



Lina Kátia Mesquita de Oliveira

**TRÊS INVESTIGAÇÕES
SOBRE ESCALAS DE PROFICIÊNCIA
E SUAS INTERPRETAÇÕES**

Tese de Doutorado

Tese apresentada ao Programa de Pós-Graduação em Educação do Departamento de Educação da PUC-Rio como parte dos requisitos parciais para obtenção do título de Doutor em Educação.

Orientador: Prof. Francisco Creso J. F. Júnior
Co-Orientador: Prof. Tufi Machado Soares

Rio de Janeiro
Agosto de 2008



Lina Kátia Mesquita de Oliveira

**TRÊS INVESTIGAÇÕES
SOBRE ESCALAS DE PROFICIÊNCIA
E SUAS INTERPRETAÇÕES**

Tese apresentada como requisito parcial para obtenção do título de Doutor pelo Programa de Pós-Graduação em Educação do Departamento de Educação do Centro de Teologia e Ciências Humanas da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof^a. Francisco Creso J.F. Júnior
Orientador
Departamento de Educação - PUC-Rio

Prof. Tufi Machado Soares
Co-orientador
UFJF

Prof^a. Alícia Maria Catalano de Bonamino
Departamento de Educação - PUC-Rio

Prof^a. Maria Aparecida C. Mamede Neves
Departamento de Educação - PUC-Rio

Prof. Manuel Palácios da Cunha e Melo
UFJF

Prof. Ruben Klein
Fundação Cesgranrio

Prof. Amaury Patrick Gremaud
USP

Prof. Paulo Fernando C. de Andrade
Coordenador Setorial do Centro de
Teologia e Ciências Humanas

Rio de Janeiro, 29 de agosto 2008.

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

Lina Kátia Mesquita de Oliveira

Graduou-se em Matemática pela Faculdade de Filosofia, Ciências e Letras de Belo Horizonte (FAFIBH) - 1976. É Especialista em Educação Matemática pela Universidade Federal de Juiz de Fora (UFJF) - 1999. É Mestre em Educação pelo Programa de Pós-Graduação da UFJF, tendo defendido a Dissertação em junho de 2002. Atualmente é Coordenadora da Unidade de Avaliação do Centro de Políticas Públicas e Avaliação da Educação da UFJF. Atua em projetos na área de Avaliação de Sistemas Educacionais e de Gestão da Educação.

Ficha Catalográfica

Oliveira, Lina Kátia Mesquita de

Três investigações sobre escalas de proficiência e suas interpretações / Lina Kátia Mesquita de Oliveira ; orientador: Francisco Creso J. F. Júnior ; co-orientador: Tufi Machado Soares. – 2008.

216 f. : il. ; 30 cm

Tese (Doutorado em Educação)–Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

Inclui bibliografia

1. Educação – Teses. 2. Avaliação educacional. 3. Escalas de proficiência. 4. Análise de conglomerado. 5. Níveis de proficiência escolar. 6. Interpretação pedagógica de itens. I. Franco, Creso. II. Soares, Tufi Machado. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Educação. IV. Título.

CDD: 370

Agradecimentos

Nesse momento, meu objetivo é externar minha gratidão, um dos mais nobres sentimentos expressos pela humanidade, e fazer com que todos aqueles que contribuíram para que eu chegasse até aqui possam permanecer na memória.

Ao meu orientador, Professor Creso Franco, de quem tive a honra de merecer a confiança e compartilhar do seu compromisso com a qualidade e a propriedade da pesquisa educacional. A ele devo a pergunta adequada, o tema desafiador, a inquietação para buscar o caminho mais rico e proveitoso. Séri o e competente, deixará sua marca no movimento das universidades brasileiras pela excelência na área da avaliação da educação.

Ao Professor Tufi Machado Soares, meu co-orientador e construtor dos modelos estatísticos adotados na pesquisa. Conhecedor profundo das análises de dados quantitativos, dispôs-se a buscar comigo o caminho mais adequado à pesquisa proposta. Seu trabalho, amplamente conhecido, certamente acrescentou dose significativa de valor a esta tese.

