

O processamento da avaliação padronizada utilizando o software CHIC: uma releitura dos resultados e a reflexão sobre a elaboração de novos instrumentos.

Thiago Fernando Ferreira Costa

Lenir Morgado da Silva

Claudio Maroja

Marcelo Rivelino Rodrigues

Fábio Oliveira Silva

Resumo

A presente pesquisa, aplicada e exploratória, evidencia a releitura dos resultados obtidos em uma avaliação padronizada, a partir do seu tratamento por meio da utilização do software CHIC (do inglês Correspondence & Hierarchical Cluster), sinalizando reflexões acerca do próprio instrumento e dos resultados obtidos, propiciando considerações e encaminhamentos, como a emissão de relatórios pedagógicos, o tratamento dos dados e a reelaboração do próprio instrumento. A pesquisa, realizada a partir de uma avaliação padronizada, como aporte, mostrou que o novo tratamento permite correlacionar os itens de forma a perceber relações significativas e hierárquicas entre eles, sinalizando a possibilidade de elaboração de instrumentos de avaliação que considerem, previamente, relações dessa natureza, permitindo maior assertividade e plausibilidade quanto aos resultados.

Introdução

A avaliação de sistemas é realizada, há tempos, por inúmeras esferas, devido às contribuições que dá no que diz respeito ao encaminhamento de ações que visam à melhoria da qualidade do ensino, considerando que os resultados ou impactos obtidos com a avaliação externa não devem ser o único fator a ser examinado durante a avaliação de um programa, projeto, ou de qualquer outro objeto de avaliação que ocorre na escola (NEVO, 1997). A avaliação de sistemas precisa ser eficaz em nível de instrumento, de processamento e geração de dados, bem como de comunicação à escola, para que possam ser feitos encaminhamentos negociados, autênticos e consistentes (ALAVARSE, 2013). Para que esse instrumento de avaliação seja de fato eficaz, seu processamento deve contar com rigor estatístico, seguindo padrões de qualidade. Um novo tratamento, no entanto, pode nos levar a outras reflexões sobre o aprimoramento do instrumento e sobre a correlação de seus resultados. Para

tecer essas reflexões, foi realizado o tratamento por meio do software CHIC, escolhido por permitir extrair, de um conjunto de dados e por meio do cruzamento de atributos, regras de associação entre variáveis, bem como fornecer um índice de qualidade de associação e representar uma estruturação das variáveis por meio dessas regras de associação. Esse tratamento permite obter classificação hierárquica entre dados e suas correlações, propiciando uma releitura dos resultados. Para a realização do novo tratamento, foi utilizada a Provinha São Paulo, edição 2017, concebida, elaborada e aplicada pela Secretaria Municipal de Educação de São Paulo, cujos resultados mostraram a eficácia do instrumento para avaliar a alfabetização matemática de estudantes do segundo ano do Ensino Fundamental.

O objetivo principal da Provinha São Paulo consistiu em diagnosticar os níveis de alfabetização em Matemática, dos estudantes, após um ano de estudo. Sua aplicação forneceu informações que podem contribuir com esse processo, e a pesquisa teceu reflexões sobre intervenções e sobre o próprio instrumento. Essa prova foi aplicada a todos os estudantes do 2º ano do Ensino Fundamental, da Rede Municipal de Ensino de São Paulo, em 2017, em caráter censitário, e se configurou como uma avaliação de segunda geração (BONAMINO; SOUZA, 2012). Os resultados foram divulgados às unidades escolares, sem ranqueamento ou bonificação, com o intuito de levar à reflexão sobre os resultados. A matriz de referência foi a mesma matriz de referência da Provinha Brasil. Diferentemente dessa provinha, no entanto, a Provinha São Paulo não foi constituída apenas por itens dicotômicos (questões de múltipla escolha), mas também por itens de resposta construída (itens abertos, que seriam, para fins de processamento, considerados politômicos dicotomizados). Foi levado em consideração que uma avaliação com itens abertos deveria ser cautelosamente elaborada e processada, para garantir validade, confiabilidade, comparabilidade e justiça, uma vez que avaliação com itens abertos tem sido uma área problemática desde a década de 1920 (TOFFOLI et al. 2016).

