A CONSTRUÇÃO E A INTERPRETAÇÃO DAS ESCALAS DE CONHECIMENTO – CONSIDERAÇÕES GERAIS E UMA VISÃO DO QUE VEM SENDO FEITO NO SARESP[1]

Raquel da Cunha Valle
Mestre em Estatística pela Universidade de São Paulo – USP
Estatístico do Departamento de Pesquisas Educacionais da Fundação Carlos Chagas, São Paulo - SP

Resumo

O artigo mostra que, com o aumento do interesse dos educadores brasileiros na Teoria de Resposta ao Item – TRI, em virtude da sua utilização cada vez maior em nossas avaliações, surge a questão da compreensão e da utilização dos resultados que vêm sendo divulgados, em especial no que se refere às escalas de habilidade. Este trabalho tem como objetivo principal fornecer uma visão geral do processo de construção das escalas de habilidade, esclarecendo algumas dúvidas básicas, como, por exemplo, como são definidos os pontos da escala, como é estabelecida a distância entre eles, como cada ponto é caracterizado e como deve ser interpretado, entre outros aspectos. Apresenta, ainda, um exemplo prático do que vem sendo feito no caso do SARESP.

[1] Artigo apresentado na 27ª Conferência Anual da IAEA (International Association for Educational Assessment), 6 a 11 de maio de 2001, Rio Othon Palace Hotel, Rio de Janeiro, Brasil.

1. Introdução

Atualmente, vem crescendo o interesse dos educadores brasileiros na Teoria da Resposta ao Item – TRI em virtude da sua aplicação em avaliações nacionais, como o **SAEB – *Sistema Nacional de Avaliação da Educação Básica*** e, também, em avaliações regionais em larga escala, como, por exemplo, o **SARESP – *Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo***, onde esta técnica vem sendo utilizada desde 1995 e 1996, respectivamente[2].

No entanto, a compreensão e a utilização dos resultados que vêm sendo divulgados ainda parecem estar muito aquém do esperado, e justamente entre a clientela para a qual essas informações seriam mais relevantes – professores, diretores, supervisores de ensino, etc. Inúmeros pontos referentes tanto ao processo de construção das escalas de habilidade quanto à sua interpretação precisam ser ainda exaustivamente discutidos.

Sem dúvida cabe ao estatístico conduzir o processo de elaboração das escalas de habilidade, no que se refere à sua parte matemática, mas a presença de profissionais da área educacional é imprescindível em praticamente todas as suas etapas. Portanto, é fundamental que tais profissionais estejam adequadamente preparados para esta tarefa, assim como para um aproveitamento, cada vez mais intensivo, das informações obtidas.

Neste trabalho, teremos como objetivo principal fornecer uma visão geral do processo de construção das escalas de conhecimento, esclarecendo algumas dúvidas básicas, como, por exemplo, como são definidos os pontos da escala, como é estabelecida a distância entre eles, como cada ponto é caracterizado e como deve ser interpretado, etc. Nossa intenção é que esses objetivos sejam alcançados através de um exemplo prático do que vem sendo feito no caso do SARESP.

2. A escala de habilidade

A utilização da TRI nas avaliações educacionais vem possibilitando uma série de avanços em termos do acompanhamento do desenvolvimento escolar que antes não eram possíveis. Hoje, pode-se avaliar o rendimento escolar de uma determinada série ou verificar

[2] A autora agradece à Secretaria de Estado da Educação de São Paulo (SEE), pelo uso parcial dos resultados do SARESP 96, SARESP 97 e SARESP 98.

se houve ganho de uma série para outra por intermédio da comparação de resultados de provas diferentes aplicadas em populações distintas, desde que haja itens comuns entre as provas, para que se possa realizar uma equalização.

Equalizar significa equiparar, tornar comparável, o que, no caso da TRI, significa colocar parâmetros de itens vindos de provas distintas ou habilidades de respondentes de diferentes grupos, na mesma métrica, isto é, numa escala comum, tornando os itens e/ou as habilidades comparáveis.

Mas, uma vez feita uma equalização, ou seja, uma vez que todos os resultados desejados são comparáveis, o próximo passo seria atribuir um significado prático aos valores obtidos. Saber que, na 3ª série do ensino fundamental, os alunos têm habilidade média em matemática 100 e que na 4ª série essa habilidade é de 150, já nos fornece uma informação quantitativa de que, na 4ª série, os alunos tiveram um ganho de 50% em relação ao conhecimento em matemática na série anterior, mas, e qualitativamente, o que eles sabem a mais em termos de conteúdo ? E o que sabiam na 3ª série ?

Com essa finalidade é que se constrói uma escala de habilidade: para buscar uma interpretação qualitativa dos valores obtidos. Mas, assim como existe uma teoria matemática que possibilita a obtenção dessas habilidades, também existe uma metodologia matemática e todo um trabalho de interpretação pedagógica na construção de uma escala de habilidade. E são alguns desses pontos que pretendemos abordar aqui.

2.1 Alguns pontos importantes

Para começar, devemos ter em mente que, para se construir uma escala, é necessária um quantidade suficiente de itens a fim de que se possa caracterizar/interpretar cada ponto da mesma.

