

INTERPRETAÇÃO DE ESCALAS DE PROFICIÊNCIA COM UTILIZAÇÃO DO MÉTODO DO MARCADOR¹

FABIANA ESMÉRIA DE CASTRO ALVES UBRIACO

RESUMO

Este artigo descreve o Método do Marcador, utilizado para interpretação de escalas de proficiência conforme a criação de Níveis de Desempenho. Mostra a utilidade prática desse método e produz uma interpretação pedagógica para a escala de matemática dos alunos avaliados pelo Projeto Geres/2005 – Estudo Longitudinal da Geração Escolar 2005. A primeira etapa do método consiste na alocação de cada item utilizado no teste em um ponto da escala de medida usada para a atribuição de notas aos alunos. Quando todos os itens estão ordenados por sua posição na escala, através de uma análise substantiva, decidem-se quais itens serão tomados como definidores dos pontos de corte dos níveis de desempenho. Além disso, como são usados níveis referenciados com termos de sentido pedagógico, e não apenas valores numéricos, o método viabiliza a produção de uma interpretação potencialmente mais clara da escala de medida, mais acessível à comunidade escolar e aos pais dos alunos.

¹ Este artigo é fruto da minha dissertação de mestrado, sob orientação do Prof. José Francisco Soares.

PALAVRAS-CHAVE ESCALA DE AVALIAÇÃO • RENDIMENTO ESCOLAR • TEORIA DE RESPOSTA AO ITEM.

RESUMEN

Este artículo describe el Método del Marcador, utilizado para la interpretación de escalas de rendimiento a través de la creación de niveles de logros. Se muestra la utilidad práctica de este método, produciendo una interpretación pedagógica para la escala de matemática de los alumnos que fueron evaluados por el Proyecto Geres/2005 – Estudio Longitudinal de la Generación Escolar 2005. La primera etapa del método introducido, consiste en la asignación de cada ítem utilizado en la prueba a un punto de la escala de medida utilizada, para la distribución de notas a los alumnos. Estando todos los ítems ordenados por su posición en la escala, a través de un análisis substantiva, se deciden cuales ítems serán tomados como definidores de los puntos de corte de los niveles de logros. Como utiliza niveles referenciados con términos de sentido pedagógico y no apenas valores numéricos, este procedimiento permite la producción de una interpretación potencialmente más clara de la escala de medida y más accesible a la comunidad escolar y a los padres de los alumnos.

PALABRAS CLAVE ESCALA DE EVALUACIÓN • RENDIMIENTO ESCOLAR • TEORÍA DE LA RESPUESTA AL ÍTEM.

ABSTRACT

This article describes the Bookmark Method utilized in the interpretation of proficiency scales through the creation of Performance Levels. The present paper shows the utility of this method, producing a pedagogic interpretation of the Mathematics scale of the students assessed by the longitudinal study, Projeto Geres/2005 – Estudo Longitudinal da Geração Escolar 2005. The first stage of the method consists of the allocation of each item used in the test to a specific point in the scale used for the attribution of students' scores. After all of the items are ordered according to their position in the scale, certain items are identified, through a substantive analysis, which define the cut scores between the performance levels. This procedure uses referenced levels with terms that have pedagogic meaning and not just numeric values, it makes viable the production of an interpretation which is potentially clearer of the scale of measurement, and thereby more accessible to the school community and to the parents of students.

KEYWORDS ASSESSMENT OF PROFICIENCY SCALES • PERFORMANCE LEVELS • ITEM RESPONSE THEORY.

INTRODUÇÃO

Nas avaliações externas, normalmente, os resultados dos alunos avaliados são apresentados em escalas de proficiência². Essas escalas representam um *continuum* ordenado crescente dos desempenhos dos alunos em uma avaliação e sintetizam em um único número a posição do aluno em relação a seu domínio em uma Competência³ e, também, em relação a seus colegas.

Nas escalas de proficiência, também é possível alocar os itens utilizados na avaliação. A Probabilidade de Resposta Correta (PRC) é o que possibilita essa alocação. Como a relação entre o item e a escala é probabilística, ou seja, qualquer item pode ser acertado ou errado por aluno de qualquer proficiência, definiu-se em 65% na literatura o valor percentual de acerto para o ponto da escala para onde deve ser alocado cada item. Se um item é fácil, o valor da PRC atinge 65% logo nos pontos de menor valor da escala; por outro lado, se o item é difícil, ele demanda maior proficiência para sua resolução e, portanto, fica alocado nos pontos finais da escala.

Entender o que significa na escala a posição dos resultados dos alunos, ou a dos itens, exige, muitas vezes, um grande

2 A Escala de Proficiência também pode ser denominada Escala de Conhecimentos, Escala de Desempenho ou Escala de Habilidades.

3 Neste texto considera-se Competência como pertencente a uma categoria conceitual ampla e como aquilo que permanece latente até o momento em que surge uma situação que demande o seu uso. Como consequência, uma Competência só está completamente definida quando as tarefas, para cuja consecução ela é necessária, são explicitadas. Dessa forma, um teste é o instrumento que permite tornar visível, corporificar e medir a Competência: um conjunto de tarefas, escolhidas entre aquelas que a definem.

esforço da comunidade escolar. Por conseguinte, para facilitar e garantir uma boa comunicação das informações contidas nas escalas de proficiência, faz-se necessário realizar uma interpretação pedagógica, ou seja, evidenciar, além dos dados numéricos, o que cada aluno sabe e pode fazer.

