



SEMINÁRIO INTERNACIONAL DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA

O PAPEL DOS SISTEMAS DE AVALIAÇÃO
NA ORIENTAÇÃO DA APRENDIZAGEM

**PAINEL DE DISCUSSÃO 4:
COMPARABILIDADE ENTRE AVALIAÇÕES**

Ruben Klein (Fundação Cesgranrio)

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.



Parceiros:


**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

Necessidade de comparabilidade e de fazer julgamento.

Diferença entre avaliação de sala de aula e avaliação externa.

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

Em uma escola, duas turmas de uma mesma série.

Provas no mesmo dia. Provas diferentes não são comparáveis. A prova precisa ser a mesma. Não importa se o professor é o mesmo ou não.

Na escala do percentual de acerto, a variação é sempre de 0 a 10 ou de 0 a 100. Mas uma turma pode ter uma média de 7 em uma prova fácil e outra turma ter um média de 4 em outra prova difícil.

Qual é a melhor turma? Não sabemos.

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

Com a Teoria Clássica dos Testes (TCT) não é possível aplicar diferentes testes em populações distintas garantindo a comparabilidade

É preciso usar a Teoria de Resposta ao Item (TRI)

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

A TRI modela a probabilidade de resposta ao item em função da proficiência do aluno, uma variável latente (não observável), e de parâmetros (características) do item.

As escalas de proficiências são obtidas a partir da calibração dos itens de teste através do modelo da TRI. Em geral, o modelo utilizado é o modelo logístico. No Saeb e no ENEM utiliza-se o modelo de 3 parâmetros, a dificuldade (posição), a discriminação (inclinação da curva) e o a probabilidade de um aluno de baixa habilidade acertar o item (acerto casual).

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

Calibrar um item de teste significa identificar seus parâmetros, sendo uma das propriedades importantes da TRI o fato dos parâmetros dos itens e as proficiências dos indivíduos serem invariantes.

Isso quer dizer que se calibramos o mesmo teste em duas populações diferentes, podemos ir de uma calibração para outra por meio de uma transformação linear como no caso de graus centígrados para graus Fahrenheit.

Na TRI, a proficiência e o parâmetro de dificuldade (posição) do item estão na mesma escala.

Organização:



Parceiro realizador:



Parceiros:



Banco de itens calibrado na escala.

Itens que meçam as habilidades descritas na matriz de referência.

Parâmetros de dificuldade distribuídos em um largo espectro da escala de proficiência.

A partir desse banco pode-se extrair diferentes testes para a mesma população e para populações diferentes e cujos resultados estarão na mesma escala.

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

É importante que esse teste meça as diversas habilidades relativas a essa população e que seja adequado para ela, isso é, que contenha “informação” na região onde se encontra a distribuição de proficiência da população.

O conceito de informação de um teste é um conceito estatístico relacionado à precisão da medida da proficiência.

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

A TRI apresenta a propriedade que a informação do teste é a soma das informações dos itens (um item é um teste com somente um item). Como toda informação é positiva, a informação do teste aumenta com o número de itens e consequentemente o erro padrão da estimativa de θ decresce.

Ao contrário da TCT, o erro padrão da estimativa de proficiência depende da proficiência θ e sabe-se também o que acontece com a informação do teste na substituição ou acréscimo de itens.

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

Itens muito fáceis ou muito difíceis para um aluno fornecem pouca informação para a estimativa de sua proficiência. Itens com parâmetro de dificuldade “b” próximos da proficiência do aluno fornecem mais informação. O aumento do parâmetro “c” diminui a informação do item, pois aumenta a chance do acerto casual.

Esse resultado permite construir testes adaptados a populações diferentes, desde que se tenha informações a priori sobre a população e se tenha um banco de itens calibrados. É desejável um teste que tenha informação alta em uma boa extensão.

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

Na prática, para ter testes “mais comparáveis” para uma mesma população, é preciso ter testes com curvas de informação semelhantes.

É preciso calibrar os itens. Pode ser feito no próprio teste ou em pré-testes em populações adequadas.

Dada uma escala, em geral, para calibrarmos novos itens nessa escala, precisamos utilizar itens já calibrados (parâmetros conhecidos) no teste ou pré-teste.

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

Para isso é preciso ter itens não divulgados no banco de itens. É por isso que o SAEB não pode divulgar todos os itens.

A TRI tem duas hipóteses que precisam ser verificadas. A uni-dimensionalidade e a independência condicional.

Há várias maneiras de se estimar a proficiência e é preciso se estimar sempre da mesma maneira.

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

Comportamento diferencial do item. O item tem que ter o mesmo comportamento para as diferentes populações.

Interpretação da escala.

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

Para haver comparação precisa haver planejamento e garantir que as condições necessárias sejam satisfeitas.

A testagem deve ser feita na mesma época do ano e garantir a presença dos alunos.

Decisões sobre quais itens divulgar e quais manter sigilosos para serem usados na calibração de novos itens, na mesma escala.

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

Itens devem ser pré-testados, preferencialmente em uma amostra com características semelhantes a da população alvo.

Todos os itens vêm de um banco calibrado ou ter itens comuns já calibrados no novo teste, para calibrar novos itens na mesma escala.

Procedimentos de testagem padronizados

Etc.

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

A qualidade do banco vai depender da calibração se seus itens.

Depende do planejamento para efetuar a calibração de itens novos.

Que itens comuns devem ser utilizados, a quantidade de itens comuns, a população na qual os itens serão testados, etc.

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

Desafio: mudança de maneira de avaliar.

Testes computadorizados: lineares e adaptativos.

Lineares: testes iguais para todos. Mais parecidos com os tradicionais.

Adaptativos. Permite escolher itens adequados ao aluno. Aumenta a precisão de sua proficiência.

Testes computadorizados permitem a introdução de novos tipos de itens. Pode ser necessária a introdução de novas escalas.

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social

Avaliações estaduais .

Colocar na escala SAEB. Grande conquista.

Cuidado no planejamento dos testes.

Complementaridade com o SAEB.

Padronização do nível (ponto da escala) considerado adequado.

Organização:

consed
Conselho Nacional de Secretários de Educação

Parceiro realizador:

Instituto
**Ayrton
Senna**
Educação do futuro,
agora.

Parceiros:

**INSTITUTO
UNIBANCO**
35 anos

Itaú Social