

APLICAÇÃO DO MODELO DE CRÉDITO PARCIAL GENERALIZADO NA AVALIAÇÃO DO PROJETO SESI - POR UM BRASIL ALFABETIZADO

*Eduardo de São Paulo, Betina Silvestri Miranda, José Gonçalves Moreira Neto y
Luiz Alexandre Rodrigues da Paixão*

1. INTRODUÇÃO

A elevação do nível de alfabetismo é vista por uma grande quantidade de países como fator crucial para o desenvolvimento humano e econômico. A melhoria da alfabetização exerce papel central no desenvolvimento, contribuindo para a prosperidade econômica, a obtenção de bons níveis de saúde, a reafirmação da identidade cultural e a participação cívica, levando os indivíduos a explorarem seus potenciais. Para os indivíduos, a falta de alfabetização adequada pode resultar em pobreza, desvantagens competitivas e, principalmente, exclusão social.

Medir o nível de alfabetismo não é somente apontar quem pode ou não ler. Habilidades de alfabetização funcional, como leitura, escrita e matemática são necessárias em diferentes níveis, desde escrever o nome em um formulário até compreender instruções em uma embalagem de remédio ou fazer cálculos. A mensuração destas habilidades constitui-se em uma importante ferramenta de planejamento e monitoramento de programas de alfabetização.

Neste sentido, são realizados, sistematicamente, levantamentos que visam mensurar as habilidades envolvidas no processo de alfabetização. O NALS (National Adult Literacy Survey) e o LAMP (Literacy Assessment and Monitoring Programme) constituem-se em duas iniciativas internacionais. Esta última é desenvolvida pelo Instituto de Estatísticas da UNESCO, situado no Canadá, em parceria com várias agências internacionais e especialistas técnicos. O principal objetivo é desenvolver e conduzir uma pesquisa para medir níveis de alfabetismo da população adulta em países em desenvolvimento, que não possuem essas informações.

No Brasil, a principal experiência é o INAF (Indicador de Alfabetismo Funcional). Desenvolvido pelo Instituto Paulo Montenegro —ligado ao IBOPE— o INAF é um indicador que revela os níveis de alfabetismo funcional da população adulta brasileira. O principal objetivo do INAF é oferecer informações qualificadas sobre as habilidades e práticas de leitura, escrita e matemática dos brasileiros entre 15 e 64 anos de idade.

O Programa Brasil Alfabetizado, em atividade desde 2003, é uma iniciativa do Governo Federal para alfabetizar brasileiros com idade superior a 14 anos e representa um portal de entrada para a cidadania, articulado diretamente com o aumento da escolarização de jovens e adultos e promovendo o acesso à educação como um direito de todos em qualquer momento da vida. Vários são os parceiros que se engajaram nesta difícil tarefa de combate ao analfabetismo e a ação em conjunto tornou-se importante para estimular a continuidade da formação no ensino de jovens e adultos. O SESI é um dos maiores parceiros deste Programa no combate ao analfabetismo, e visa o desenvolvimento pessoal do aluno, tornando-o capaz de se integrar a novas situações e em condições de qualificar-se permanentemente.

Com o objetivo de elevar a escolaridade de trabalhadores com idade superior a 14 anos, o Projeto SESI – Por um Brasil Alfabetizado adota metodologias de ensino adaptadas às necessidades do

aluno adulto, tornando-o um cidadão consciente e, em parceria com instituições públicas e privadas, oferece cursos de alfabetização em unidades escolares do SESI e outros espaços da comunidade. O Projeto é uma iniciativa dos dirigentes da indústria que verificaram a necessidade de ampliação e fortalecimento do atendimento ao trabalhador da Indústria Brasileira fundamentada na preocupação com a modernização de máquinas e procedimentos e com a construção de bases para atuar com responsabilidade social.

O Sistema de Avaliação de Competências consiste de um conjunto de práticas e procedimentos de caráter externo aos programas educacionais do SESI e tem como papel primordial sua retroalimentação, por meio da coleta e análise sistemática de dados que possibilitem a elaboração de informações e indicadores. Este Sistema teve sua criação proposta pela UNESCO, em 1999, e acompanha as mais modernas tendências metodológicas em uso em sistemas de avaliação em diversos países no mundo.

Como o INAF e o LAMP, o Sistema de Avaliação de Competências propõe-se a mensurar as habilidades de leitura, escrita e numerização. Para tal, o Sistema utiliza um método estatístico denominado de Teoria de Resposta ao Item (TRI). A TRI conduz a medidas muito mais informativas do que as baseadas em pontuações brutas, além de oferecer maior facilidade na interpretação das escalas de medidas das habilidades cognitivas e uma melhor análise de cada item, fornecendo elementos para a construção de testes cada vez mais informativos e mais adequados para o grupo que será avaliado.

No mês de março de 2004, o Sistema de Avaliação de Competências, sob a assistência da Coordenação de Alfabetização do Departamento Nacional do SESI – COALFA e dos Departamentos Regionais do SESI nos estados de Goiás, Pará, Pernambuco e Rio de Janeiro e no Distrito Federal, promoveu uma experiência piloto para aplicação de metodologias de avaliação em nível de alfabetização.

Essa experiência se constituiu na primeira fase de pré-testagem de metodologias, em cujo escopo se inclui a definição da sistemática de aplicação mais adequada à avaliação do Projeto SESI - Por um Brasil Alfabetizado. Ademais, tal aplicação possibilitou a observação, à luz da TRI, do comportamento estatístico dos itens apresentados aos alunos. Tais informações são fundamentais para a seleção de itens adequados aos testes a serem aplicados a estudantes deste nível de ensino.

Como característica, os itens elaborados para esta avaliação puderam receber diferentes pontuações em seu processo de correção, de acordo com o nível de desenvolvimento das respostas auferidas (itens politômicos). Surge daí a necessidade da utilização de um modelo elaborado para respostas graduadas que possibilite a obtenção da estimativa de dificuldade e discriminação dos itens.