Ao Manuel Palacios da Cunha e Mello, responsável por minha formação acadêmica, profissional e, especialmente, por minha capacidade de sonhar e acreditar, generosamente, que o mundo pode ser melhor. Tenho muito orgulho de participar com ele da construção de um centro de estudos e pesquisas que se constituiu como o projeto mais inovador

na área da avaliação educacional no Brasil, justamente por ser um espaço de formação, de parceria e de excelência científica e tecnológica.

Aos professores participantes das bancas de qualificação, Alícia Bonamino, Fátima Alves, Aparecida Mamede, Amaury Patrick Gremaud, agradeço a preciosa colaboração, as idéias criativas e a disponibilidade.

Tenho muita honra de poder contar, em minha banca, com o Professor Rubem Klein, referência obrigatória e autoridade reconhecida pelo seu trabalho precursor na construção e desenvolvimento de programas de avaliação educacional em larga escala.

Não poderia deixar de registrar, nos meus agradecimentos, a contribuição institucional dada

- pelo Departamento de Educação da PUC-Rio pela qualidade dos cursos e excelência na formação acadêmica e profissional que propicia a seus alunos;
- pela Diretoria de Avaliação da Educação Básica do Instituto Nacional de Pesquisa Anísio Teixeira, do Ministério da Educação, pela disponibilização da base de dados do SAEB;
- pelo CAEd / UFJF, meu local de trabalho, referência nacional na área da avaliação educacional, onde teoria e prática se associam em busca do melhor debate e da mais cuidadosa interpretação dos temas relevantes da educação pública.

À equipe de pesquisa e análise de dados do CAEd, agradeço pela acolhida carinhosa e sempre disponível. Agradeço especialmente ao Neimar, ao Wellington e ao Luís Fajardo, que me acompanharam mais de perto, me apoiando nesta caminhada. Muitas vezes transformaram minhas dificuldades no campo da estatística em idéias claras, objetivas e promissoras. Ao Ailton e ao Helson, agradeço pelo desenvolvimento dos sistemas operacionais utilizados.

A Magda, Ederaldo, Janine, Verônica, Isabela, Regina, Lulude, Claytom e Márcio pelo apoio e ajuda constante em todos os momentos.

À minha amiga Wanda Alves, pela disponibilidade para ler, observar e opinar na área da educação matemática.

À Luiza que, com muito carinho e competência nas artes gráficas, editorou o trabalho, não se poupando mesmo nos fins de semana.

À Shirley Torres e à Josiane Toledo pela disponibilidade, apoio e revisão dos textos.

Aos meus queridos, Tio Raimundo e Tia Zezé, e às queridas amigas Lucy, Marlene, Bia e Letícia. Obrigada pela confiança, palavras de incentivo e pela torcida.

À Eleuza, minha irmã emprestada, companheira e amiga, que me ofereceu o olhar confiante, a palavra de incentivo e a mais rica oportunidade da vida: poder contar com a pessoa certa na hora certa, o meu agradecimento pela valiosa colaboração na revisão desse trabalho.

Por fim, de modo muito especial e carinhoso agradeço à minha família pelo companheirismo, solidariedade e força com que me presentearam. Apoio incondicional, paciência e segurança. Obrigada, José Francisco, Fernando e Mayra, vocês realmente são uns amores.

Dedico esse trabalho à minha mãe. Tenho certeza que, onde estiver, olha por mim, me abençoa e protege.

Resumo

Oliveira, Lina Kátia Mesquita de; Franco Jr., Francisco Creso Junqueira. **Três Investigações sobre Escalas de Proficiência e suas Interpretações**. Rio de Janeiro, 2008. 216 p. Tese de Doutorado - Departamento de Educação, Pontifícia Universidade Católica do Rio de Janeiro.

