

# IMPACTO DA CONGLOMERAÇÃO NA ESTIMAÇÃO DO COEFICIENTE $H$ DA ESCALA DE MOKKEN<sup>1</sup>

MARCIA ANDRADE

CRISTIANO FERNANDES

PEDRO LUÍS DO NASCIMENTO SILVA

## RESUMO

*A finalidade deste estudo é avaliar o efeito do plano de amostragem por conglomerados na estimação do coeficiente  $H$  que desempenha papel importante tanto na construção quanto na avaliação de uma Escala de Mokken. Para tal, considerou-se a população de referência formada pelos alunos que frequentavam o 9º ano do Ensino Fundamental, na rede pública, em área urbana do estado do Rio de Janeiro que participaram da Prova Brasil 2007. As respostas a um conjunto de 10 itens dicotomizados que mensuram o capital econômico da família dos alunos foram usadas no cálculo de  $H$ . Foram selecionadas 2.000 amostras independentes dessa população de referência sob dois planos amostrais: amostragem por conglomerados em único estágio e amostragem aleatória simples. O estudo aponta o efeito significativo da amostragem por conglomerados na estimação do coeficiente  $H$ .*

Nossos agradecimentos ao Prof. Dr. Álvaro de Lima Veiga Filho, do Departamento de Engenharia Elétrica da Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio -, pela sugestão do tema deste artigo. Agradecemos, também, ao Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) pela concessão dos microdados da Prova Brasil 2007, ao Laboratório de Inteligência Computacional da PUC-Rio (ICA), à Escola Nacional de Ciências Estatísticas (ENCE) e ao CNPq.

**PALAVRAS-CHAVE** COEFICIENTES DE ESCALONABILIDADE •  
TEORIA DE RESPOSTA AO ITEM NÃO PARAMÉTRICA •  
AMOSTRAGEM COMPLEXA • NÍVEL SOCIOECONÔMICO.

## RESUMEN

*El propósito de este estudio es evaluar el efecto de los conglomerados para estimar el coeficiente H que desempeña un papel importante tanto en la construcción como en la evaluación de una Escala Mokken. Para tal fin, se consideró una población de referencia formada por alumnos del noveno grado de la enseñanza fundamental del sistema público, de un área urbana del Estado de Rio de Janeiro, que participaron en la Prova Brasil 2007. Para calcular H se usaron las respuestas dadas a un conjunto de 10 ítems dicotómicos que miden el capital económico de la familia de los estudiantes. Se seleccionaron 2000 muestras independientes de esa población de referencia a partir de procedimientos por muestreo: muestreo por conglomerados en una sola etapa y el muestreo aleatorio simple. El estudio muestra el efecto significativo del conglomerado para estimar el coeficiente H.*

**PALABRAS CLAVE** COEFICIENTES DE ESCALABILIDAD • TEORÍA DE LA RESPUESTA AL ÍTEM NO PARAMÉTRICA • DISEÑO MUESTRAL COMPLEJO • NIVEL SOCIOECONÓMICO.

## ABSTRACT

*The purpose of this paper is to evaluate the impact of cluster sampling design on the estimation of coefficient H, which plays a key role in the construction and quality evaluation of a Mokken scale. For this purpose we considered as our reference population junior high school students, in public schools, in the urban area of the state of Rio de Janeiro, who had taken the Prova Brasil 2007. The responses to a set of 10 dichotomized items that measure the economic capital of the students' families were used to calculate H. Repeated samples were selected from the same reference population using two sampling designs: simple random without replacement cluster sampling and simple random sampling without replacement. The results show a significant effect of cluster sampling design on the estimation of the Loevinger H coefficient. The possibility of using multilevel models must also be evaluated.*

**KEYWORDS** SCALABILITY COEFFICIENTS • NON PARAMETRIC ITEM RESPONSE THEORY • COMPLEX SAMPLING DESIGN • SOCIOECONOMIC LEVEL.

**1** Este procedimento apresenta as versões exploratória e confirmatória que estão disponíveis no pacote estatístico *R* apenas para dados independentes e identicamente distribuídos.

## INTRODUÇÃO

Com o intuito de mensurar construtos latentes, com base em um conjunto de itens dicotômicos ou politômicos ordinais, Mokken (1971) desenvolveu, no âmbito da Teoria de Resposta ao Item não Paramétrica (TRIN), um procedimento<sup>1</sup> eficiente e rápido para construção de escalas de Mokken.

Essa técnica oferece uma opção bastante flexível para a construção de escalas ordinais em diversas áreas do conhecimento, em que há dificuldade de se obterem itens com qualidade atestada, principalmente quando o conjunto de itens ou o tamanho da amostra de respondentes são pequenos, tornando assim inadequado o uso de modelos paramétricos.

Apesar de pouco divulgada no Brasil, alguns estudos empregaram a Escala de Mokken para a mensuração de construtos como: participação política (BORBA, 2012), ansiedade em matemática (CHAGAS et al., 2012), autocuidado de idosos (ANDRADE et al., 2011), hábitos de sedentarismo e atividade física (ABBES et al., 2011), prática pedagógica (ORTIGÃO, 2011), capital econômico (BONAMINO et al., 2010; ANDRADE; FRANCO; PITOMBEIRA, 2003), qualidade de vida na demência (INOUE

et al., 2009), prática docente em matemática (ORTIGÃO, 2009), nível socioeconômico (FRANCO et al., 2007), dentre outros. Além disso, esse procedimento de construção de escalas de Mokken pode ser usado também para avaliar o aspecto da dimensionalidade de um conjunto de itens (SHANSIS et al., 2004; SIJTSMA; MOLENNAR, 2002).