Se desejamos que a escala tenha vários níveis, isto significa que devemos trabalhar com diferentes níveis de habilidade, o que, por sua vez, pode significar que deveremos ter diferentes séries envolvidas.

Mas, para diferentes séries, teremos diferentes provas, e para construir qualquer escala de habilidade, o primeiro passo é sempre que todos os itens estejam numa mesma escala. Logo, deveremos ter alguns itens comuns entre as provas das diferentes séries ou, então, provas de "ligação" entre elas.

3. Passos para a construção de uma escala de habilidades

Primeiramente, iremos descrever os passos para a construção e interpretação de uma escala de habilidades. A seguir, ilustraremos esses passos descrevendo o procedimento adotado e os resultados obtidos no SARESP.

3.1 Definição das séries e disciplinas a serem estudadas.

Uma escala será elaborada para cada disciplina, mas o ideal é que seja para todas as séries envolvidas no estudo. Para tanto, é necessário que todos os itens estejam numa mesma métrica.

Deve-se planejar com bastante critério as populações que serão estudadas, pois deve haver um número suficiente de indivíduos avaliados em cada população. Por outro lado, uma vez que quanto maior a diferença entre as habilidades dos indivíduos avaliados mais pontos poderá ter a escala, seria desejável trabalhar com indivíduos de diferentes séries. No entanto, uma vez que será necessário equalizar todos os itens envolvidos, trabalhar com séries muito distantes pode ser um problema, pois, em algum momento, um grupo de alunos deverá responder itens que sejam das séries anteriores, para que se possa fazer a equalização, e se esses itens forem fáceis demais para esses alunos e o índice de acertos for quase total, o processo de equalização será prejudicado.

3.2 Elaboração e aplicação dos instrumentos (provas)

Não podemos esquecer que a escala é caracterizada pelos itens, logo, a qualidade da escala depende da qualidade dos itens. Além disso, sempre devemos ressaltar que a quantidade de itens envolvidos deve ser suficiente tanto para caracterizar bem cada ponto da escala, quanto para possibilitar que a escala possa ter vários níveis. Basicamente, será o bom planejamento e a boa execução dos passos 3.1 e 3.2 que definirão a qualidade da escala. Uma boa escala de habilidade será fruto da aplicação de um número razoável de itens de boa qualidade (com altos níveis de discriminação, diferentes níveis de dificuldade, etc), em um número suficiente de indivíduos com os mais variados níveis de habilidade.

3.3 Eqüalização

Para que seja feita a equalização, há a necessidade de itens comuns entre as diferentes provas ou populações. Pode-se adotar os mais diferentes planos experimentais, envolvendo diferentes séries, resultando em diferentes tipos de equalização, mas o que se busca é que, ao final do processo, todos os itens que formarão a escala devem estar numa mesma métrica.

3.4 Definição da escala

Após a calibração e a equalização dos itens, devido aos procedimentos matemáticos e aos recursos computacionais utilizados, tanto os parâmetros dos itens quanto as habilidades dos alunos estarão numa escala que pode ser pouco conveniente em termos práticos. Em geral, os programas computacionais utilizam a escala (0;1) , em que a média dos valores obtidos é 0 e o desvio padrão (variabilidade) é 1. Sendo assim, grande parte dos valores resultantes são negativos e há a necessidade de se trabalhar com números com várias casas decimais.

Logo, por praticidade e também para facilitar o entendimento, é usual que se defina uma escala mais conveniente. Escolhe-se, por exemplo, um valor para a habilidade média de uma das populações ou, então, se define que a escala deve variar apenas num determinado *range* de valores. Uma vez definida a escala, faz-se uma transformação linear em todos os valores originais, para colocá-los na escala desejada.

É comum se trabalhar com escalas que variam de 0 a 100, mas é importante que fique claro que esses valores serão habilidades e não "porcentagens de acerto", confusão bastante comum em escalas com esse tipo de variação. Por isso, muitas vezes é mais aconselhável definir escalas em intervalos de variação bem distintos, por exemplo, com média 200 ou 500, que não apresentam valores negativos ou o valor zero, que também costumam levam a equívocos do tipo "alunos com habilidade nula ou negativa".

3.5. Escolha dos níveis âncora

Uma vez que o *range* de variação da escala está definido, o próximo passo é definir seus níveis âncora. Os níveis âncora são os pontos da escala que serão interpretados pedagogicamente. Esses

pontos são caracterizados por um conjunto de itens, denominados de itens âncora, que são conjuntos de itens que apresentam determinadas propriedades matemáticas. Tais propriedades estão relacionadas com características do item, tais como índice de discriminação e de dificuldade, e serão apresentadas a seguir. Assim, não se pode caracterizar todos os pontos da escala e a escolha da distância entre seus pontos âncora também é importante. Se escolhermos níveis âncora muito próximos, não conseguiremos encontrar itens âncora para caracterizá-los, ou seja, não será possível encontrar itens que satisfaçam às condições matemáticas necessárias. Por outro lado, escolhendo níveis âncora muito distantes, teremos poucos níveis, e a escala será uma escala "pobre", pedagogicamente falando.