Metodologia

Foram realizados dois tratamentos diferentes na Provinha São Paulo. O primeiro, pela Teoria de Resposta ao Item, e o segundo, que se deu a partir dos resultados do primeiro, foi feito pelo uso do software CHIC. Foram analisados os resultados e feitas algumas problematizações para a construção de novos instrumentos e emissão de resultados.

Foi feita a aplicação de pré-teste, da Provinha, em duas versões do instrumento, cada uma composta por 15 itens de múltipla escolha e 5 de resposta construída. Não foram considerados os 5 itens da Prova São Paulo, do 3º Ano, para compor o pré-teste. Eles foram utilizados somente no instrumento final, para fins de processamento, por serem já calibrados. O instrumento final foi composto, portanto, por 25 itens. Posteriormente foi elaborado o Guia de Aplicação do Pré-teste da Provinha São Paulo e o Guia de Correção, tanto das questões de múltipla escolha como das questões de resposta construída.

Resultados e discussão

O desempenho dos estudantes, segundo análise clássica, foi alto para os itens de múltipla escolha. O percentual de acertos, na ordem dos itens de múltipla escolha, de 1 a 20, foi 94,40; 95,00; 88,10; 90,20; 84,10; 87,70; 86,80; 88,40; 90,40; 74,40; 53,90; 72,90; 54,80; 51,50; 44,10; 46,40; 68,00; 47,80; 41,50, 62,80. Quanto aos itens de resposta construída, conforme hipótese levantada, o nível de desempenho foi extremamente baixo. Para composição do instrumento final houve estudo e reflexão sobre aspectos pedagógicos e técnicos do pré-teste. Foi gerada a escala de proficiência, a partir dos mesmos critérios que geraram as escalas do SAEB, com média 250 e desvio padrão 50. Foram cobertas as 15 habilidades na matriz contidas e os itens apresentaram diferentes graus de complexidade. Para gerar a escala de proficiência, com base nos mesmos critérios que a escala SAEB, foi utilizado como balizador o valor mínimo de 65% (critério SAEB) de acerto do item em cada faixa de proficiência para a ancoragem.

Quadro 1: Percentuais de Acerto por Faixa de Proficiência

Itens	Descritores	25	50	75	100	125	150	175	200	225	250	275	Níveis
Item 01	D 4.1	30%	63%	78%	88%	93%	96%	97%	98%	99%	100%	100%	1
Item 02	D 5.1	14%	51%	82%	89%	94%	97%	98%	99%	99%	100%	100%	
Item 04	D 1.3	14%	43%	67%	79%	86%	91%	95%	97%	98%	99%	100%	
Item 08	D 1.1	11%	46%	70%	79%	83%	89%	93%	95%	96%	98%	100%	
Item 09	D 2.1	4%	22%	56%	76%	86%	93%	97%	98%	99%	99%	100%	
Item 03	D 6.2	7%	30%	59%	75%	81%	90%	95%	97%	98%	100%	100%	
Item 06	D 6.1	6%	28%	54%	73%	81%	89%	94%	97%	98%	98%	100%	
Item 07	D 4.2	9%	30%	60%	74%	81%	88%	92%	95%	96%	99%	100%	2
Item 05	D 1.2	7%	18%	39%	62%	71%	86%	93%	97%	98%	98%	100%	
Item 12	D 5.2	23%	41%	50%	62%	65%	68%	76%	83%	87%	94%	100%	
Item 10	D 6.1	8%	18%	35%	47%	57%	71%	85%	91%	95%	96%	100%	
Item 17	D 2.1	4%	12%	18%	36%	45%	62%	80%	90%	96%	97%	100%	
Item 20	D 6.2	7%	16%	19%	31%	41%	56%	72%	85%	92%	95%	100%	
Item 21*	D 2.1	0%	1%	3%	12%	34%	59%	77%	88%	94%	96%	100%	
Item 11	D 2.2	25%	21%	24%	31%	34%	41%	59%	74%	85%	92%	100%	3
Item 13	D 3.1	10%	20%	28%	32%	33%	39%	60%	79%	90%	96%	100%	
Item 14	D 3.2	8%	14%	18%	26%	30%	38%	56%	75%	86%	92%	100%	
Item 15	D 5.3	14%	19%	18%	21%	24%	25%	43%	70%	89%	95%	100%	
Item 19	D 5.3	7%	11%	12%	14%	17%	22%	40%	71%	88%	100%	100%	
Item 23*	D 3.1	0%	0%	1%	5%	14%	33%	61%	77%	90%	95%	100%	
Item 24*	D 3.2	0%	1%	1%	2%	6%	19%	47%	70%	87%	92%	100%	
Item 25*	D 2.1	0%	0%	1%	2%	9%	31%	57%	75%	88%	94%	100%	4
Item 16	D 2.2	16%	25%	25%	27%	28%	32%	47%	63%	80%	88%	100%	
Item 18	D 2.1	11%	22%	25%	29%	31%	38%	50%	62%	70%	96%	100%	
Item 22*	D 2.1	0%	0%	1%	3%	7%	20%	39%	56%	72%	82%	100%	