É cada vez mais freqüente o uso dos modelos da TRI para análise de itens e estimação da proficiência dos alunos. A experiência mais conhecida em nível nacional é o SAEB – Sistema Nacional de Avaliação da Educação Básica, que desde 1995 utiliza a Teoria de Resposta ao Item —mais especificamente o modelo logístico de três parâmetros— em suas análises.

Méndez, Inglês e Hidadalgo (2002) comprovaram a adequação do Modelo de Crédito Parcial Generalizado aos dados de 799 indivíduos que responderam a 30 itens referentes a um questionário de confiança para falar em público. Além de verificar a adequação do modelo, os autores apresentaram informação acerca da precisão do questionário para diferentes níveis de ansiedade de falar em público. Flores-Mendoza *et al* (2005) utilizou o modelo de resposta graduada para modelar dados referentes ao Desenho da Figura Humana na prática de avaliação psicológica de crianças no Brasil. A autora demonstra haver problemas de dificuldade e discriminação de grupo de itens por meio da utilização do modelo politômico. Por fim, Gempp, Denegri, Caprile *et al* (2006) mensuraram a alfabetização

econômica de crianças entre 10 e 15 anos de idade a partir do Modelo de Crédito Parcial (MCP). Em todos estes estudos, ficou patente a adequação dos modelos politômicos aos dados e sua contribuição na avaliação da qualidade dos itens elaborados.

Este artigo utiliza o Modelo de Crédito Parcial Generalizado (MCPG) —teoricamente adequado para lidar com itens de natureza politômica, extraíndo-lhes a maior informação possível— para modelar os dados de um conjunto de aproximadamente 2.800 alunos concluintes do Projeto SESI - Por um Brasil Alfabetizado. A proposta do presente trabalho é validar a metodologia de análise avaliando-se a adequação do modelo teórico aos dados empíricos. É objetivo também apresentar a análise estatística dos itens e a curva de informação do teste com o intuito de avaliar-se a qualidade dos itens elaborados. Esse aspecto é importante na medida em que estes itens serão utilizados em rodada posterior de avaliação e certamente comporão a escala de proficiência em alfabetização, a ser implementada. O pacote estatístico utilizado para trabalhar com o modelo proposto foi o PARSCALE.

Na seção 2, apresentada-se um breve detalhamento da Teoria da Resposta ao Item e sua importância no contexto do estudo. A seção 3 apresenta as medidas utilizadas para verificar a adequação do modelo proposto. A análise estatística de alguns itens selecionados é apresentada na seção 4, enquanto que as curvas de informação dos testes, que identificam os níveis de proficiência nos quais os testes são particularmente informativos, são apresentadas na seção 5.

2. METODOLOGIA

2.1. A teoria de Resposta ao Item e suas aplicações

Em muitas situações nas áreas de Educação e Psicologia faz-se necessária a mensuração de uma variável de interesse, que no caso da vertente de alfabetização do Sistema de Avaliação de Competências é definida como competência em letramento¹. Tal medida torna-se extremamente útil no acompanhamento e na avaliação de conjunto(s) de indivíduos no que se refere a sua evolução no processo ensino-aprendizagem, bem como na efetividade das propostas pedagógicas delineadas pelo Projeto SESI – Por um Brasil Alfabetizado. Essa variável, apresentada como competência em leitura e numerização, recebe na Psicometria a denominação genérica de traço latente². Portanto, uma meta inicial do Sistema de Avaliação é a determinação da “quantidade” (mensuração) do traço latente que os alfabetizando possuem.

Os testes usados em avaliação educacional são compostos de itens que avaliam o domínio de um conjunto de habilidades. Este conjunto de habilidades caracteriza a competência que se quer medir e é usualmente apresentado na forma de uma matriz de competência. Em seguida, parte-se para a preparação do instrumento de medida. A partir da montagem do instrumento de medida, idealmente, devem-se produzir evidências de sua qualidade. Para isso, é fundamental que os itens sejam pré-testados, o que permite a identificação de problemas de construção. Busca-se, por meio da utilização de técnicas psicométricas adequadas, que viabilizem a validação dos itens e conseqüentemente do teste, uma representação fidedigna do traço latente que se deseja avaliar.

¹ Segundo Tfouni (2002), competência em letramento é “...compreendida como a procura em estudar e descrever o que ocorre nas sociedades quando adotam um sistema de escritura de maneira restrita ou generalizada; procura ainda saber quais práticas psicossociais substituem as práticas “letradas” em sociedades ágrafas. Desse modo, o letramento tem por objetivo investigar não somente quem é alfabetizado, mas também quem não é alfabetizado, e, nesse sentido, desliga-se de verificar o individual e concentra-se no social”.

² Segundo Pasquali (2003), a Psicometria trabalha com o conceito fatorista do traço latente, que o define como o conjunto de processos cognitivos necessários para a execução de uma tarefa.

Nesse contexto, a Psicometria procura compreender as competências e habilidades adquiridas pelas respostas dadas pelos sujeitos a uma série de tarefas, tipicamente denominadas de itens, utilizando-se de modelos estatísticos adequados à problemática inerente a esse tipo de situação.

As técnicas estatísticas usualmente utilizadas para modelagem de dados oriundos dessa situação são: Teoria Clássica dos Testes (TCT) e Teoria de Resposta ao Item (TRI). Com características distintas, essas duas técnicas são capazes de estimar a proficiência de cada sujeito e de produzir importantes informações sobre cada item de um teste, verificando sua adequação quanto aos objetivos propostos. Dificuldade e discriminação são alguns, talvez os mais relevantes, parâmetros relacionados aos itens apresentados por esses dois modelos paramétricos.