Dentre os coeficientes de escalonabilidade  $H_{ij}$ ,  $H_i$  e  $H$  usados para construir uma Escala de Mokken, o coeficiente  $H$  de Loevinger (1948) desempenha papel fundamental também na avaliação da consistência dessa escala (MOKKEN, 1971; ANDRADE, 2012).

A mensuração de variáveis latentes via escalas de Mokken é, geralmente, construída sob a hipótese de que as respostas dadas a um conjunto de itens são realizações de variáveis aleatórias independentes e identicamente distribuídas (SÄRNDAL et al., 1992). Entretanto, essa hipótese é violada quando os respondentes – os quais provêm de uma amostra probabilística – estão organizados em conglomerados (escolas, turmas, hospitais, setores censitários, domicílios etc.) e, além disso, são muito homogêneos dentro do conglomerado em termos das respostas dadas aos itens no instrumento de medida.

Dessa forma, torna-se necessária a incorporação dessa estrutura de correlação positiva dentro dos conglomerados<sup>2</sup> existentes na amostra para a estimação do coeficiente  $H$ .

<sup>2</sup> Efeito de conglomeração.

Existem sérias consequências nas estimativas pontuais de quantidades populacionais de interesse e na precisão dessas estimativas, caso seja feita a análise estatística de dados amostrais com estruturas de conglomeração como se fossem observações amostrais independentes e identicamente distribuídas. Assim, de modo geral, os resultados da estimação dos parâmetros de interesse ficam incorretos (SKINNER; HOLT; SMITH, 1989).

Diante da construção de escalas de Mokken, com dados oriundos de amostras com conglomeração desde 1961 (COSCO et al., 2012; MOKKEN, 1971), até agora não existem, no entanto, estudos sobre o impacto da amostragem de conglomerados na estimação dos coeficientes de escalonabilidade. Há apenas um único trabalho desenvolvido por Cohen *et al.*

(2008) no cenário da Teoria de Resposta ao Item (TRI) sobre a estimação dos parâmetros dos itens do modelo de Rasch.

Sendo assim, o propósito deste artigo é avaliar o efeito do plano amostral com conglomeração em um único estágio na estimação do coeficiente  $H$  com itens dicotômicos.

Este trabalho apresenta, inicialmente, as propriedades e as hipóteses de adequação de uma Escala de Mokken a um conjunto de itens – a estimação do coeficiente  $H$  –, sob uma amostra com conglomeração em único estágio, e a descrição dos estudos de simulação adequados ao contexto da amostragem de populações finitas, bem como a análise dos resultados da estimação de  $H$  sob dois planos amostrais de interesse.

## **METODOLOGIA**

### **POPULAÇÃO DE REFERÊNCIA**

Desde 2005, a Avaliação Nacional do Rendimento Escolar (Prova Brasil) é um dos instrumentos de avaliação da educação básica, aplicado anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Participam dessa avaliação censitária os alunos matriculados no 5º e 9º anos do Ensino Fundamental nas escolas públicas urbanas e rurais do território nacional, com pelo menos 20 alunos na série. O objetivo dessa avaliação está direcionado para o desempenho dos estudantes em leitura e resolução de problemas em matemática. Com os resultados obtidos por escolas, municípios etc. são apresentados diagnósticos da relação que envolve gestores, diretores e professores como corresponsáveis pelo rendimento dos estudantes que auxiliam na tomada de decisão de políticas educacionais.

A escolha do cadastro dessa avaliação educacional como um sistema de referência foi adequada para a realização dos estudos de simulação nessa população, pois disponibilizou informações atualizadas sobre o contexto escolar brasileiro e, além disso, um questionário socioeconômico que permitiu a mensuração do capital econômico por meio de uma Escala de Mokken.

Desta forma, considerou-se uma população de referência

formada pelos alunos do estado do Rio de Janeiro que frequentavam o 9º ano do Ensino Fundamental nas escolas públicas urbanas que participaram da Prova Brasil 2007. Nessa população, o total de 109.314 alunos estava organizado em 1.719 escolas públicas e 4.370 turmas.

### CAPITAL ECONÔMICO

A mensuração do capital econômico como a posse de bens de consumo (BORDIEU, 1985; ORTIGÃO, 2011; SOARES, 2005) foi considerada neste artigo, devido a sua utilização, para medir a condição socioeconômica dos alunos. Essa prática é comum em estudos educacionais que visam à avaliação do impacto do nível socioeconômico, por exemplo, na proficiência em matemática (ANDRADE; FRANCO; PITOMBEIRA, 2003), na eficácia escolar (FRANCO et al., 2007), nos programas de larga escala de avaliação da educação básica, como o Sistema de Avaliação da Educação Básica (Saeb), a Prova Brasil e o *Programme for International Student Assessment* (Pisa),<sup>3</sup> e no acesso ao ensino superior por meio do Exame Nacional do Ensino Médio (Enem) e do Programa Universidade Para Todos (Prouni).