Está é, sem dúvida, uma tarefa que requer paciência e bom senso e muitas vezes chegamos na melhor escolha dos níveis âncora por tentativa e erro. Mas, em geral, pode-se tomar como base a média e o desvio padrão de uma das populações em estudo. Uma boa escolha é definir a média de tal população como um dos níveis âncora, e definir a distância entre eles como sendo algum múltiplo do desvio padrão desta população, por exemplo, meio desvio, um desvio, um desvio e meio, etc.

3.6. Identificação dos itens âncora

Depois de estabelecidos os níveis âncora da escala, o próximo passo é identificar os itens que caracterizam cada um destes níveis, ou seja, identificar os itens âncora. Para que um item seja âncora em determinado nível, ele deve satisfazer a certas condições matemáticas, que são dadas a seguir.

Definição de item âncora:

Considere dois níveis âncora consecutivos Y e Z, com $Y < Z$. Dizemos que um determinado item é <u>âncora para o nível Z</u> se, e somente se, as três condições abaixo forem satisfeitas simultaneamente:

1. $P(X=1/\theta=Z) \geq 0,65$
2. $P(X=1/\theta=Y) < 0,50$
3. $P(X=1/\theta=Z) - P(X=1/\theta=Y) \geq 0,30$

Em outras palavras, para um item ser âncora em um determinado nível âncora da escala, ele precisa ser respondido

corretamente por uma grande proporção de indivíduos (pelo menos 65%) com este nível de habilidade e por uma pequena proporção de indivíduos (no máximo 50%) com o nível de habilidade imediatamente anterior. Além disso, a diferença entre a proporção de indivíduos com esses níveis de habilidade que acertam a esse item deve ser de pelo menos 30%. Assim, para um item ser âncora ele deve ser um item *típico* daquele nível, ou seja, bastante acertado por indivíduos com aquele nível de habilidade e pouco acertado por indivíduos com um nível de habilidade imediatamente inferior.

3.7 Interpretação de cada ponto da escala por especialistas

Após identificar matematicamente o conjunto de itens âncora em cada nível da escala, o passo seguinte será realizado por especialistas na área de conhecimento em questão. Caberá a esses especialistas caracterizar cada ponto da escala, a partir do estudo do conteúdo abordado no conjunto de itens que definem cada nível âncora.

Nesse momento, a escala está finalmente pronta para ser utilizada.

4. A utilização da escala

Mesmo após a construção da escala de habilidade, ainda existem algumas dúvidas sobre qual a maneira correta de utilizá-la. O primeiro passo seria posicionar às populações que foram estudadas e verificar em que pontos da escala elas se encontram, utilizando sua habilidade média. Assim, a primeira informação que a escala nos fornecerá é a identificação do que os alunos sabem e do que não são capazes de fazer, ou seja, quais os conteúdos que eles dominam e em quais conteúdos ainda precisam melhorar. Outra informação interessante é a porcentagem de alunos de cada população distribuída em cada faixa de habilidade. A partir dessa informação, pode-se verificar, por exemplo, qual a porcentagem de alunos de uma determinada série que domina os conteúdos abordados naquele nível de habilidade e como essa porcentagem evolui de uma série para outra.

5. Considerações finais

Mesmo observando todos os passos descritos até aqui, podem ocorrer problemas na construção ou na interpretação de uma escala de habilidades. Por exemplo, os níveis extremos da escala, referentes às habilidades mais baixas e às mais altas, são, de modo geral, mal caracterizados, por serem definidos, respectivamente, por itens muito fáceis ou muito difíceis, que em geral, são poucos.

Outro problema que freqüentemente ocorre, é que, nos níveis extremos superiores da escala, há poucos alunos, isto é, é possível *interpretar pedagogicamente um nível da escala, mas uma* porcentagem muito baixa dos alunos avaliados domina os conhecimentos descritos nesse nível.

6. Um dos principais sistemas de avaliação brasileiros: o SARESP

O SARESP é um sistema de avaliação considerado modelo, dentre as avaliações regionais. Aplicado em 1996[3] (3ª e 7ª séries do Ensino Fundamental), em 1997 (4ª e 8.ª séries do Ensino Fundamental) e em 1998 (5ª série do Ensino Fundamental e 1ª série do Ensino Médio) em todas as escolas públicas estaduais do Estado de São Paulo. Em 1999 não houve aplicação. Já em 2000 foram avaliadas três séries: a 5ª e a 7ª séries do Ensino Fundamental e a 3ª série do Ensino Médio.

6.1 Uma breve descrição das características do SARESP

As provas do SARESP são elaboradas a partir de matrizes curriculares, ou seja, tabelas de especificação de conteúdos e objetivos, que indicam os temas e metas do currículo a serem desenvolvidos em cada série e disciplina. Esses parâmetros fundamentam-se nas Propostas Curriculares elaboradas pela Coordenadoria de Estudos e Normas Pedagógicas – CENP e, desde 1997, os itens que compõem as provas vêm sendo construídos pelos professores da Rede Estadual de Ensino.