Os resultados provenientes das duas teorias são apresentados neste estudo de forma a se complementarem. No entanto, é importante salientar que existem diferenças teóricas consideráveis entre as duas abordagens. Enquanto a teoria clássica se preocupa em formar testes válidos (em contraposição com a abordagem por itens) e se mostra dependente da amostra de sujeitos, a TRI constrói itens válidos capazes de constituir vários testes conseqüentemente válidos e aplicados a vários sujeitos.

Cada item do teste é um identificador da competência do sujeito avaliado. Uma pessoa de alta competência tem grande probabilidade de acertar um item difícil e, por sua vez, uma pessoa de baixa competência tem pequena probabilidade de acertá-lo. Essa relação entre o escore no item e a competência do examinando é descrita por funções cujo conjunto é denominado de Teoria de Resposta ao Item. Com base em algumas características, a TRI vem sendo amplamente utilizada na análise e evolução dos testes. Em geral, ela assume que existe uma função de regressão não linear que descreve a relação entre a probabilidade de dar uma resposta correta ao item e o nível de habilidade do respondente. Denominada Curva Característica do Item (CCI), esta função se constrói em toda escala de traço latente que, geralmente, varia de -3 a 3, com média 0 e desvio padrão 1. Assim, valores negativos e próximos de -3 indicam baixo nível do traço, enquanto valores positivos e próximos a 3 sugerem altos níveis. A TRI permite obter ainda índices de precisão do item (Função de informação do item) e do teste (Função de informação do teste), importantes ferramentas para seleção de itens e construção de testes a serem elaborados para determinado nível do traço latente dos sujeitos.

Além disto, e como a sua mais importante característica quando comparados à TCT, esses modelos permitem a comparabilidade dos escores estimados para os alunos em momentos distintos do tempo, mesmo quando instrumentos de avaliação parcialmente diferentes são empregados. Este procedimento, denominado de equalização, pode ser empregado aos dados aqui apresentados, uma vez que itens de ligação foram incluídos nos instrumentos de avaliação desta e da subsequente etapa de avaliação.

2.2. Modelo utilizado

A TRI proporciona uma vasta quantidade de modelos, produzindo diferentes formas de CCI, que permitem trabalhar com diferentes formatos de resposta, como por exemplo, itens dicotômicos e itens politômicos.

Utilizam-se itens politômicos no intuito de apresentar uma maior quantidade de informação sobre as habilidades, pela mensuração de competências ou habilidades que estejam parcialmente desenvolvidas. Além disso, busca-se controlar os efeitos da escolha ao acaso, problema apresentado principalmente pelos itens de múltipla escolha.

Em geral, podem-se separar os modelos para respostas politômicas em dois grupos: os modelos de resposta que assumem uma certa ordem, conhecidos como de respostas graduadas, e os de resposta nominais, que, em princípio, não apresentam uma ordem entre as respostas. Entre os primeiros encontram-se os modelos de crédito parcial de Masters (1982) e o de crédito parcial generalizado de Muraki (1992).

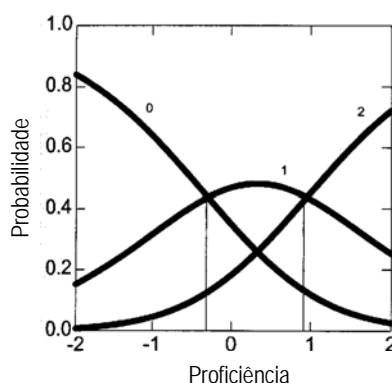
Um dos modelos de respostas graduadas é o de crédito parcial, introduzido por Masters em 1982 para a análise de itens que não apresentam somente as categorias correto e errado ou verdadeiro e falso. O item é pontuado parcialmente e o respondente ganha mais créditos na medida em que sua resposta se aproxima da completa. Porém, uma limitação deste modelo é que ele não permite que o parâmetro de discriminação varie para cada item, tendo apenas o parâmetro de dificuldade considerado para a caracterização do item com relação aos fatores que influenciam no desempenho do respondente.

O modelo de crédito parcial generalizado desenvolvido por Muraki em 1992 é uma generalização do primeiro e permite que os itens se distingam pelo parâmetro de discriminação. Possui os mesmos pressupostos do modelo logístico de dois parâmetros, no qual os parâmetros de dificuldade e de discriminação são considerados como fatores que influenciam a escolha das respostas e caracterizam os respondentes, proporcionando uma grande flexibilidade quando se trabalha com itens politômicos. Sendo assim, de acordo com o modelo de crédito parcial generalizado, a probabilidade de resposta é expressa pela função:

$$P_{ix}(\theta) = \frac{\exp \sum_{j=0}^x Da_i(\theta - b_{ij})}{\sum_{r=0}^{m_i} [\exp \sum_{j=0}^r Da_i(\theta - b_{ij})]}$$

Esta função permite a incorporação de itens com respostas politômicas (com m categorias), onde a e b são os parâmetros de discriminação e dificuldade, respectivamente. O parâmetro b é composto por outros dois parâmetros, o parâmetro de localização do item e o parâmetro de interseção da categoria (o ponto de interseção de duas curvas de categorias de respostas adjacentes). Este modelo indica que a probabilidade do aluno responder a categoria de resposta m , é dada como a probabilidade condicional sobre as categorias $(m-1)$ e m . Em resumo, o modelo explicita os pontos na escala de proficiência onde uma categoria de resposta se torna relativamente mais provável que a resposta precedente dado que o respondente completou os passos anteriores.

FIGURA 1. CURVA DE RESPOSTAS PARA UM ITEM COM 3 CATEGORIAS



A figura 1 mostra um item com três categorias (0, 1 e 2). O respondente ganha créditos (pontuações) na medida em que completa as respostas atingindo a pontuação máxima do item (categoria 2) indicando que possui totalmente a habilidade medida.