Nesse sentido, com base no questionário do aluno da Prova Brasil 2007, foram usadas as respostas dadas a um conjunto de 10 itens dicotomizados (14, 5, 6, 12, 15, 8, 13, 7, 10 e 11) que mensuram o capital econômico da família dos estudantes no cálculo do coeficiente *H*. Esses itens foram rotulados, nesta ordem, como item 1, ..., item 10 (Tabela 1).

**3** Sob a coordenação do Inep, esses programas utilizam amostragem complexa para a coleta dos dados.

**TABELA 1 - Popularidades dos itens do capital econômico na população de referência**

ITEM	NA SUA CASA TEM... ? (RESPOSTA: 1: SIM; 0: NÃO)	POPULARIDADE (Pi)
1	Banheiro	0,980
2	Televisão em cores	0,978
3	Rádio	0,94
4	Carro	0,36
5	Empregada doméstica	0,09
6	Geladeira	0,98
7	Computador	0,48
8	Vídeocassete ou DVD	0,86
9	Máquina de lavar roupa	0,70
10	Aspirador de pó	0,18

Fonte: Elaboração do autor com base em dados do MEC/INEP/Prova Brasil, 2007 (ANDRADE, 2012).

## ESCALA DE MOKKEN

Para que a Escala de Mokken seja adequada para a mensuração do construto latente, o conjunto de itens precisa satisfazer algumas hipóteses, a saber: unidimensionalidade, independência local e monotonicidade das curvas características dos itens (CCI); além das seguintes condições sobre os coeficientes de escalonabilidade:  $H_{ij} > 0, \forall i \forall j$ ;  $H_i > 0,30^4 \forall i$  e  $H > 0,40$  (SIJTSMA; MOLENAAR, 2002, p. 51, 59). Sendo assim, o pesquisador pode utilizar o escore total<sup>5</sup> como uma estimativa do construto latente.

**4** O valor de 0,30 é sugerido por Sijtsma e Molenaar (2002), mas na prática valores maiores podem ser especificados com intuito de produzir itens com maior poder de discriminação.

**5** A soma das respostas atribuídas aos itens selecionados pela Escala de Mokken. Para um conjunto de  $J$  itens dicotômicos, esta soma varia de zero até  $J$ .

## HIPÓTESES DE ADEQUAÇÃO DA ESCALA DE MOKKEN UNIDIMENSIONALIDADE

De modo geral, os traços latentes que estão relacionados a comportamentos, atitudes ou aptidões apresentam várias dimensões (ANDRIOLA, 2009; PASQUALI, 2009). Mensurar aspectos multidimensionais, apesar de desejável do ponto de vista prático, é ainda um problema não solucionado de modo

satisfatório pela Psicometria moderna. Para lidar com esse detalhe, na maioria das vezes, o pesquisador pressupõe, no momento da construção do conjunto de itens, o enfoque unidimensional. Posteriormente, quando esse conjunto de itens for aplicado a uma amostra de respondentes, verificar-se-á a validade da hipótese de unidimensionalidade.<sup>6</sup>

O pressuposto de unidimensionalidade se refere à presença de um fator dominante (dimensão única ou principal) em detrimento dos outros fatores secundários (possíveis dimensões do traço latente) que estejam em vigor no momento da mensuração e que possam ser considerados suficientemente pequenos para serem descartados da medição (PASQUALI, 2009; GESSAROLI, 1994).

**6** Se a hipótese de unidimensionalidade for violada, verificar-se-á a adequação da Escala de Mokken às outras dimensões evidenciadas do traço latente.

### INDEPENDÊNCIA LOCAL

Segundo Pasquali (2009), a suposição da independência local pressupõe que a probabilidade de o indivíduo responder “corretamente” ao item  $i$  não é afetada pelas respostas dadas aos demais itens.

Considere um conjunto de  $J$  itens dicotômicos tal que  $i = 1, 2, \dots, J$  e seja  $\beta$  o traço latente dominante do indivíduo. A independência local das respostas de dado indivíduo ao conjunto unidimensional com  $J$  itens dicotômicos é definida por:

$$1. \quad \Pr(U_1, U_2, \dots, U_J | \beta) = \Pr(U_1 | \beta) \Pr(U_2 | \beta) \dots \Pr(U_J | \beta) = \prod_{i=1}^J \Pr(U_i | \beta)$$

em que  $U_i$  é a resposta positiva do indivíduo a um item dicotômico  $i$  e  $\Pr(U_i | \beta)$  é a probabilidade da resposta positiva desse indivíduo ao item  $i$  dado seu traço latente  $\beta$ .

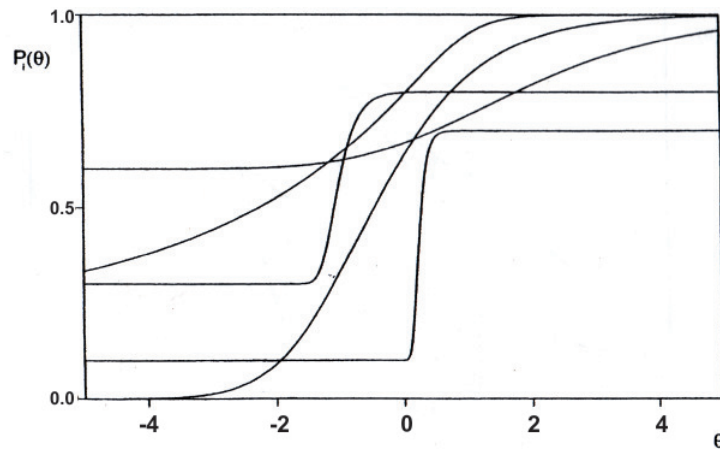
### MONOTONICIDADE NÃO DECRESCENTE DA CURVA CCI

A hipótese da monotonicidade da curva característica de um dado item  $i$  estabelece que a probabilidade de responder “corretamente” ao item  $i$  deve ser monótona e não decrescente em relação ao traço latente  $\beta$ , ou seja:

$$2. \quad \beta_a < \beta_b \Rightarrow \Pr(U_i | \beta_a) \leq \Pr(U_i | \beta_b)$$



**FIGURA 1** - Comportamento monótono não decrescente de algumas curvas CCI



Fonte: Sijtsma e Molenaar (2002).

