- Irá procurar um curso de reforço ou irá abandonar o colégio?

2ª Face - Valores e significados das estimativas de avaliação

- As estimativas do teste refinam as teorias relativa ao ensino e a aprendizagem?
- Os professores percebem que as informações sobre avaliação são válidas para auxiliar a indicar direções para o aperfeiçoamento da instrução em classe?
- Os professores enxergam nos testes informações úteis para compreender porque os estudantes progridem ou fracassam em seu progresso?

3ª Face – Impacto dos exames estaduais/nacionais

- “Se a medição é uma ciência, então a pesquisa e o uso das medida é ciência (política) aplicada.” Messick

- Qual o impacto desses exames sobre os estudantes e os professores e as escolas?

3ª Face – Impacto dos exames estaduais/nacionais

- Uma mesma questão pode ser apresentada de forma contextualizada, com muita leitura e composição com outras áreas de conhecimento, ou de forma isolada.

- Qual o impacto desses distintos exames sobre a comunidade escolar?

3ª Face – Impacto dos exames estaduais/nacionais

- ***Assessments and Accountability, Robert L. Linn***
- Analisando ondas educacionais desde a segunda guerra, Linn observou que as ondas educacionais estavam, de certa forma, associadas às avalições e às medidas das estimativas educacionais.

3ª Face – Impacto dos exames estaduais/nacionais

Sua proposta para explicar essa associação seria:

a) É relativamente fácil impor testes e práticas de avaliação;

b) Testes e medidas educacionais podem ser rapidamente implementadas;

c) Os resultados são visíveis;

d) O custo de uma modificação é muito baixo;

e) O impacto nas práticas escolares é alto.

4ª Face – Validade consequencial

- Os exames de larga escala tem um grande apelo midiático e político, como observamos no caso do Enem no Brasil.
- Porém qual é o efeito dos testes para os professores?
- Em geral, uma escola deveria ter um patamar estável, ou cair de vez em quando em sua avaliação, pois os professores tendem a ensinar primeiro as coisas mais fáceis, e depois as mais difíceis (se houver tempo).

4ª Face – Validade consequencial

- Porém, o crescimento dos indicadores pode estar associado ao fato de que os estudantes e os professores estão mais acostumados com o formato e com as solicitações dos testes.
- Mais ainda: os professores, para atender às mais variadas demandas, trocam seu papel de formadores para se aproximarem de treinadores dos estudantes para os testes.