Conforme a figura acima, o indivíduo com proficiência maior que 1 tem mais chance de acertar completamente o item. Uma vez que o respondente não possui a habilidade total e, sim, habilidade intermediária, ele atinge até a categoria 1 e ganha um crédito por ter respondido parcialmente à questão proposta, indicando que a habilidade em questão deve ser mais trabalhada para este grupo de pessoas. Nesse sentido, indivíduos que não possuem alguma habilidade (valores mais baixos na escala de proficiência) tem uma probabilidade alta de não conseguirem avançar nas categorias (não passando da categoria 0) e, assim, não obtêm pontuação alguma na questão. Ou seja, indivíduos com baixa habilidade possuem maior probabilidade de não alcançarem alguma pontuação e aqueles com maior habilidade possuem maior probabilidade de pontuação no item.

3. ADEQUAÇÃO DOS MODELOS

Como primeiro objetivo do estudo, a verificação do ajustamento do MCPG aos dados foi feita por meio da estatística qui-quadrado (χ^2). O programa utilizado proporciona esta estatística de ajuste para cada item mediante a qual se pode estabelecer o nível de significância estatística das discrepâncias entre as frequências de escolha das opções teóricas e as observadas para cada nível de habilidade.

Entretanto, no caso de pesquisas de larga escala com respondentes de muitos lugares diferentes o valor do χ^2 sofre efeitos relativos aos agrupamentos (cluster) podendo seu valor ser aumentado em função disto, levando a conclusões equivocadas a favor do não ajuste dos itens. Para eliminação destes efeitos é recomendável dividir o valor do χ^2 por uma constante igual a 2 ou 3. Neste estudo, o valor do χ^2 foi dividido por 3. As tabelas 1 e 2 apresentam os valores do χ^2 e os graus de liberdade associados para todos os itens de Linguagens e Códigos e Numerização. Uma razão χ^2 /gl inferior a três é um indício de boa adequação do item ao modelo.

TABELA 1. ESTATÍSTICAS DE AJUSTE DO MODELO – LINGUAGENS E CÓDIGOS

Item	χ^2	g.l.	χ^2 / g.l.
1	20,18	9	2,24
2	4,84	5	0,97
3	164,52	32	5,14
4	27,54	9	3,06
5	24,48	24	1,02
6	0,59	7	0,08
7	54,13	22	2,46
8	18,21	18	1,01
9	4,86	7	0,69
10	3,31	8	0,41
11	45,77	21	2,18
TOTAL	368,44	162	2,27

TABELA 2. ESTATÍSTICAS DE AJUSTE DO MODELO – NUMERIZAÇÃO

Item	χ^2	g.l.	$\chi^2 / \text{g.l.}$
12	15,83	9	2,26
13	9,26	5	1,32
14	3,45	32	0,49
15	8,12	9	1,01
16	70,61	24	3,53
17	23,49	7	2,94
18	13,18	22	1,65
19	2,54	18	0,42
20	20,08	7	2,51
21	29,59	8	3,70
22	33,91	21	4,24
TOTAL	230,06	162	2,42

Pode-se verificar a partir da análise das tabelas acima que somente dois itens de Linguagens e Códigos e três itens de numerização parecem não ajustarem-se bem ao modelo proposto. Quando são analisadas as estatísticas gerais dos modelos, conclui-se que ambos apresentam um bom ajuste global aos dados.

4. ANÁLISE DOS ITENS

4.1. Parâmetros dos itens

A tabela 3 apresenta os parâmetros estimados e seus respectivos erros padrão segundo o modelo de crédito parcial generalizado de Muraki (1992) para cada um dos itens utilizados. Os primeiros onze itens da prova são aqueles utilizados para auferir a proficiência dos alunos em Linguagens e Códigos, enquanto os itens de 12 a 22, em Numerização. O parâmetro ‘a’ representa a discriminação (que, em geral, varia de 0 a 2) associada ao item e o parâmetro ‘b’ (que varia de -3 a 3) representa a sua dificuldade. O número de créditos parciais na tabela representa as categorias de pontuação que o aluno poderia obter em cada item. Sendo assim, se este número for 4, quer dizer que, para este item, a pontuação varia de 0 (erro), 1 (25% de acerto), 2 (50% de acerto), 3 (75% de acerto) a 4 (100% de acerto). A faixa de dificuldade indica em que nível o item pode ser considerado para aquela população, se está muito fácil (faixa I), fácil (faixa II), mediano (faixa III), difícil (faixa IV) ou muito difícil (Faixa V).

Os itens, em geral, apresentam uma discriminação média de 0,840 para Linguagens e Códigos e 1,133 para Numerização. O item mais discriminativo para Linguagens foi o item 2 ($a=1,456$) e os itens 10 e 11 apresentam também uma discriminação acima de um. Para a prova de numerização, o item 14 ($a=1,531$) é o que mais distingue os alunos, sendo que o item 16 ($a=0,649$) é o menos discriminativo. Com estas informações podemos diferenciar os alunos que possuem a habilidade medida pelo item dos que ainda precisam ter a habilidade trabalhada.

Com relação ao parâmetro b dos itens, a dificuldade média é de -0,724 e -0,761 para as provas de Linguagens e Códigos e Numerização, respectivamente. Destaca-se o item 2 da prova de Linguagens e Códigos como o mais fácil ($b=-1,676$) e o item 3 como um de dificuldade mediana ($b=-0,292$). Neste contexto, para a prova de Numerização, o item 18 se destaca como mais fácil ($b=-$

1,483), enquanto o item 21 apresenta-se como o mais difícil dessa prova ($b = 0,202$). Assim, em testes futuros, pode-se balanceá-los escolhendo itens fáceis e difíceis para sua composição.

TABELA 3. PARÂMETROS DE DISCRIMINAÇÃO (A), PARÂMETRO DE DIFICULDADE (B) E FAIXA DE DIFICULDADE E Nº DE CRÉDITOS PARCIAIS EM CADA ITEM.