Em 1996 iniciou-se a utilização da TRI no SARESP. Em 1999 não houve SARESP, mas em 2000 o estudo foi retomado, incorporando *novas características. Assim, descreveremos neste trabalho algumas das*

[3] Primeiro ano de aplicação em que foram utilizadas técnicas derivadas da TRI na análise dos resultados.

características do SARESP referentes às aplicações feitas em 1996, 1997 e 1998.

A aplicação das provas nesse período foi feita segundo o quadro a seguir.

Tabela 1

Esquema da aplicação das provas do SARESP no período de 1996 a 1998

Ano de aplicação	Séries e períodos avaliados	Provas feitas nas disciplinas
1996	3ª série diurna do Ensino Fundamental	1-Língua Portuguesa e 2-Matemática
	7.ª série diurna e noturna do Ensino Fundamental	1-Língua Portuguesa, 2-Matemática, 3-Ciências e 4-História e Geografia
1997	4ª série diurna do Ensino Fundamental	1-Língua Portuguesa e 2-Matemática
	8.ª série diurna e noturna do Ensino Fundamental	1-Língua Portuguesa, 2-Matemática, 3-Ciências e 4-História e Geografia
1998	5ª série diurna e noturna do Ensino Fundamental	1-Língua Portuguesa e 2-Matemática
	1ª série diurna e noturna do Ensino Médio	1-Língua Portuguesa, 2-Matemática, 3-Ciências e 4-História e Geografia

Como as avaliações foram sempre realizadas no início do ano letivo, as provas de cada uma das séries-alvo são baseadas em conteúdos abordados no ano anterior. Exemplificando, em 1996, as provas dos alunos da 3ª e 7ª séries foram elaboradas com base nos conteúdos relativos ao Ciclo Básico e à 6ª série, respectivamente.

Em todos os anos foram avaliados todos os alunos que freqüentavam as séries envolvidas: trata-se, portanto, de uma avaliação de caráter censitário. Cada aluno, entretanto, é avaliado em apenas uma disciplina, ou seja, na 3ª , 4ª e 5ª séries metade dos alunos responde à prova de Língua Portuguesa e a outra metade, à de Matemática. Essa divisão é feita de maneira aleatória. Nas demais séries, os alunos são divididos, também aleatoriamente, em 4 partes e então cada uma delas é submetida a um tipo de prova: Língua Portuguesa, Matemática, Ciências ou História e Geografia. Essa última prova é a única onde aparecem duas disciplinas. No entanto, em termos de análise, as duas disciplinas são obviamente consideradas separadamente.

Como resultado desses três anos de aplicação, foram obtidas 4 escalas de habilidade: uma escala de habilidade em língua portuguesa e outra em matemática para a 3ª, 4ª e 5ª séries do ensino fundamental,

e uma escala de habilidade em língua portuguesa e outra em matemática para a 7ª e 8ª séries do ensino fundamental e 1ª série do ensino médio.

Para as demais disciplinas não foram construídas escalas pois apenas língua portuguesa e matemática foram avaliadas todos os anos em todas as séries. Além disso, como os itens de 3ª, 4ª e 5ª séries puderam ser colocados em uma métrica e os itens de 7ª, 8ª e 1ª séries em outra, foram construídas 2 escalas separadas para os dois conjuntos de itens.

Para ilustrar o processo de construção e interpretação de uma escala de habilidades e exemplificar o tipo de resultados obtidos, vamos descrever neste trabalho como foi feita a Escala de Habilidades em Matemática do SARESP, desenvolvida a partir dos itens aplicados na 3ª, 4ª e 5ª séries do ensino fundamental.

7. O processo de construção da escala de habilidades em matemática do SARESP da 3ª, 4ª e 5ª séries do ensino fundamental

7.1 Definição da séries e disciplinas a serem estudadas

As séries e disciplinas avaliadas no SARESP, no período de 1996 a 1998, já foram descritas anteriormente, mas desse ponto em diante vamos manter nossa atenção voltada somente às séries e à disciplina que usaremos para exemplificar o processo, ou seja, a disciplina de matemática na 3ª, 4ª e 5ª séries do ensino fundamental.

Assim, no nosso caso, como o período noturno só foi avaliado a partir da 5ª série, consideramos 4 populações distintas, e o número de alunos que foi submetido às provas de matemática nessas populações é dado na tabela a seguir.

Tabela 2
Número de alunos que fizeram provas de matemática em cada ano de aplicação

Série	Número de alunos
3ª série diurna de 1996	290594
4ª série diurna de 1997	270265
5ª série diurna de 1998	269942
5ª série noturna de 1998	13937

Como podemos observar, no SARESP não enfrentamos problemas em relação ao número de alunos avaliados.

7.2. Elaboração e aplicação dos instrumentos

Foram elaboradas provas com 30 itens cada para cada uma das 4 populações descritas. Em 1998, havia 7 itens comuns entre as provas do diurno e do noturno. Além disso, 2 itens da prova da 5ª série noturna foram anulados. Logo, a escala de matemática foi desenvolvida a partir de um total de 111 (= 30 + 30 + 30 + 30 – 7 – 2) itens distintos.

7.3 Equalização

Para que a escala pudesse ser construída, seria necessário que esses 111 itens fossem comparáveis, ou seja, estivessem na mesma métrica. E isto pode ser conseguido através de uma equalização.