obs(*): percentual de alunos que alcançaram o nível mais alto - nível 4. (itens politômicos)

Início do desenvolvimento da habilidade

Auge desenvolvimento da habilidade - faixa de ancoragem

Máximo desenvolvimento da habilidade

Fonte: Relatório Provinha São Paulo -2017.

Quadro 2 - Divisão da escala de proficiência em níveis de desempenho e suas respectivas habilidades, segundo a interpretação pedagógica da ancoragem dos itens.

Nível de desemp.	Faixa de Profici.	Habilidades
1	50-100 pontos	Associar a contagem de coleções de objetos à representação numérica das suas respectivas quantidades; Comparar ou ordenar quantidades pela contagem para identificar igualdade ou desigualdade numérica; Resolver problemas que demandam as ações de juntar, separar, acrescentar e retirar quantidades; Identificar figuras geométricas planas; Reconhecer as representações de figuras geométricas espaciais;

		Comparar e ordenar comprimentos; Identificar informações apresentadas em tabelas; Identificar informações apresentadas em gráficos de colunas.
2	100-175 pontos	Todas as habilidades descritas no nível 1 mais: Associar a denominação do número a sua respectiva representação simbólica; Identificar e relacionar cédulas e moedas.
3	175-200 pontos	Todas as habilidades descritas nos níveis 1 e 2 mais: Resolver problemas que demandam as ações de comparar e completar quantidades; Resolver problemas que envolvam as ideias da multiplicação; Resolver problemas que envolvam as ideias da divisão; Identificar, comparar, relacionar e ordenar tempo em diferentes sistemas de medida
4	Acima de 200 pontos	Todas as habilidades descritas nos níveis 1, 2 e 3 mais: Resolver problemas que demandam as ações de juntar, separar, acrescentar e retirar quantidades.

O tratamento seguinte, utilizando o software CHIC, conforme amostra de análises apresentadas ao final desse texto, mostraram correlações entre itens e de hierarquia entre eles. O estudo desses e outros resultados apontaram para a possibilidade de elaborar um instrumento considerando, previamente, resultados condicionados a outros, por meio do acerto ou erro, para cada item.