Item	Nº créditos parciais	TRI				Faixa de dif.
		a	erro a	b	erro b	
1	1	0,637	0,037	-0,346	0,048	III
2	1	1,456	0,09	-1,676	0,059	I
3	4	0,336	0,011	0,292	0,033	III
4	1	0,791	0,044	0,152	0,039	III
5	4	0,56	0,015	-0,888	0,031	II
6	1	0,98	0,058	-1,603	0,067	I
7	3	0,569	0,019	0,449	0,03	III
8	3	0,845	0,036	-1,346	0,04	I
9	1	0,906	0,054	-1,57	0,07	I
10	1	1,007	0,052	-0,942	0,045	II
11	4	1,152	0,053	-0,482	0,025	III
12	1	1,076	0,059	-1,468	0,059	I
13	1	1,279	0,066	-0,83	0,04	II
14	1	1,531	0,081	-0,926	0,037	II
15	1	1,262	0,068	-0,503	0,035	III
16	3	0,649	0,026	-0,863	0,035	II
17	1	0,993	0,052	-0,393	0,036	III
18	1	0,788	0,046	-1,483	0,073	I
19	1	1,886	0,112	-0,958	0,036	II
20	1	0,666	0,04	-1,186	0,068	II
21	1	1,324	0,073	0,202	0,031	III
22	1	1,004	0,055	0,038	0,034	III

As tabelas 4 (itens de Linguagens e Códigos) e 5 (itens de Numerização), a seguir, apresentam os dados estatísticos apurados para cada um dos itens da avaliação selecionados, a saber: (1) o número de créditos parciais de cada item; (2) a porcentagem de acerto em cada uma das possibilidades; (3) a correlação bisserial do item com o rendimento geral dos alunos; (4) os parâmetros da TRI de discriminação e dificuldade dos itens com sua respectiva faixa de alocação que categoriza a dificuldade do item em níveis. Em seguida, será descrita a habilidade exigida para que o aluno acertasse o item.

4.2. Análise de itens selecionados de linguagens e códigos

TABELA 4. SUMÁRIO DAS ESTIMAÇÕES DOS PARÂMETROS DOS ITENS 9 E 11 DE LINGUAGENS E CÓDIGOS

Item	Nº de créditos parciais	% de acerto					Bisserial	TRI		
		0	1	2	3	4		a	b	Faixa de dificuldade
9	1	14,8	85,3	-	-	-	0,501	0,906	-1,57	I
11	4	15,7	11,8	10,9	16,2	45,4	0,773	1,152	-0,482	III

Item 9

Habilidade avaliada: Buscar informações para o desempenho social na lista telefônica, nos manuais, nos formulários, em entrevistas, cadastros e outros.

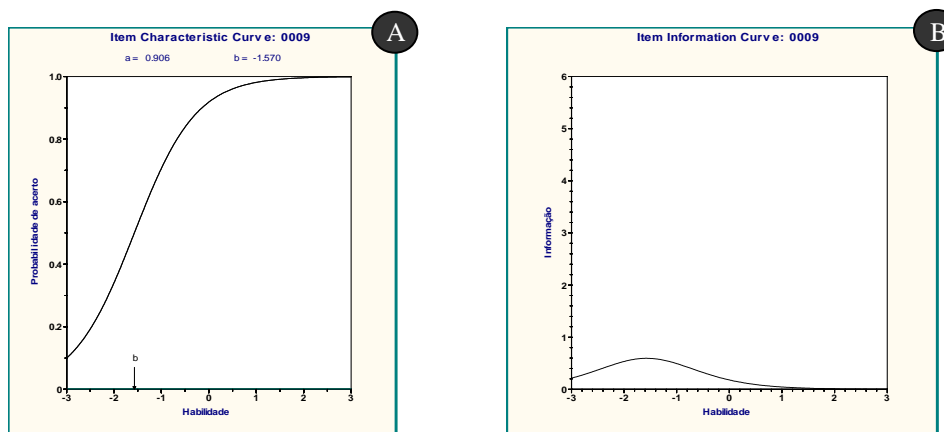
Você e sua família resolveram mudar-se para uma casa nova. Vocês querem uma casa com três quartos. Circle o anúncio que melhor atende à necessidade de sua família.

FIGURA 2. ITEM 9



Guia de Marcação: UM ponto para o anúncio circulado corretamente.

FIGURA 3. CURVA CARACTERÍSTICA (CCI) E CURVA DE INFORMAÇÃO (CII) DO ITEM 9



Este item foi respondido corretamente por aproximadamente 85% dos alunos, como pode ser verificado na tabela 4. Pode-se, então, considerá-lo um item muito fácil, o que é confirmado pelo parâmetro b (mostrado na tabela 4) e pela observação do eixo horizontal da figura 3 (A). Note que como o parâmetro de dificuldade b varia numa escala de -3 a 3, sendo que quanto mais próximo de -3, menor a dificuldade do item, o valor obtido de -1,570 traduz a “facilidade” dessa questão. Quanto à sua discriminação, analisando-se a correlação bisserial (que representa a correlação entre o resultado no item e a sua pontuação total no teste), percebe-se que os alunos que acertaram o item, de forma geral, alcançaram altas pontuações no teste, indicando uma boa discriminação do item. Este fato é confirmado quando se analisa o parâmetro a da TRI. Como seu valor varia normalmente na escala de 0

a 2, sendo que quanto mais próximo de 2 mais discriminante é o item, o valor estimado de 0,906 confirma a boa discriminação desta questão. A curva de informação (figura 3 (B)) demonstra que este item apresenta uma melhor informação para alunos com níveis de proficiência situados entre -2,5 e -0,5.

Item 11

Habilidade avaliada: Compreender ou interpretar textos não-literários.

FIGURA 4. ITEM 11

QUAL É O ANIMAL MAIS ALTO?

A GIRAFA TEM COMO CARACTERÍSTICA PRINCIPAL A ALTURA. É O MAIS ALTO DE TODOS OS ANIMAIS. QUANDO ADULTA, MEDE POR VOLTA DE 5 METROS. COMO SEU PESCOÇO É MUITO COMPRIDO, ELA CONSEGUE COMER AS FOLHAS E AS FRUTAS QUE FICAM NO ALTO DAS ÁRVORES. AS GIRAFAS VIVEM NA ÁFRICA.