Como as provas de um ano para outro não apresentavam itens comuns, a solução encontrada no caso do SARESP foi a criação de provas adicionais, que serviriam de "ligação" entre duas séries consecutivas, uma vez que seriam compostas de itens que haviam sido submetidos a essas duas populações.

Exemplificando, a prova de matemática aplicada em 1997, na 4ª série, não tinham itens comuns com a prova aplicada no ano anterior, na 3ª série. Assim, foi montada uma prova de ligação, composta de itens que haviam sido submetidos à 3ª e à 4ª séries. Essa prova adicional foi aplicada, no final do ano de 1997, a uma amostra de alunos da 3ª série. Cabe ressaltar, que este grupo adicional foi introduzido no estudo com o único objetivo de possibilitar a equalização, não havendo nenhum interesse em estudar o desempenho desta população (3ª série de 1997). O número de alunos que fizeram essa prova de ligação foi apenas o suficiente para atender às exigências da TRI, no que se refere ao número mínimo de sujeitos necessários para obter-se boas estimativas dos parâmetros dos itens.

Também é importante notar que a população escolhida para fazer a prova de ligação foi a 3ª série de 1997, pois, como já foi dito, os itens das provas da 3ª série de 96 e da 4ª série de 97 foram elaboradas com base nos conteúdos dos anos anteriores, ou seja, eram referentes aos conteúdos do Ciclo Básico e da 3ª série,

respectivamente. Como a prova de ligação foi aplicada no final do ano letivo de 1997, a série mais indicada para ser submetida a tal prova era, portanto, a 3ª série.

Assim, todos os 60 itens, respondidos pelos alunos das 3 populações envolvidas, foram então calibrados, simultaneamente, através do programa computacional BILOG-MG.

Dando prosseguimento ao estudo, em 1998, as provas aplicadas na 5ª série do ensino fundamental, nos períodos diurno e noturno, novamente, não tinham itens comuns com as provas dos anos anteriores.

Mais uma vez, foi montada uma prova de ligação, composta de itens utilizados nas provas de três das quatro populações de interesse: 4ª série de 1997, 5ª série diurna de 1998 e 5ª série noturna de 1998. Essa prova foi aplicada, então, a uma amostra de alunos que cursavam a 4ª série em 1998. Essa população adicional também foi introduzida no estudo apenas com o objetivo de possibilitar a equalização.

Cabe ressaltar que a meta agora é colocar os alunos da 3ª série de 96, 4ª série de 1997 e 5ª séries diurna e noturna de 98, todos, na mesma escala. Nessa nova equalização, os itens da 3ª série não precisaram mais entrar na prova de ligação, pois a 3ª e a 4ª séries já haviam sido colocadas na mesma métrica. Na verdade, agora é como se fossemos apenas "colar" a 5ª série nas séries anteriores. Assim, essa segunda equalização foi realizada de uma maneira bastante distinta da primeira. Os itens calibrados no ano anterior foram mantidos fixos durante o processo de estimação e apenas os itens aplicados à 5ª série foram calibrados, resultando, ao final do processo, num conjunto de itens de 3ª à 5ª séries, todos na mesma escala. Novamente, o programa computacional utilizado foi o BILOG-MG.

7.4 Definição da escala

Uma vez que os 111 itens foram equalizados, ou seja, foram colocados na mesma métrica, o próximo passo foi definir o *range* de variação da escala. Primeiramente, foi definido que a série de menor habilidade (que no caso era a 3ª série) teria sua habilidade média fixada em cerca de 50 pontos, com um desvio padrão de cerca de 16 pontos. Dessa maneira, em geral, não seriam esperados valores negativos para as habilidades dos alunos.

Então, foi feita uma transformação linear nas estimativas dos parâmetros dos itens e das habilidades dos alunos. Após essa transformação, os valores obtidos para a média e o desvio padrão das habilidades dos alunos das 4 populações consideradas foram :

Tabela 3
Habilidades médias em matemática obtidas no SARESP

Ano/série	média	desvio padrão
1996 – 3ª série	49,5	16,3
1997 – 4ª série	60,4	16,2
1998 – 5ª série diurna	59,8	14,8
1998 – 5ª série noturna	60,3	15,7

Podemos observar que da 3ª para a 4ª série houve um ganho na habilidade média, mas da 4ª para a 5ª série a habilidade média ficou praticamente inalterada.

7.5 Escolha dos níveis âncora

Com todos os 111 itens na mesma métrica, e uma vez definida a escala, o passo seguinte foi a escolha dos níveis âncora.

Lembrando que a habilidade media de uma das populações (no caso, a 3ª série) era cerca de 50 pontos, com um desvio padrão próximo de 16, optou-se por definir um nível âncora que fosse próximo desse valor, e estabelecer que a distância entre eles deveria ser próxima desse desvio. Assim, um dos níveis âncora estabelecidos foi o 55 e definiu-se que haveria uma distância de 15 pontos entre eles.