As avaliações de proficiência em larga escala vêm exercendo um papel relevante no contexto educacional contemporâneo, pois servem, entre outros propósitos, para retratar mais precisamente o nível de desempenho escolar de uma grande quantidade de alunos avaliados, como também para fornecer subsídios mais sólidos ao aperfeiçoamento das políticas educacionais que vêm sendo implementadas. As escalas de proficiência utilizadas nessas avaliações revestem-se de uma grande importância nesse processo, visto que são utilizadas pelos avaliadores e pelos gestores educacionais como um instrumento capaz de aferir e interpretar a proficiência dos alunos nas diversas habilidades que compõem os domínios de conhecimento investigados. O objeto de estudo da presente tese são precisamente as escalas de proficiência utilizadas nos sistemas de avaliação em larga escala, que são tratadas em três diferentes abordagens: primeiro, há uma avaliação das diferentes abordagens para se obter uma associação entre os itens de um teste e os níveis ou pontos significativos de proficiência de diversas escalas atualmente empregadas. Depois, utilizando uma escala de proficiência em Matemática para a quarta série do Ensino Fundamental, é desenvolvida uma abordagem alternativa para caracterizar os níveis representativos de uma escala de proficiência escolar. Para isso foram utilizados métodos baseados em análise de conglomerados (*cluster analysis*) para a escolha do subconjunto de itens que caracteriza cada nível de proficiência. Por fim, é realizada uma interpretação pedagógica dos *clusters* formados, com base nos conhecimentos do ensino e da aprendizagem das habilidades básicas da educação matemática para a primeira etapa do ensino fundamental. A abordagem utilizada será a análise da natureza das operações mentais que caracterizam cada uma das competências e habilidades que compõem os *clusters*, tendo como pressupostos básicos as referências ao processo de aprendizagem da matemática.

Palavras-chave:

Avaliação educacional; escalas de proficiência; análise de conglomerado; níveis de proficiência escolar; interpretação pedagógica de itens.

Abstract

Oliveira, Lina Kátia Mesquita de; Franco Jr, Francisco Creso Junqueira. **Three Investigations of Proficiency Scales and Their Interpretation**. Rio de Janeiro, 2008. 216 p. Tese de Doutorado - Departamento de Educação, Pontifícia Universidade Católica do Rio de Janeiro.

Large-scale evaluations have played a relevant role in the contemporary field of education, not only because of their importance in providing a more precise assessment of the proficiency of large numbers of students, but also because of their use as a means of producing more reliable elements that are used in the improvement of educational policies. The proficiency scales adopted in these evaluation programs are a very important part of this process, for they are employed by educational evaluators and decision-makers as an instrument able to gauge and interpret the students' levels of proficiency in the many different abilities that compose the knowledge domains under investigation. The object of this thesis is precisely the proficiency scales used in large scale evaluation systems, which are treated according to three different approaches: First an investigation is made about the relative merit of different methods of association between the items that make up a test and the significant points or levels of proficiency in several scales that have been employed by different evaluation programs. Then, by using a 4th-grade Mathematics proficiency scale, an alternative approach to characterize levels that are representative of a proficiency scale is developed. This is done by the use of a cluster analysis method aimed at choosing a subset of items that characterize each proficiency level. Finally, a pedagogical interpretation of the clusters formed along the proficiency scale is obtained through the consideration of teaching and learning of Mathematics education basic skills, related to the beginning series of the fundamental level. The methodology used in this process is the analysis of the nature of mental operations that characterize each one of the competencies and skills that are typical of each cluster, by adopting references to Mathematics learning processes as basic assumptions.

Key-words:

Educational evaluation; proficiency scales; cluster analysis; school proficiency levels; pedagogical interpretation of items.

Sumário

1	Introdução	17
2	Avaliação das diferentes abordagens para a associação entre itens e níveis ou pontos de uma escala de proficiência.....	22
2.1	Conceitos básicos: dos itens à Teoria da Resposta ao Item	24
2.1.1	As principais características dos itens utilizados nas avaliações de larga escala	26
2.1.2	Aspectos básicos da Teoria da Resposta ao Item (TRI)	27
2.1.3	Os parâmetros da TRI	30
2.2	As Escalas de Proficiência	34
2.2.1	Considerações gerais	34
2.2.2	Transformações lineares	36
2.2.3	Dando significado às Escalas de Proficiência	37
2.2.4	Construção da Escala de Proficiência: critérios de seleção de itens representativos	39
2.3	Principais referências de abordagens para a associação entre itens e níveis ou pontos característicos de uma escala de proficiência	40
2.3.1	O SAEB	40
2.3.1.1	Objetivo e ciclo de avaliação	40
2.3.1.2	A seleção de itens representativos utilizada pelo SAEB a partir de 1999	41