Complete a ficha com as informações sobre o animal.

NOME DO ANIMAL: Girafa

CARACTERÍSTICA PRINCIPAL: Altura

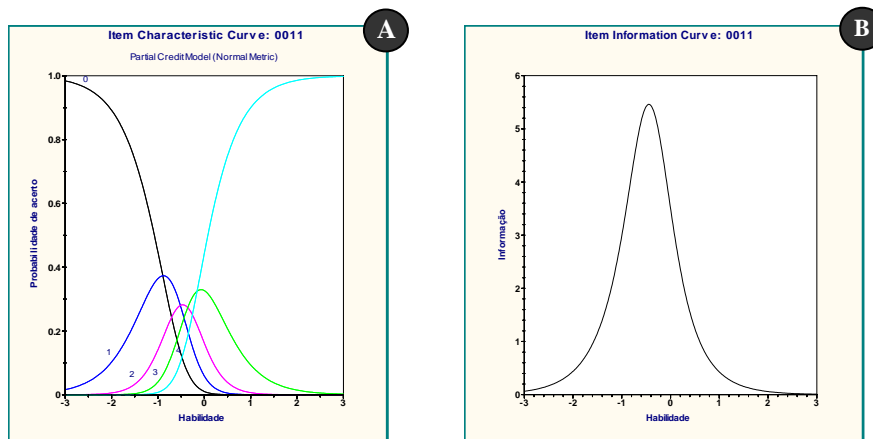
ALIMENTAÇÃO: Folhas e frutas

ONDE VIVE: África

Guia de Marcação: UM ponto para cada resposta corretamente preenchida. Pontuação máxima de 4 pontos.

Observação: Se o aluno colocar só “folhas” ou só “frutas” não lhe será atribuída pontuação no campo Alimentação.

FIGURA 5. CURVA CARACTERÍSTICA (CCI) E CURVA DE INFORMAÇÃO (CII) DO ITEM 11



Neste item, que permitia acertos parciais, 45,4% dos alunos acertaram completamente a questão. Em contrapartida, cerca de 15,7% erraram e 38,9% obtiveram acertos parciais no item. Com base nessa informação, juntamente com a análise do parâmetro *b* (-0,482), é possível concluir que o item apresenta uma dificuldade mediana (faixa III). Em relação à sua discriminação, o item pode ser considerado discriminante, distinguindo os alunos que possuem a habilidade necessária para a resolução da questão daqueles que não a possuem ou que ainda necessitam aprimorar a habilidade requerida pelo item (acertos parciais). Ressalta-se o significativo valor da correlação bisserial (0,773), corroborando a análise acima referente à discriminação do item.

A CCI deste item (figura 5 (A)) apresenta as cinco curvas referentes aos créditos parciais recebidos pelos alunos (0 “errou ou não respondeu”; 1 “25% de acerto”; 2 “50% de acerto”; 3 “75% de acerto”; e 4 “acerto completo”). Analisando-se os pontos de inflexão destas curvas, onde é maximizada a probabilidade de acerto, verifica-se que a proficiência necessária para a obtenção de pontuações mais altas aumenta de curva para curva. Portanto, percebe-se claramente que, à medida que o nível de proficiência do aluno aumenta, a probabilidade de obtenção de um melhor resultado cresce. A curva de informação (figura 5 (B)) reflete uma ótima acurácia para níveis de proficiência intermediários (-2 a 1, aproximadamente).

4.3. Análise de itens selecionados de numerização

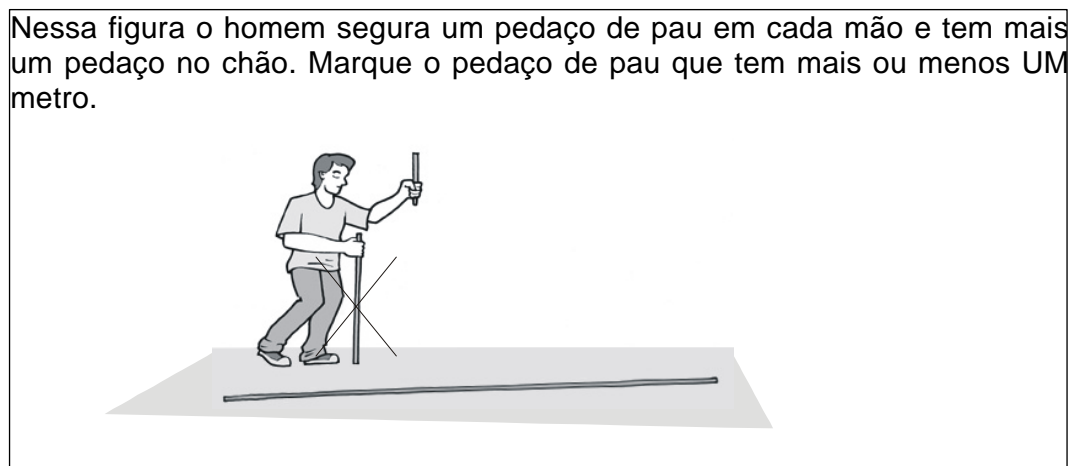
TABELA 5. SUMÁRIO DAS ESTIMAÇÕES DOS PARÂMETROS DOS ITENS 20 E 22 DE NUMERIZAÇÃO

Item	Nº de créditos parciais	% de acerto					Bisserial	TRI		
		0	1	2	3	4		a	b	Faixa de dificuldade
20	1	25,2	74,8	-	-	-	0,455	0,666	-1,186	II
22	1	50,6	49,4	-	-	-	0,586	1,004	0,038	III