7.6 Identificação dos itens âncora

Definidos quais seriam os níveis âncora, foram calculadas as probabilidades descritas na seção 3.6 para todos os 111 itens. Alguns desses itens não puderam ser considerados âncora em nenhum ponto da escala, mas cerca de 1 em cada 3 itens pode ser considerado âncora em algum dos níveis âncora que puderam ser identificados. Assim, foi possível a caracterização de 6 níveis âncora (nos pontos 25, 40, 55, 70, 85 e 100) na escala de habilidades de Matemática da 3ª , 4ª e 5ª séries.

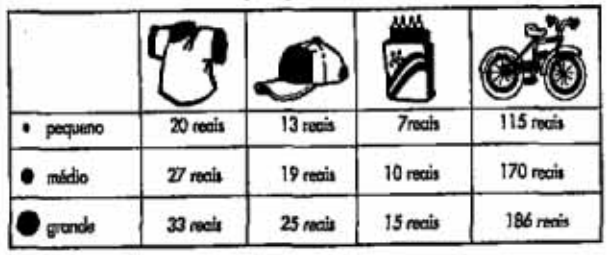
A seguir, apresentamos um exemplo de item que foi identificado como âncora em cada um dos 6 níveis âncora definidos. Juntamente com cada item aparece a habilidade envolvida em sua resolução, descrita por especialistas.

NÍVEL 25

Habilidade	Prova/96 Diurno – item 20
▪ Efetuar operações de multiplicação.	**Calcule:** a) 31 x 2 = b) 42 x 10 = c) 25 x 4

Fonte: Secretaria de Estado da Educação de São Paulo (1999). **Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – SARESP: Descrição das escalas de habilidades do SARESP 96/97/98.** São Paulo: SEE, p. 28.

NÍVEL 40

Habilidade	Prova/96 Diurno – item 23
▪ Revelar familiaridade com atividades que implicam leitura de dados organizados em tabelas, utilizando essa habilidade na solução de problemas do cotidiano.	Observe esta lista de preços  **Qual o preço de uma camiseta grande?** 33 Reais

• pequeno	20 reais	13 reais	7reais	115 reais
● médio	27 reais	19 reais	10 reais	170 reais
● grande	33 reais	25 reais	15 reais	186 reais

Fonte: Secretaria de Estado da Educação de São Paulo (1999). **Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – SARESP: Descrição das escalas de habilidades do SARESP 96/97/98.** São Paulo: SEE, p. 28.

NÍVEL 55

Habilidade	Prova/98 Diurno – questão 7
▪ Comparar números racionais expressos sob notação decimal.	**A tabela a seguir contém as medidas de altura de alguns alunos da 5ª série. Identifique os alunos do mais alto para o mais baixo.** ALUNOS — ALTURAS FLÁVIO — 1,45 metros LEANDRO — 1,50 metros CLÁUDIO — 1,57 metros JOÃO — 1,05 metros JOSÉ — 1,54 metros (A) Cláudio, José, Leandro, Flávio, João. ■ (B) José, João, Cláudio, Leandro, Flávio. (C) Leandro, Cláudio, José, Flávio, João. D) Cláudio, Flávio, João, José, Leandro.

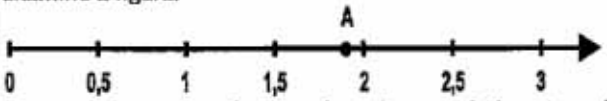
Fonte: Secretaria de Estado da Educação de São Paulo (1999). **Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – SARESP: Descrição das escalas de habilidades do SARESP 96/97/98.** São Paulo: SEE, p. 29.

NÍVEL 70

Habilidade	Prova/98 Diurno – questão 6
▪ Compreender os conceitos de metade e triplo de um número, solucionando situação-problema que envolve os diferentes significados da multiplicação e/ou divisão com números naturais.	**Eu tenho 1.320 figurinhas. Meu primo tem a metade do que tenho. Minha irmã tem o triplo das figurinhas do meu primo. Quantas figurinhas minha irmã tem?** (A) 1.900 (B) 1.930 (C) 1.940 (D) 1.980 ■

Fonte: Secretaria de Estado da Educação de São Paulo (1999). **Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – SARESP: Descrição das escalas de habilidades do SARESP 96/97/98.** São Paulo: SEE, p. 31.

NÍVEL 85

Habilidade	Prova/98 Diurno – questão 8
▪ Localizar, na reta numérica, números racionais na forma decimal.	Examine a figura:  **O ponto A corresponde a um dos números abaixo. A qual deles?** (A) 0,25 (B) 0,85 (C) 1,25 (D) 1,85 ■

Fonte: Secretaria de Estado da Educação de São Paulo (1999). **Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – SARESP: Descrição das escalas de habilidades do SARESP 96/97/98.** São Paulo: SEE, p. 32.

NÍVEL 100

Habilidade	Prova/98 Noturno – questão 4
▪ Reconhecer e aplicar as propriedades das operações como facilitadoras na construção das técnicas operatórias, no exercício da estimativa e do cálculo mental, sem no entanto nomeá-las.	**Em um estacionamento há motos e automóveis, perfazendo um total de 120 rodas. Sabendo-se que 24 veículos são motos, o número de carros é:** (A) 18 ■ (B) 24 (C) 48 (D) 96

Fonte: Secretaria de Estado da Educação de São Paulo (1999). **Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – SARESP: Descrição das escalas de habilidades do SARESP 96/97/98.** São Paulo: SEE, p. 36.