Figura 1: Análise dos itens e níveis de proficiência com o Grafo Implicativo

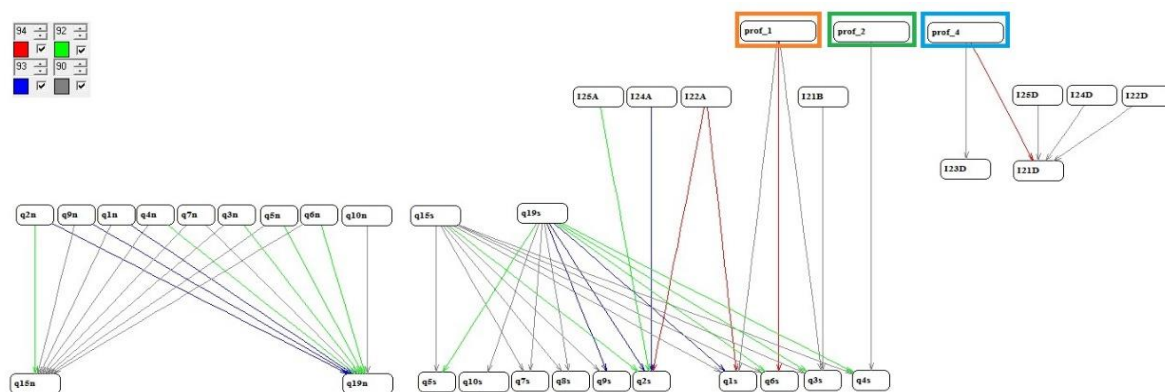
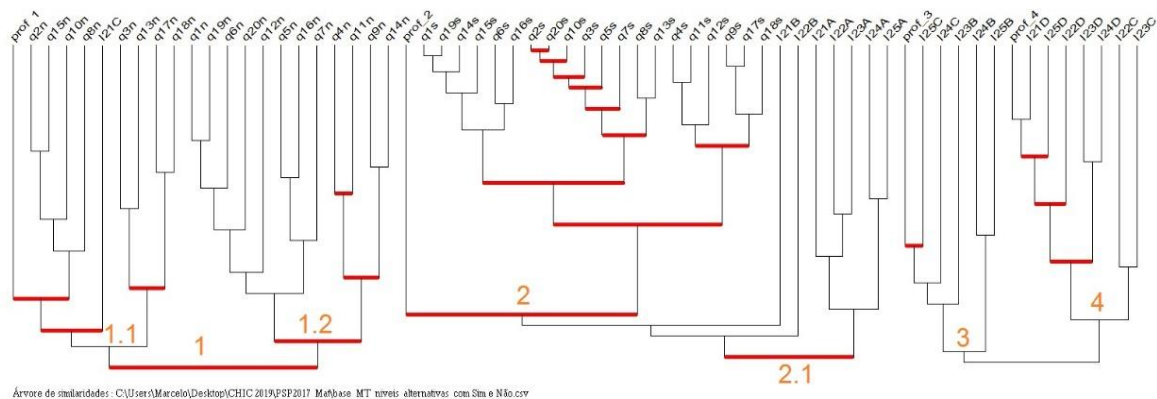
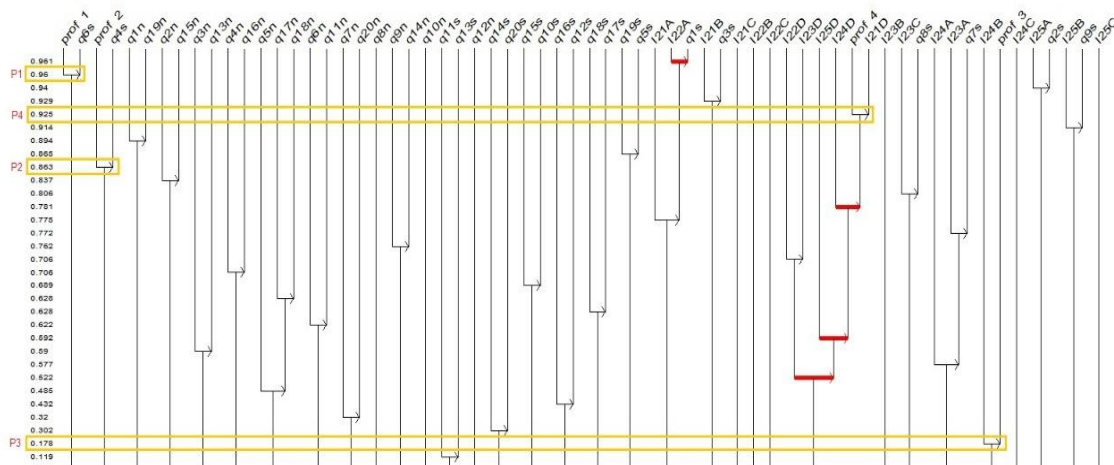