Item 20

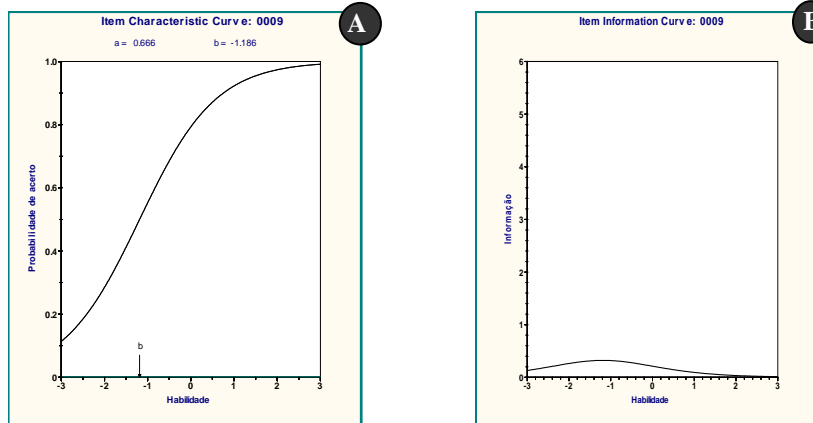
Habilidade avaliada: Identificar a dimensão real das principais unidades de medida.

FIGURA 6. ITEM 20



Guia de marcação: conferir UM ponto para a resposta correta.

FIGURA 7. CURVA CARACTERÍSTICA (CCI) E CURVA DE INFORMAÇÃO (CII) DO ITEM 20



Este item pode ser considerado fácil (faixa II) e discriminativo. Cerca de 75% dos alunos acertaram este item com um parâmetro $b=-1,186$ e $a=0,666$. A Curva Característica (figura 7 (A)) apresenta uma relação diretamente proporcional entre o nível de proficiência e a probabilidade de acerto da questão, ou seja, quanto maior a proficiência maior a probabilidade de acerto. Este item dá maior informação para alunos que possuam níveis de proficiência inferiores a 0, como pode ser observado na figura 7 (B).

Item 22

Habilidade avaliada: Coletar, organizar e interpretar dados e informações.

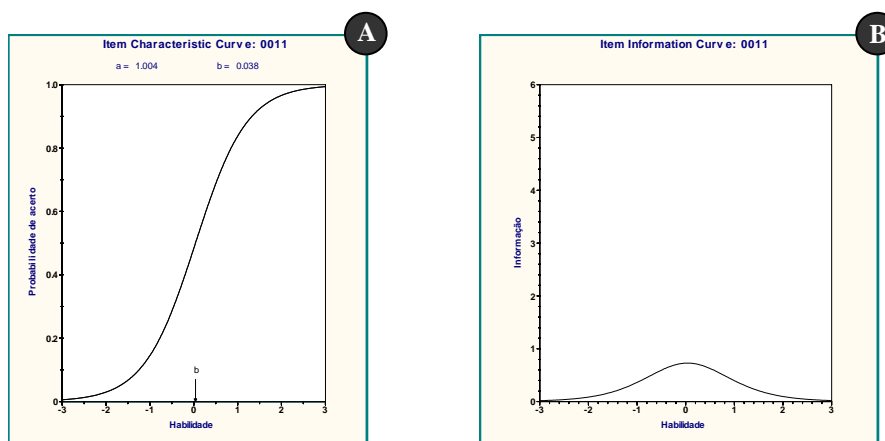
FIGURA 8. ITEM 22

Sebastião foi à lanchonete com a namorada e comprou um suco e um pastel para cada um. Observe a tabela de preços e escreva no espaço quanto Sebastião gastou.

LANCHONETE	
PRODUTO	PREÇO
Suco	R\$ 1,40
Refrigerante	R\$ 1,00
Água	R\$ 0,80
Pastel	R\$ 1,10
Coxinha	R\$ 1,20
Pão de queijo	R\$ 1,00

Guia de marcação: Conferir UM ponto para a resposta correta.

FIGURA 9. CURVA CARACTERÍSTICA (CCI) E CURVA DE INFORMAÇÃO (CII) DO ITEM 22



Este item apresenta dificuldade mediana, uma vez que 49,4% dos alfabetizandos solucionaram o item corretamente, o que é corroborado ao observarmos o valor do parâmetro b (0,038). Verifica-se, ainda, que o item apresenta boa discriminação ($a=1,004$). A curva de informação (figura 9 (B)) indica que o item fornece mais informação para os sujeitos cuja proficiência encontra-se em níveis intermediários da escala.

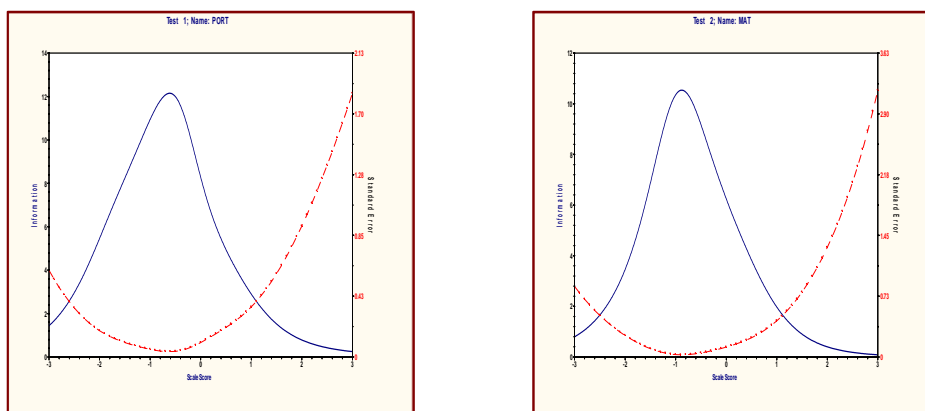
5. CURVAS DE INFORMAÇÃO DOS TESTES

As curvas de informação dos testes (CIT) possibilitam indicar os níveis de proficiência onde o teste é mais discriminativo, ou seja, em que faixas o teste é particularmente fidedigno³ e para as quais ele não o é. Além disso, proporciona uma medida global de erro padrão para cada nível de proficiência. Em outras palavras, a CIT indica para cada nível de proficiência a quantidade de imprecisão esperada para a estimativa do desempenho do sujeito. Assim, a curva de informação constitui-se em uma forma de orientar a aplicação de um instrumento de avaliação e decidir em que situações e para que propósitos esse será mais adequado.