7.7 Interpretação de cada ponto da escala por especialistas

Após a determinação dos níveis âncora e da identificação de seus respectivos itens âncora, um grupo de especialistas nos conteúdos de matemática das séries avaliadas analisou e interpretou o conjunto de itens que compunham cada nível, a fim de caracterizá-los.

A seguir, exemplificamos como ficou a caracterização dos nível âncora da escala de habilidades em Matemática da 3ª , 4ª e 5ª séries do SARESP.

DESCRIÇÃO DA ESCALA DE HABILIDADES DE MATEMÁTICA 3ª, 4ª e 5ª SÉRIES DO ENSINO FUNDAMENTAL

NÍVEL 25

Neste nível, os alunos são capazes de :

- Efetuar operações de multiplicação.

NÍVEL 40

Neste nível, os alunos são capazes de :

- Revelar familiaridade com atividades que implicam leitura de dados organizados em tabelas, utilizando essa habilidade na solução de problemas do cotidiano;
- Solucionar problemas simples, que envolvem as operações de adição e subtração.

NÍVEL 55

Neste nível, os alunos são capazes de :

- Ler e interpretar um esquema, associando-o com a situação proposta, bem como identificar as informações necessárias para, por exemplo, comparar distâncias percorridas em um trajeto representado por desenho figurativo;
- Solucionar problemas concretos simples, que envolvem valor monetário, aplicando a operação de adição com reserva na ordem das dezenas;
- Comparar números racionais expressos sob notação decimal;
- Resolver problemas que implicam tanto leitura de dados organizados em tabelas, como cálculos que requerem a operação de adição.

NÍVEL 70

Neste nível, os alunos são capazes de :

- Utilizar as regras do sistema de numeração decimal para leitura, escrita e comparação de números naturais de qualquer ordem de grandeza;
- Efetuar a divisão exata de um número de 3 algarismos por um de 1 algarismo, demonstrando domínio sobre a multiplicação e a subtração;
- Dominar o conceito de resto;
- Compreender os conceitos de metade e triplo de um número, solucionando situação-problema que envolve os diferentes significados da multiplicação e/ou divisão com números naturais;
- Resolver, via multiplicação, problema que envolve o sistema monetário.

NÍVEL 85

Neste nível, em relação aos temas abaixo discriminados, os alunos são capazes de :

Números – Sistema de Numeração decimal :

- Compreender e utilizar as regras do Sistema de Numeração decimal para leitura e comparação de números racionais escritos na forma decimal, revelando domínio do valor posicional dos algarismos;
- Calcular frações de quantidade;
- Localizar, na reta numérica, números racionais na forma decimal.

Operações :

- Resolver problemas simples do cotidiano, que envolvem mais de uma operação.

Geometria – Medidas:

- Identificar, em um grupo de diversos quadriláteros, os que são losangos.
- Resolver problema envolvendo figuras não-planas.
- Interpretar registros de medidas apresentados por meio de símbolos convencionais, estabelecendo relações entre as unidades usuais de medida de massa;

- Transformar unidades de medida de comprimento;
- Solucionar situação-problema do cotidiano, utilizando conhecimentos a respeito do sistema monetário brasileiro;
- Estabelecer relações entre unidades usuais de medidas de capacidade;

Estatística:

- Interpretar dados ou informações em representações gráficas, para resolver situação-problema;
- Identificar a porcentagem como uma fração de denominador 100;
- Interpretar dados apresentados em gráfico de colunas, para resolver uma situação-problema;
- Revelar familiaridade com a leitura de dados apresentados em forma de tabela, resolvendo problemas mais complexos, que exigem mais de uma operação.

NÍVEL 100

Neste nível, os alunos são capazes de :

- Reconhecer e aplicar as propriedades das operações como facilitadoras na construção das técnicas operatórias, no exercício da estimativa e do cálculo mental, sem, no entanto, nomeá-las;
- Comparar e ordenar números racionais expressos na representação fracionária de uso mais freqüente, como meios, terços, quartos e décimos;
- Compreender a representação decimal dos números racionais, comparando números representados com diferentes quantidades de casas decimais;
- Resolver problema que envolve raciocínio combinatório, chegando a determinar sua solução por representações diversas.

Em relação aos temas específicos abaixo relacionados, os alunos desse nível são capazes de:

Geometria – Medidas:

- Resolver problemas que envolvem medida de comprimento;
- Descrever e interpretar a representação da movimentação de um objeto no plano cartesiano;

- Calcular a área de regiões determinadas por paralelogramos, triângulos ou trapézios por redução ao retângulo equivalente, utilizando a composição e a decomposição.

Estatística:

- Interpretar tabelas de modo a identificar regularidades para resolver uma situação-problema.

Fonte: Secretaria de Estado da Educação de São Paulo (2000). **Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – SARESP 98: Conhecendo os resultados da avaliação, volume I .** São Paulo: SEE, p. 100-104.