Vale destacar que a curva CCI, satisfazendo a relação de ordem em (2), pode não ter uma forma simétrica, mas apresentar intervalos nos quais a função seja linear (Figura 1).

Para a avaliação das hipóteses de adequação de uma Escala de Mokken, empregando a *library Mokken* disponível no programa estatístico R (R DEVELOPMENT CORE TEAM, 2011), consulte o estudo de Van der Ark (2012).

**OS COEFICIENTES DE ESCALONABILIDADE  $H_{ij}$ ,  $H_i$  E  $H$**   
Mokken (1971) apresentou três coeficientes de escalonabilidade com o objetivo de construir escalas de Mokken com boas propriedades psicométricas.

O primeiro deles,  $H_{ij}$ , mede o grau de associação entre cada par de itens  $i$  e  $j$ . O segundo,  $H_i$ , avalia o poder de discriminação do item  $i$  em relação aos demais itens. O terceiro,  $H$ , mede a consistência (grau de homogeneidade) da escala como um todo. Para detalhes sobre o coeficiente de escalonabilidade entre pares de itens ( $H_{ij}$ ) e o de escalonabilidade de um item  $i$  ( $H_i$ ), veja Andrade (2012).

### **O COEFICIENTE $H$**

Dentre os coeficientes de escalonabilidade,  $H$  desempenha um papel importante tanto na construção de uma Escala de

Mokken quanto na avaliação da consistência dessa escala (MOKKEN, 1971; SIJTSMA; MOLENAAR, 2002; ANDRADE, 2012).

Sob o enfoque da amostragem de populações finitas, esse coeficiente pode ser definido como uma função de proporções populacionais de acertos individuais ( $P_i$ ) e simultâneos aos itens  $i$  e  $j$  ( $P_{ij}$ ), em que  $i = 1, 2, \dots, J$  e  $j = 2, \dots, J$ . Segundo Mokken (1971), a popularidade de um item  $i$  é definida como a proporção de acertos ao item. A relação de ordem entre as popularidades dos itens  $i$  e  $j$ , denotada por " $i < j$ ", significa que  $P_i < P_j$ , ou seja, o item  $i$  é mais difícil que o item  $j$ . Assumindo que os itens estejam ordenados em termos de suas popularidades populacionais, a expressão do coeficiente  $H$  é dada por:

$$3. \quad H = \frac{\sum_{i \neq j} (P_{ij} - P_i P_j)}{\sum_{i < j} (P_i - P_i P_j) + \sum_{i > j} (P_j - P_i P_j)}, \quad 0 \leq H \leq 1$$

Supondo que as hipóteses de adequação sejam válidas, uma Escala de Mokken pode ser classificada em fraca ( $0,30 \leq H < 0,40$ ), razoável ( $0,40 \leq H < 0,50$ ) ou forte (para valores de  $H$  acima de  $0,50$ ). Assim, de acordo com a classificação obtida da escala, o escore total pode ser empregado como uma estimativa do construto latente  $\beta$  (MOKKEN, 1971; SIJTSMA; MOLENAAR, 2002).

## ESTIMADOR DE $H$

### SOB DESENHO AMOSTRAL COMPLEXO

Ao mensurar uma variável latente mediante uma Escala de Mokken, o pesquisador pode deparar com três situações a respeito do plano amostral utilizado para a seleção da amostra na qual foram coletadas as respostas dadas a um conjunto de  $J$  itens dicotômicos que compõem um instrumento de medida. Esse plano amostral pode ser amostragem aleatória simples com reposição (AASC), amostragem aleatória simples (AAS) ou um plano amostral complexo. Este último apresenta algumas das seguintes características: conglomeração em um ou mais estágios, estratificação, probabilidades desiguais de sorteio para as unidades envolvidas etc.

Quando a amostra  $s$  for selecionada segundo um plano amostral complexo, a estimação do coeficiente  $H$  deve incorporar as características de identificação desse plano amostral. Assim, baseado na abordagem de modelos de superpopulação, o estimador  $\hat{H}_w$  pode ser definido (ANDRADE, 2012).

Como o plano de amostragem por conglomerados simples em único estágio (AC1S) é equivalente a uma AAS para valores agregados do conglomerado (BOLFARINE; BUSSAB, 2005), o estimador de  $H$  ( $\hat{H}$ ), decorrente da expressão (3), pode ser empregado nessa situação. Desse modo,  $\hat{H}$  é dado pela seguinte expressão:

$$4. \quad \hat{H} = \frac{\sum_{i=1}^{J-1} \sum_{j=i+1}^J (\hat{P}_{ij} - \hat{P}_i \hat{P}_j)}{\sum_{i=1}^{J-1} \sum_{j=i+1}^J \hat{P}_i (1 - \hat{P}_j)}, \quad i = 1, 2, \dots, J; j = 2, 3, \dots, J$$

em que  $\hat{P}_i$ : estimador da proporção de acertos do item  $i$  e  $\hat{P}_{ij}$ : estimador da proporção de acertos simultâneos aos itens  $i$  e  $j$ .