A figura 10 apresenta a curva de informação para os dois testes aplicados. Pode-se observar que o teste de Linguagens apresenta mais informação para sujeitos com proficiência estimada entre -1 e 0, ao passo que há um acréscimo do erro padrão da estimativa de proficiência para níveis da escala acima de 1. Este teste, portanto, contribui muito pouco para discriminar sujeitos com alta proficiência. Conclusão similar pode ser auferida observando-se a curva de informação do teste de Numerização; a diferença é que para Numerização o teste é ligeiramente mais informativo para sujeito com baixa proficiência. Cabe ressaltar que o teste com informação alta apresenta uma variância baixa.

³ A fidedignidade ou precisão de um teste diz respeito à característica que ele deve possuir de medir sem erros, daí os nomes precisão, confiabilidade ou fidedignidade. Medir sem erros significa que o mesmo teste, medindo os mesmos sujeitos em ocasiões diferentes, ou testes equivalentes, medindo os mesmos sujeitos na mesma ocasião produzem resultados idênticos. Dentro da TCT o coeficiente de fidedignidade é definido estatisticamente como a correlação entre os escores dos mesmos sujeitos em duas formas paralelas de um teste. Similarmente, o conceito de informação na TRI está relacionado com os níveis de proficiência para os quais o teste apresenta os menores erros-padrão, ou seja, a maior precisão.

FIGURA 10. CURVA DE INFORMAÇÃO DOS TESTES APLICADOS (CIT)



6. Conclusões

As provas de Linguagens e Códigos e de Numerização aplicadas em alunos concluintes do curso de alfabetização do Projeto SESI – Por um Brasil Alfabetizado mostraram um ajuste adequado do modelo de crédito parcial generalizado.

A análise estatística possibilitou a estimação dos parâmetros de discriminação e dificuldade dos itens utilizados. Itens com problemas referentes a estes parâmetros podem ser retirados ou reformulados para aplicação dos testes da próxima etapa de avaliação. Posteriormente, com a replicação deste estudo, será possível a montagem de um banco de itens referente à alfabetização, com itens devidamente testados e calibrados. A constituição deste banco é de extrema importância, pois permitirá uma grande flexibilidade na elaboração de testes de qualidade.

A utilização da TRI permite a criação de testes adaptados aos sujeitos, uma vez que fornece informação suficiente em relação a que níveis de proficiência os itens e o teste são mais precisos. As provas das duas disciplinas relatadas neste artigo podem ser aplicadas a uma população com proficiência um pouco abaixo de níveis intermediários da escala de proficiência.

Este estudo foi o primeiro passo para a produção de uma escala de proficiência de alfabetização. A possibilidade de comparação futura dos escores produzidos aqui e aqueles que serão estimados na próxima etapa de avaliação está garantida pela inserção de itens de ligação entre os instrumentos das duas avaliações. Este fato acarretará em um maior número de itens com boas características que constituirão a escala de proficiência o que, certamente, aumentará a sua confiabilidade.

Finalmente, é importante ressaltar a importância deste estudo no incentivo à utilização destas técnicas para a análise de dados provenientes de avaliações educacionais, principalmente no que tange à alfabetização de jovens e adultos, nível de ensino ainda carente de estudos desta natureza.

REFERENCIAS BIBLIOGRÁFICAS

Beaton, A.E. y Gonzalez, E.J. (1995). Reporting Achievement in Mathematics and Science Content Areas. En: *Timss Technical Report Volume II: Implementation and Analysis (Primary and Middle School Years)*. Massachusetts: TIMSS International Study Center.

- Childs, R.A. y Chen, W.H. (1999), Obtaining comparable item parameter estimates in Multilog and Parscale for tow polytomous IRT models. *Applied Psychological Measurement*, 23 (4), pp. 371-379.
- Embretson, S.E. y Reise, S.P. (2000). *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Flores-Mendonza, Carmen E., Abad, F.J. y Lelé, A.J (2005). Análise de Itens do Desenho da Figura Humana: Aplicações de TRI. *Psicologia: Teoria e Pesquisa*, 21(2), pp. 243-254.
- Gempp, R., Denegri, M., Caprile, C., *et al* (2006). Medición de la Alfabetización Económica em Niños: Oportunidades Diagnósticas com el Modelo de Crédito Parcial. *Psyche*, 15(1), pp13-27.
- Hidalgo, M.D., López-Pina, J.A, Inglés, C.J. y Méndez, F.X. (2002). Análisis psicométrico Del Cuestionario de confianza para hablar em público, usando la teoría d respuesta al ítem. *Anales de Psicología*, 18(2), pp. 333-349.
- Muraki, E. y Bock, R. D. (1997). *Parscale: IRT Item Analysis and Test Scoring for Rating-scale Data*. Chicago: Scientific Software International.
- Pasquali, L. (2004). *Psicometria: Teoria dos Testes na Psicologia e na Educação*. Petrópolis: Vozes.
- Pasquali, L. (no prelo). *Análise Fatorial para Pesquisadores*.
- Soares, J. F. y Júdice, R. (2004). A medida da competência matemática no estudo do analfabetismo funcional. *Letramento no Brasil*. Org. M.C.F.R. Fonseca. São Paulo: Global.
- Soares, M. (2003). Letramento e escolarização. En: Ribeiro, V.M. (Org.). *Letramento no Brasil*. São Paulo: Global.
- Tfouni, L.V. (2002). *Letramento e alfabetização*. São Paulo: Cortez.