8. Interpretação dos resultados

Além da interpretação de cada ponto que caracteriza a escala de habilidades, também foi calculada a porcentagem de alunos em cada série que dominavam os assuntos descritos em cada nível, visando avaliar os ganhos, em termos de conhecimentos, de um ano para outro.

Tabela 4
Porcentagem de alunos da Rede Estadual em cada nível de habilidade, segundo a série e o período

Nível	3ª série (%)	4ª série (%)	5ª série Diurno (%)	5ª série Noturno (%)
25	93	99	99	99
40	72	90	91	90
55	37	63	63	63
70	10	28	25	27
85	1	6	4	6
100	0	1	0	1

Por exemplo, para o nível 55, descrito anteriormente, podemos tirar as seguintes conclusões, em relação aos anos de 1996 e 1997 :

Em 1996, a porcentagem de estudantes que respondiam questões desse nível era de 37%. Em 1997, essa porcentagem passa a ser de 63%. Ou seja, houve um ganho de 26% (pontos percentuais) da 3ª para a 4ª série.

Por fim, foi estimada a habilidade média (e respectivo erro padrão) em Matemática, para cada escola. Assim, cada uma delas recebeu um boletim, indicando o desempenho médio da escola, da delegacia de ensino da qual ela faz parte e, também, o resultado médio

geral (ou seja, da população toda, que, no caso, são todas as escolas públicas estaduais de São Paulo). Com base nessas informações, cada instituição de ensino pode verificar qual sua situação em relação às demais, além de avaliar os ganhos de seus alunos de um ano para outro, e de ter indicações sobre quais os assuntos em que seus alunos ainda estão deficientes.

Obviamente, todos os resultados obtidos são também enviados para as Delegacias de Ensino e para a Secretaria de Estado da Educação de São Paulo. Assim, a partir das informações fornecidas pelo SARESP, as ações podem ser tomadas tanto a nível de cada instituição de ensino, quanto em proporções estaduais.

Concluindo, o SARESP, além de avaliar o desempenho da rede estadual de São Paulo ano a ano, também vem fornecendo indicadores quantitativos e qualitativos de como as intervenções no ensino público têm afetado o conhecimento dos alunos de uma série para outra, e esse tipo de questão só pode ser respondida através das ferramentas fornecidas pela TRI, dentre as quais o uso das escalas de conhecimento tem sido de grande importância.

Referências Bibliográficas

BAKER, F. B. *Item Response Theory - Parameter Estimation Techniques*. New York: Marcel Dekker, Inc., 1992.

BEATON, A. E.; ALLEN, N. L. Interpreting scales through scale anchoring. *Journal of Educational Statistics*, 17, 191-204, 1992.

BOCK, R. D.; ZIMOWSKI, M. F. Multiple Group IRT. In: W. J. van der LINDEN; R. K. HAMBLETON (eds.) *Handbook of Modern Item Response Theory*. New York: Spring-Verlag. 1997.

HAMBLETON, R. K.; COOK, L. L. Latent trait models and their use in the analysis of educational test data. *Journal of Educational Measurement*, (14) p. 75-96. 1997.

HAMBLETON, R. K.; SWAMINATHAN, H.; ROGERS, H. J. *Fundamentals of Item Response Theory*. Newbury Park: Sage Publications. 1991.

HEDGES, L. V.; VEVEA, J. L. *A study of equating in NAEP*. Paper presented at The NAEP Validity Studies Panel. Palo Alto: American Institutes for Research. 1997.

KOLEN, M. J.; Brennan, R. L. *Test Equating - Methods and Practices*. New York: Springer. 1995.

LINDEN, W. J. van der; HAMBLETON, R. K. *Handbook of Modern Item Response Theory*. New York: Springer-Verlag. 1997.

MISLEVY, R. J. *Linking Educational Assessments: concepts, issues, methods and prospects*. Princeton: Educational Testing Service. 1992.

SECRETARIA DE ESTADO DA EDUCAÇÃO DE SÃO PAULO/SEESP. *Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – SARESP 96: relatório final dos resultados*. 3v. São Paulo: SEE. 1996.

SECRETARIA DE ESTADO DA EDUCAÇÃO DE SÃO PAULO/SEESP. *Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – SARESP 97: relatório final dos resultados*, 4v. São Paulo: SEE. 1997.

SECRETARIA DE ESTADO DA EDUCAÇÃO DE SÃO PAULO/SEESP. *Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – SARESP 98: relatório final dos resultados*, 6v. São Paulo: SEE. 2000.

SECRETARIA DE ESTADO DA EDUCAÇÃO DE SÃO PAULO/SEESP. *Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo - SARESP: relatório final dos resultados 96/97/98*. São Paulo: SEE. 1999.

VALLE, Raquel da C. Teoria da resposta ao item. *Estudos em Avaliação Educacional*. (21), p.7-91. São Paulo: Fundação Carlos Chagas. 2000.

ZIMOWSKI, M. F.; MURAKI, E.; MISLEVY, R. J.; BOCK, R. D. *BILOG-MG: Multiple-Group IRT Analysis and Test Maintenance for Binary Items*. Chicago: Scientific Software, Inc., 1996.