Vale destacar que as notações  $\hat{H}_{ac1s}$  e  $\hat{H}_{aas}$  são referentes aos estimadores de  $H$  sob os planos de AC1S e AAS, respectivamente.

## ESTUDOS DE SIMULAÇÃO

Para investigar o efeito da conglomeração em único estágio na estimação de  $H$ , além dos aspectos de vício, precisão e normalidade dos estimadores pontuais propostos, foram desenvolvidos dois estudos de simulação, com 2.000 réplicas de amostras independentes selecionadas da mesma população de referência, sob dois planos amostrais: AC1S e AAS.

### PLANO AC1S

No primeiro estágio foram selecionadas, por meio da amostragem aleatória simples, 90 escolas do 9º ano de um total de 1.719. Em seguida, todos os alunos nas escolas selecionadas foram incluídos na amostra.

Sob AC1S, os pesos amostrais atribuídos às escolas e aos alunos são iguais e desnecessários no cálculo das estimativas do coeficiente  $H$ .

Para a comparação dos planos amostrais adotados neste estudo, foi necessário fixar o tamanho da AAS em 5.723 alunos, uma vez que as escolas possuem tamanhos diferentes em termos do total de alunos do 9º ano que participaram da avaliação (KISH, 1965).

### QUALIDADE DOS ESTIMADORES SOB O PLANO AMOSTRAL

Foram calculadas as estimativas de medidas estatísticas para avaliar a qualidade de um estimador  $\hat{\theta}$  segundo um plano amostral A, a saber: valor esperado ( $E_A$ ), variância ( $V_A$ ), vício ( $B_A$ ), razão de vício ( $BR_A$ ) e erro quadrático médio ( $EQM_A$ ). Essas estimativas descrevem características importantes do comportamento dos estimadores pontuais por meio de uma aproximação da sua distribuição amostral (SÄRNDAL et al., 1992). Essas distribuições foram obtidas pelo processo de amostragem repetida e foram utilizadas no desenvolvimento deste estudo conforme o trabalho de Bean (1975).

Cabe destacar que a variância da aproximação da distribuição amostral do estimador  $\hat{\theta}$  foi considerada uma estimativa da variância verdadeira de  $\hat{\theta}$  (BEAN, 1975).

O impacto da conglomeração em único estágio na estimação da variância de  $\hat{H}_{ac1s}$  foi avaliado pela razão entre as estimativas de variâncias de  $\hat{H}_{ac1s}$  e  $\hat{H}_{aas}$ , nessa ordem (SÄRNDAL et al., 1992).

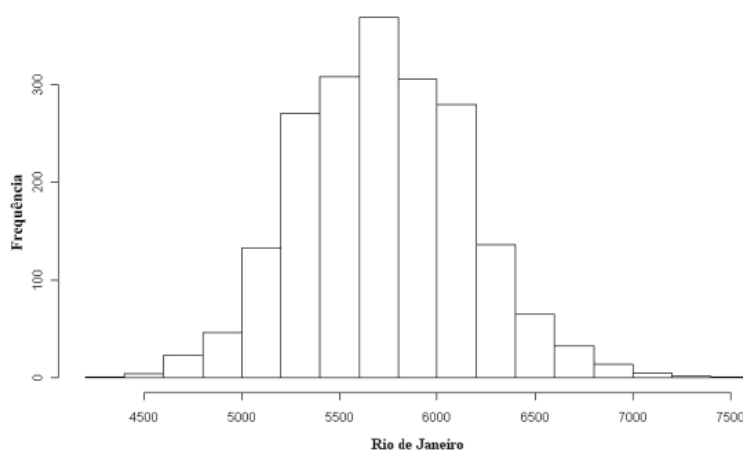
Para testar as hipóteses de igualdade entre as médias e as variâncias das distribuições amostrais bem como a hipótese de normalidade, foram empregados os testes: *Welch two sample t-test*, *Bartlett test of homogeneity of variances* e *Shapiro-Wilk normality test* (ROYSTON, 1982). Além disso, para testar se o estimador pontual é viciado, foi usado *one sample t-test*.

Nos processos de amostragem repetida – avaliação das hipóteses de adequação de uma Escala de Mokken e da estimação do coeficiente  $H-$ , foram usados os programas R (versão 2.13) e *Complex Mokken* (ANDRADE, 2012), respectivamente.

## RESULTADOS

Nesta seção são apresentados os resultados da estimação de  $H$  obtidos pelos estudos de simulação na população de referência do Rio de Janeiro. O tamanho das réplicas de amostras conglomeradas de alunos pode ser aproximado por uma distribuição normal com média 5.726 e desvio padrão 433 (Figura 2).

**FIGURA 2** - Histograma do número de respondentes selecionados sob AC7S na população de referência



Fonte: Andrade (2012).