2.3.1.3	A seleção de itens representativos utilizada pelo SAEB em 1995 e 1997	44
2.3.2	Análise comparativa entre os critérios de seleção utilizados pelo SAEB a partir de 1999 e os utilizados em 1995 / 1997	46
2.4	O NAEP: objetivo e ciclo de avaliação	49
2.4.1	A ancoragem de itens utilizada pelo NAEP	50
2.4.2	Análise comparativa entre os critérios de seleção utilizados pelo SAEB a partir de 99 e o NAEP	59
2.5	O TIMSS: objetivo e ciclo de avaliação	60
2.5.1	A seleção de itens representativos utilizada pelo TIMSS	62
2.5.2	Considerações sobre a seleção dos itens segundo os critérios do TIMSS	65
2.5.3	Análise comparativa entre os critérios de seleção utilizados pelo SAEB a partir de 99 e o TIMSS	66
2.6	O PISA: objetivo e ciclo de avaliação	67
2.6.1	A seleção de itens representativos utilizada pelos PISA	68
2.6.2	Análise comparativa entre os critérios de seleção utilizadas pelo SAEB a partir de 1999 e os critérios de interpretação de escala do PISA	73
2.7	O Projeto GERES: Objetivos e ondas de aplicação	75
2.7.1	Metodologia de construção da escala de proficiência adotada pelo Projeto GERES	76
2.7.2	Análise comparativa entre os critérios de seleção utilizados pelo SAEB a partir de 99 e o GERES	80
2.7.2.1	Comparações entre o SAEB 1999-2007 e o GERES no ponto de auge do desenvolvimento de habilidade	81
2.7.2.2	Comparações entre o SAEB 1999-2007 e o GERES no ponto de consolidação da habilidade	83
3	O uso da análise de conglomerados na seleção de um conjunto de itens característicos dos níveis de proficiência ...	92
3.1	Análise de <i>clusters</i> : considerações gerais	93
3.1.1	Formação dos agrupamentos	93

3.2	Construção de <i>clusters</i> para a seleção de itens característicos dos níveis da escala de proficiência	97
3.2.1	A seleção de itens	97
3.2.2	O agrupamento inicial dos itens selecionados	98
3.2.3	Características gerais do agrupamento de itens após a obtenção da solução com 6 <i>clusters</i>	109
4	Interpretação pedagógica da Escala de Proficiência a partir dos resultados dos <i>clusters</i>	118
4.1	Espaço e Forma	121
4.2	Medidas e Grandezas	124
4.3	Números e Operações	126
4.4	Tratamento da Informação	128
4.5	Interpretação Pedagógica dos <i>clusters</i> de proficiência	129
5	Considerações finais	139
6	Referências bibliográficas	146
	Anexos	148

Lista de tabelas

Tabela 1	Número de itens selecionados por nível de proficiência, segundo os critérios do SAEB 1999-2007	43
Tabela 2	Número de itens selecionados por nível de proficiência, segundo os critérios do SAEB 1995-1997	45
Tabela 3	Número de itens-âncora por nível de proficiência conforme o primeiro critério do NAEP	57
Tabela 4	Número de itens-âncora por nível de proficiência conforme o segundo critério do NAEP.....	58
Tabela 5	Pontos selecionados na escala TIMSS	62
Tabela 6	Níveis selecionados na escala TIMSS	63
Tabela 7	Correspondência entre os níveis de proficiência do TIMSS e do SAEB nos níveis de seleção definidos pelos critérios do TIMSS	64
Tabela 8	Número de itens selecionados por nível de proficiência conforme os três critérios do TIMSS considerados separadamente	65
Tabela 9	Número de itens selecionados por nível de proficiência conforme os três critérios do TIMSS tomados conjuntamente	66
Tabela 10	Níveis e intervalos de proficiência no PISA	70
Tabela 11	Níveis de proficiência no PISA	72