Figura 2: Análise dos itens e níveis de proficiência com a Árvore de Similaridade



Árvore de similaridades: C:\Users\Marcelo\Desktop\CHIC 2019\SP2017\Matbase MT níveis alternativos com Sim e Não.csv

Figura 3: Análise dos itens e níveis de proficiência com a Árvore Coersitiva



Árvore coesiva: C:\Users\Marcelo\Desktop\CHIC 2019\PSP2017_Mat\base MT_níveis alternativas com Sim e Não.csv

Partir da análise possibilitada pelo software C.H.I.C., observa-se uma definição das proficiências dos Níveis P1 e P4, ou seja, os de menores e maiores proficiências respectivamente. Por outro lado, os níveis intermediários P2 e P3, aparentemente não apresentam uma consistência significativa em relação aos grupos de alunos enquadrados nesses níveis de proficiências.

De acordo com o Grafo Implicativo o Nível P4 é o que aparece mais bem definido, enquanto na Árvore de Similaridade observa-se a relação consistente entre o Nível P1 com os respondentes que erraram a maior parte dos itens fechados e, que conseqüentemente, não apresentaram nenhuma resolução para itens de Respostas Construídas, fato este que ratifica a proficiência estabelecida para este grupo de respondentes.

Conclusões

Por hipóteses temos: primeiro a existência de uma possível interferência por parte dos corretores, que nesta edição da Provinha São Paulo foram os próprios professores (as) das turmas que participaram da avaliação; segundo que o guia de correção não possibilitou aos corretores uma clara definição dos respondentes que deveriam se enquadrar no Nível P2 ou no Nível P3.

A pesquisa sinalizou para a necessidade da realização de instrumentos eficazes de avaliação, bem como para a realização de processamento com rigor estatístico e tratamentos que permitem a aferição de resultados mais assertivos no que diz respeito àquilo que se pretende avaliar. O tratamento utilizando o software CHIC também levou à reflexão da possibilidade de realização de

análises pedagógicas mais detalhadas, possíveis com a hierarquização de dados e a correlação entre eles. Essa análise nos leva à problematização de elaboração de instrumentos de avaliação, incluindo o “Manual para as marcações dos itens de resposta construída” que otimizem o uso dessas correlações.

Referência bibliográfica

ALAVARSE, O. M. **Desafios da avaliação educacional: ensino e aprendizagem como objetos de avaliação para a igualdade de resultados.** In: Cadernos Cenpec/ Centro de Estudos e Pesquisas em Educação, Cultura e Ação Comunitária. São Paulo: CENPEC, v. 3, n. 1, jun. 2013, p. 135-153.

BONAMINO, A.; SOUZA, S. Z. **Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola.** Educação e Pesquisa, São Paulo, v. 38, n.2, p. 373-388, abr./jun. 2012.

NEVO, D. **Avaliação por diálogos: uma contribuição possível para o aprimoramento escolar.** In TIANA, Alejandro (coord). Anais do Seminário Internacional de Avaliação Educacional, 1 a 3 de dezembro de 1997. Tradução de John Stephen Morris. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais (Inep), 1998, p. 89-97.

TOFFOLI, S. F. L.; ANDRADE, D. F de; BORNIA, A. C; QUEVEDO-CAMARGO, G. **Avaliação com itens abertos: validade, confiabilidade, comparabilidade e justiça.** Educ. Pesquis., São Paulo, v. 42, n. 2, p. 343-358, abr./jun. 2016.