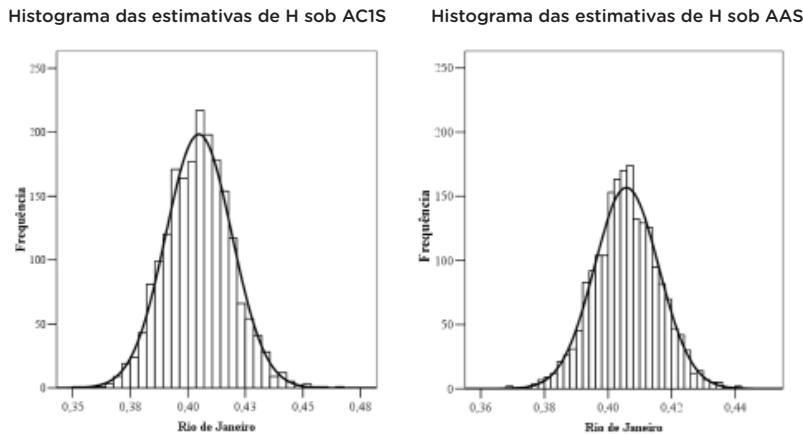
De modo geral, as popularidades dos itens e as proporções de acertos simultâneos foram bem estimadas, considerando-se que o coeficiente de variação da distribuição amostral das proporções de acertos são inferiores a 1%.

Em cada réplica, as hipóteses de adequação de uma Escala de Mokken são válidas no conjunto de itens fixados.

Em ambos os planos amostrais, a ordenação incorreta (inversa) das popularidades populacionais ocorreu apenas nos pares: (1,2) e (2,6), em torno de 8,4% e 2,3% do total de réplicas, respectivamente.

Uma aproximação da distribuição amostral de  $\hat{H}_{aa}$  e de  $\hat{H}_{acl}$ , obtida com 2.000 réplicas de amostras selecionadas sob cada plano amostral, mostrou-se estável para a avaliação do comportamento desses estimadores (Figura 3).

**FIGURA 3** - Histogramas da distribuição amostral de  $\hat{H}_{ac1s}$  e  $\hat{H}_{aas}$  com densidade Normal superposta



Fonte: Andrade (2012).

As distribuições amostrais diferem em termos da média e da variância ( $p < 0,05$ ). Além disso, a distribuição amostral de  $\hat{H}_{aas}$  pode ser aproximada por uma distribuição normal ( $p > 0,05$ ) (Tabela 2).

Ao selecionar uma amostra conglomerada sob o plano AC1S com tamanho entre 4.362 a 7.424 alunos nessa população de referência, a probabilidade da variável aleatória  $\hat{H}_{ac1s}$  assumir um valor no intervalo [0,30; 0,41] é igual a 0,53.

A Tabela 2 apresenta os resultados da estimação de  $H$  obtidos pela simulação, considerando cada plano amostral  $A$  adotado na população de referência.

**TABELA 2** - Resultados da estimação de  $H$  sob AC1S e AAS na população de referência

PLANO AMOSTRAL A	$E_A$	$V_A$	$B_A$	$EQM_A$	$W^i$	H
AC1S	0,405	2,062	-0,001	2,071	0,998	0,406
AAS	0,406	1,037	0,000	1,037	0,999	

Fonte: Andrade (2012).

Nota: As estimativas de  $V_A$  e  $EQM_A$  foram multiplicadas pelo fator  $10^4$ ; (i) A estatística do teste de normalidade de Shapiro - Wilk.

Conforme a Tabela 2, sob AC1S,  $\hat{H}_{ac1s}$  é viciado ( $p < 0,05$ ) e a estimativa do vício (em valores absolutos) corresponde a 6,46%

da estimativa do erro padrão desse estimador. Além disso, a estimativa de  $EQM_{AC1S}(\hat{H}_{ac1s})$  é 99,5% maior quando comparada com a estimativa de  $EQM_{AAS}(\hat{H}_{aas})$ .

## DISCUSSÃO

Os resultados aqui apresentados evidenciam que o plano amostral AC1S influencia a estimação do coeficiente de escalonabilidade  $H$ .

Na população de referência do Rio de Janeiro, as escolas (unidades primárias de amostragem) possuem tamanhos diferentes em termos dos alunos participantes da Prova Brasil em 2007. Dessa forma, no processo de amostragem repetida de escolas sob AC1S, o tamanho das réplicas de amostras conglomeradas de alunos é uma variável aleatória (SÄRNDAL et al., 1992), que apresentou uma distribuição equilibrada de respondentes que, por sua vez, foram consideradas na estimação de  $H$ .

Em face dos achados, o estimador  $\hat{H}_{aas}$  pode ser classificado como um estimador acurado (SÄRNDAL et al., 1992), uma vez que, em média, as estimativas  $\hat{H}_{aas}$  estão muito próximas do parâmetro populacional  $H$  e, além disso, apresentam pouca variabilidade quando comparadas com as estimativas de  $H$  calculadas sob o plano AC1S. Uma justificativa para esse fato pode estar relacionada ao tamanho fixado das réplicas de AAS, ao uso de estimadores de proporções (de acertos individuais e simultâneos aos itens) com elevada precisão e a inexistência do efeito da conglomeração.

Além disso, apesar de o estimador  $\hat{H}_{aas}$  apresentar as principais qualidades procuradas pela amostragem: pequeno vício e pequena variância (BOLFARINE; BUSSAB, 2005), esse estimador do coeficiente  $H$  não é adequado em situações em que a amostra selecionada de respondentes apresenta elevado grau de homogeneidade em relação às respostas aos itens utilizados do capital econômico.