Tabela 12	Número de itens selecionados de acordo com cada um dos três critérios do PISA	73
Tabela 13	Disparidade entre os níveis de seleção segundo os critérios do SAEB (1999-2007) e do GERES (no auge do desenvolvimento da habilidade)	82
Tabela 14	Número de itens selecionados por nível de proficiência, segundo os critérios do GERES (no auge do desenvolvimento da habilidade)	82
Tabela 15	Disparidade entre os níveis de seleção segundo os critérios do SAEB (1999-2007) e do GERES (consolidação da habilidade)	84
Tabela 16	Número de itens selecionados por nível de proficiência, segundo os critérios do GERES (na consolidação da habilidade)	85
Tabela 17	Resumo dos 11 <i>clusters</i> originais	104
Tabela 18	Os 6 <i>clusters</i> formados	110
Tabela 19	Distribuição dos alunos pelos níveis de habilidades	117

Lista de quadros

Quadro 1	Posicionamento dos itens na escala/ critérios de seleção do SAEB - 1999 A 2007	42
Quadro 2	Posicionamento dos itens na escala/ critérios de seleção do SAEB - 1995 a 1997	44
Quadro 3	Critérios/Posicionamento dos itens na Escala de Proficiência - NAEP	53
Quadro 4	Posicionamento dos itens na escala de proficiência/critérios de seleção - TIMSS	64
Quadro 5	Avaliação das três fases de desenvolvimento das habilidades	77
Quadro 6	Distribuição dos alunos da 4ª série EF - Prova Brasil 2005 ..	97
Quadro 7	Níveis de habilidade e proficiências características	116
Quadro 8	Variação observada entre os diferentes programas associada à quantidade de itens selecionados por cada critério	140
Quadro 9	Amplitude dos níveis de seleção constituídos de forma contínua, ou descontínua	141
Quadro 10	Adoção de relaxamentos dos critérios de seleção	141
Quadro 11	Critério para a Seleção dos Itens Característicos dos Níveis de Proficiência	142

Lista de figuras, gráficos e diagramas

Figura 1	ICIs de quatro itens de diferentes níveis de dificuldade, segundo o modelo de um parâmetro da TRI	29
Figura 2	CCIs de quatro itens com variações de dificuldade e de discriminação, segundo o modelo de dois parâmetros da TRI	31
Figura 3	CCIs de quatro itens com variações de dificuldade e de discriminação, levando-se em conta também o acerto casual, segundo o modelo de três parâmetros da TRI	33
Figura 4	A curva normal de proficiência	35
Figura 5	Representação gráfica de Beaton e Allen	51
Figura 6	Item 13099	55
Figura 7	Item 25099	56
Figura 8	Fases de desenvolvimento das habilidades	77
Figura 9	Exemplo hipotético de Curva Característica do Item	78
Figura 10	Curva Característica do Item e Curva de Informação do Item (CII) para um item testado no GERES	79
Figura 11	Dendograma	95
Figura 12	Aplicação de <i>k-means</i>	96
Figura 13	Exemplo de uma Curva Característica do Item	99
Gráfico 1	Número de itens selecionados segundo os diferentes critérios considerados	88

Gráfico 2	CCI's dos itens 25471 e 26325	89
Gráfico 3	Itens 25205 e 24967	90
Gráfico 4	Itens com seus respectivos pontos notáveis próximos, o que produz uma distância euclidiana menor e torna-os similares ..	101
Gráfico 5	Itens com seus respectivos pontos notáveis mais distantes entre si e, conseqüentemente, menos similares	101
Gráfico 6	Distribuição do número de itens ao longo dos 11 <i>clusters</i> formados	105
Gráfico 7	Distribuição dos itens ao longo dos 11 <i>clusters</i> formados .	105
Gráfico 8	Distribuição dos 11 <i>clusters</i> formados	106
Gráfico 9	Representação bidimensional das centróides dos 11 <i>clusters</i> formados e seus respectivos itens	108
Gráfico 10	Distribuição do número de itens ao longo dos 6 <i>clusters</i> formados	110
Gráfico 11	Distribuição dos itens ao longo dos 6 <i>clusters</i> formados ..	111
Gráfico 12	Distribuição dos 6 <i>clusters</i> formados	112
Gráfico 13	Representação bidimensional das centróides dos 6 <i>clusters</i> formados e seus respectivos itens	114
Diagrama 1	Escala de Proficiência em Matemática - Primeira etapa do Ensino Fundamental	119