Segundo Bolfarine e Bussab (2005), a presença de vício estatisticamente significativo no estimador  $\hat{H}_{ac1s}$  pode ser explicada pela natureza não linear do estimador (*vício técnico*),

pelo plano amostral adotado e pelo tamanho das réplicas. Além disso, a ordem de grandeza observada no vício não compromete, em média, a classificação final da escala.

Somente sob AC1S, a estimativa da razão de vício de  $\hat{H}_{ac1s}$  foi superior a 5%. De acordo com Särndal *et al.* (1992), esse resultado tem repercussão na probabilidade de cobertura nominal do intervalo de confiança e também no cálculo desse intervalo; uma vez que deve incorporar a estimativa de  $EQM_{AC1S}$  ao invés da estimativa de variância.

Em virtude da alta homogeneidade dentro dos conglomerados, a perda de eficiência do plano amostral AC1S comparado ao plano AAS na estimação do coeficiente  $H$  era esperada (BOLFARINE; BUSSAB, 2005). Isso foi observado no estudo de Cohen *et al.* (2008), que mostrou o impacto de um plano com conglomeração na estimação da variância dos estimadores dos parâmetros dos itens do modelo de Rasch. Além disso, demonstra que, ao ignorar a estrutura de conglomeração no cálculo da estimativa de variância do coeficiente  $H$ , os resultados, os testes de hipótese, o nível nominal dos intervalos de confiança e as conclusões a respeito da construção e classificação de uma Escala de Mokken ficam inválidos.

Como foi visto anteriormente, o estimador  $\hat{H}$  pode ser escrito como uma função de estimadores de proporções de acertos individuais (popularidades) e simultâneos aos itens. Segundo Van Onna (2004), em amostras suficientemente grandes, a propriedade de normalidade de  $\hat{H}$  é verificada quando os estimadores das proporções de acertos individuais e simultâneos aos itens também apresentam essa propriedade. A afirmação não se mostrou válida neste estudo. Apesar do tamanho grande das réplicas de AAS ( $n=5723$ ) e a propriedade de normalidade dos estimadores das popularidades serem observadas em apenas 8 itens, a distribuição amostral de  $\hat{H}_{aas}$  pode ser aproximada por uma distribuição normal.

Segundo Sijtsma e Molenaar (2002), em decorrência do processo de amostragem repetida, do tamanho das réplicas e dos pares de itens com popularidades populacionais muito próximas, é possível a ordenação inversa nas popularidades estimadas dos itens em algumas réplicas. Quando esse fenômeno ocorre com frequência elevada, o



processo de inferência baseado em estudos de simulação pode gerar resultados incorretos. Entretanto, neste estudo, o fenômeno da ordenação inversa ocorreu nos pares: (1,2) e (2,6), de forma balanceada em ambos os processos de seleção de réplicas, sem comprometer a qualidade dos resultados obtidos pela simulação.

### **CONSIDERAÇÕES FINAIS**

Neste estudo, o uso da amostragem de conglomerados em único estágio (AC1S) teve impacto significativo na estimação do coeficiente  $H$ . Isso mostrou que a organização dos alunos em conglomerados naturais existentes na população e o grau elevado de similaridade dos respondentes em relação às respostas aos itens que mensuram o capital econômico não devem ser ignorados na estimação desse coeficiente.

Recomenda-se, com o intuito de melhorar a eficiência do uso da amostragem de conglomerados, sem aumentar o tamanho da amostra de respondentes, por exemplo, o uso da amostragem em dois estágios (sorteio de turmas dentro das escolas selecionadas) ou a seleção da amostra de conglomerados com probabilidades proporcionais a uma medida de tamanho dos conglomerados.

Finalmente, a realização deste artigo contribui também para indicar o pouco desenvolvimento de pesquisa na área da Teoria de Resposta ao Item não Paramétrica (TRIN), apesar dos estudos recentes desenvolvidos por Andrade (2012) apresentarem uma nova abordagem baseada na amostragem complexa para os estimadores dos coeficientes de escalonabilidade e seus respectivos erros padrão, ao contrário do que acontece no âmbito da Teoria de Resposta ao Item (TRI).

Dessa forma, fica em aberto, como uma linha de pesquisa para estudos futuros, a construção passo a passo de uma Escala de Mokken que incorpore as características de uma amostra complexa (ANDRADE, 2012).

## REFERÊNCIAS

- ABBES, P. T. et al. Sedentarismo e variáveis clínico-metabólicas associadas à obesidade em adolescentes. *Revista de Nutrição*, v. 24, n. 4, p. 529-538, 2011.
- ANDRADE, M. S.; FRANCO, C.; PITOMBEIRA, J. B. Gênero e desempenho em matemática ao final do ensino médio: quais as relações? *Estudos em Avaliação Educacional*, São Paulo, n. 27, p. 77-96, 2003.
- ANDRADE, M. S. et al. Uma proposta para a mensuração do autocuidado em idosos. *Cogitare Enfermagem*, Curitiba, v.16, n.4, p. 661-666, 2011.
- ANDRADE, M. S. *A new approach for estimating the coefficients of scalability associated with Nonparametric Item Response Theory*. 2012. 213 p. PhD (Thesis) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2012.
- ANDRIOLA, W. B. Psicometria moderna: características e tendências. *Estudos em Avaliação Educacional*, São Paulo, v. 20, n. 43, p. 319-340, 2009.
- BEAN, J. A. Distribution and properties of variance estimators for complex multistage probability samples: an empirical distribution. *Vital Health Statistical*, v. 65, p. 1-50, 1975.
- BONAMINO, A. et al. Os efeitos das diferentes formas de capital no desempenho escolar: um estudo à luz de Bordieu e de Coleman. *Revista Brasileira de Educação*, Rio de Janeiro, v. 15, n. 45, p. 487-594, 2010.
- BOLFARINE, H.; BUSSAB, W. O. *Elementos de amostragem*. São Paulo: Edgard Blucher, 2005.
- BORBA, J. Participação política: uma revisão dos modelos de classificação. *Revista Sociedade e Estado*, Brasília, v. 27, n. 2, p. 263-288, 2012.
- BORDIEU, P. The forms of capital. In: RICHARDSON, J. G. (Ed.). *Handbook of theory and research for the sociology of education*. New York: Greenwood, 1985. p. 241-258.
- CHAGAS, P. P. et al. Math anxiety questionnaire: similar latent structure in Brazilian and German school children. *Child Development Research*, v. 2012, p. 1-11, 2012.
- COHEN, J. et al. Consistent estimation of Rasch item parameters and their standard errors under complex sample designs. *Applied Psychological Measurement*, v. 32, n. 4, p. 289-310, 2008.
- COSCO, T. D. et al. Mokken scaling analysis of the Hospital Anxiety and Depression Scale in individuals with cardiovascular disease. *General Hospital Psychiatry*, n. 34, p. 167-172, 2012.
- FRANCO, C. et al. Qualidade e equidade em educação: reconsiderando o significado de “fatores intra-escolares”. *Ensaio: Avaliação Política Pública Educacional*, Rio de Janeiro, v. 15, n. 55, p. 277-298, 2007.
- GESSAROLI, M. E. The assessment of dimensionality via local and essential independence: a comparison in theory and practice. In: LAVEAULT, D. et al. (Ed.). *Modern theories of measurement: problems and issues*. Ottawa, Canada: Faculty of Education, University of Ottawa, 1994. p. 93-104.

INOUYE, K. et al. Instrumentos específicos para mensurar a qualidade de vida na demência: levantamento, descrição, análise e comparação. *Temas em Psicologia*, Ribeirão Preto, v. 17, n. 2, p. 541-552, 2009.

KISH, L. *Survey sampling*. Nova Iorque: Wiley, 1965.

LÖEVINGER, J. The technique of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychological Bulletin*, v. 45, p. 507-529, 1948.

MOKKEN, R. J. *A theory and procedure of scale analysis*. Mouton: The Hague, 1971.

ORTIGÃO, M. I. A sala de aula de matemática: avaliação das práticas docentes. *Bolema*, Rio Claro, v. 22, n. 33, p. 117-140, 2009.

\_\_\_\_\_. Análise das práticas de professores de matemática da educação básica. *Estudos em Avaliação Educacional*, São Paulo, v. 22, n. 48, p. 29-52, 2011.

PASQUALI, L. *Psicometria: teoria dos testes na psicologia e educação*. 3. ed. Petrópolis: Vozes, 2009.

R DEVELOPMENT CORE TEAM. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, 2011. Disponível em: <<http://www.R-project.org/>>. Acesso em: 1 jan. 2010.

ROYSTON, P. An extension of Shapiro and Wilk's *W* test for normality to large samples. *Applied Statistics*, n. 31, p. 115-124, 1982.

SÄRNDAL, C. E.; SWENSSON, B.; WRETMAN, J. H. *Model assisted survey sampling*. New York: Springer-Verlag, 1992.

SHANSIS, F. et al. Desenvolvimento da versão em português da Escala de Avaliação de Mania de Bech-Rafaelsen (EAM-BR). *Revista Psiquiatria*, São Paulo, v. 26, n. 1, p. 30-38, 2004.

SIJTSMA, K.; MOLENAAR, I. W. *Introduction to nonparametric item response theory*. Netherlands: Sage, 2002.

SKINNER, C. J.; HOLT, D.; SMITH, T. M. F. *Analysis of complex surveys*. Chichester: Wiley, 1989.

SOARES, F. J. Utilização da teoria de resposta ao item na produção de indicadores socioeconômicos. *Pesquisa Operacional*, Rio de Janeiro, v. 25, n. 1, p. 83-112, 2005.

VAN DER ARK, L. A. New developments in Mokken scale analysis. *Statistical Software*, v. 48, n. 5, p. 1-27, 2012.

VAN ONNA, M. J. H. Estimates of the sampling distribution of scalability coefficient *H*. *Applied Psychological Measurement*, v. 28, n. 6, p. 427- 449, 2004.

**MARCIA ANDRADE**

Doutora pelo Departamento de Engenharia Elétrica da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Professora do Departamento de Engenharia de Produção da Universidade Cândido Mendes (UCAM)  
*armsandrade@gmail.com*

**CRISTIANO FERNANDES**

Professor doutor do Departamento de Engenharia Elétrica da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)  
*cris@ele.puc-rio.br*

**PEDRO LUÍS DO NASCIMENTO SILVA**

Professor doutor da Escola Nacional de Ciências Estatísticas (ENCE-IBGE)  
*pedronsilva@gmail.com*

**Recebido em:** ABRIL 2013

**Aprovado para publicação em:** JULHO 2013