A pretensão com essa interpretação é a de proporcionar dados para que as equipes pedagógicas das escolas possam analisar e refletir sobre os resultados de seus alunos, bem como planejar intervenções pedagógicas orientadas para a melhoria da qualidade do ensino e da aprendizagem.

O presente artigo tem a intenção de contribuir e introduzir uma alternativa ainda pouco utilizada nas avaliações brasileiras para a interpretação de escalas de proficiência, o Estabelecimento de Níveis de Desempenho através da aplicação do Método do Marcador⁴. Além da apresentação desse método, o seu uso será demonstrado empiricamente, com a realização da interpretação da escala de proficiência de matemática do Projeto Geres⁵ – Estudo Longitudinal da Geração Escolar 2005.

4 Método do Marcador é uma tradução nossa para o termo inglês *The Bookmark Method*. Uma referência detalhada da aplicação desse método é encontrada em Cizek e Bunch (2007).

ESTABELECIMENTO DE NÍVEIS DE DESEMPENHO

Estabelecer Níveis de Desempenho consiste em dividir a escala de proficiência em partes, por meio da escolha de pontos de corte.

Segundo Cizek e Bunch (2007, p. 5), determinar Níveis de Desempenho é “o processo de estabelecer um ou mais pontos de corte nas escalas das avaliações. O ponto de corte é a marca que separa um nível de outro, dividindo a distribuição do desempenho dos alunos nos testes em duas ou mais categorias” (tradução nossa)⁶.

As variações do número de pontos de corte usados são vinculadas ao objetivo da avaliação, que pode ser a certificação ou a classificação do desempenho.

Quando o intuito é certificar, apenas um ponto de corte deve ser usado, já que o essencial é criar dois padrões de desempenho que separem os não aprovados dos aprovados.

No entanto, quando se pretende com a avaliação apontar a classificação do desempenho, mais de um ponto de corte é necessário a fim de se delimitarem níveis. Para isso, a escala de proficiência é seccionada em intervalos, os níveis, os quais,

5 O Projeto Geres foi uma pesquisa longitudinal, desenvolvida de 2005 a 2008, em Belo Horizonte (MG), Rio de Janeiro (RJ), Salvador (BA), Campo Grande (MS) e Campinas (SP), cujos participantes eram alunos do primeiro segmento do Ensino Fundamental de escolas públicas e privadas. Esses alunos tiveram sua aprendizagem acompanhada do 2º ao 5º ano (na época, 1ª à 4ª série). Os dados cognitivos foram obtidos durante os quatro anos de ocorrência do projeto, através das respostas que os alunos deram às questões de múltipla escolha que compunham os testes de matemática e leitura. A medição da aprendizagem cognitiva dos alunos se deu em cinco momentos específicos, denominados ondas de aplicação dos testes.

6 Texto original em inglês. (2007).

7 Incorporar significado normativo à escala é uma maneira de aumentar sua compreensibilidade através de conceitos ou do uso de etiquetas para certos níveis ou categorias.

posteriormente, são nomeados considerando a qualidade do desempenho a que se referem. O processo de intitular os níveis incorpora informações normativas⁷ à escala, já que são usados termos com sentido explícito, como nível básico, intermediário e avançado.

Os itens usados na avaliação são alocados nos níveis, caracterizados com a descrição dos aspectos pedagógicos envolvidos em cada item. Dessa maneira, assume-se a possibilidade de comunicar com clareza o que cada aluno sabe e pode fazer, pois ao se relacionarem as especificações dos níveis com o desempenho do aluno, revela-se seu conhecimento.

Outro aspecto relevante, ao se usarem Níveis de Desempenho, é o caráter de cumulatividade, ou seja, o aluno detém tanto o conhecimento dos níveis inferiores, como aquele referente ao nível em que está alocado.

8 Texto original em inglês.

Cabe enfatizar a distinção existente, colocada por Kane (1994) (citado por CIZEK, BUNCH, 2007, p. 15, tradução nossa⁸), entre ponto de corte e padrão de desempenho. O primeiro secciona a escala em partes ou categorias. Já o segundo pode ser definido de acordo com a finalidade da avaliação, seja para o propósito de certificação, conforme o nível minimamente adequado de desempenho, seja com a intenção de classificação em categorias descritivas da Competência. O ponto de corte pode ser tratado como uma versão operacional, ao passo que o padrão de desempenho é uma versão conceitual de um desejado nível de Competência.

O MÉTODO DO MARCADOR

9 Cizek e Bunch (2007) apresentam vários métodos para o estabelecimento de Níveis de Desempenho.

O Método do Marcador é um dos métodos⁹ cuja utilização possibilita estabelecer os pontos de corte e, por conseguinte, os Níveis de Desempenho.

O emprego desse método específico consiste na execução de atividades para eleger itens cujos pontos de alocação definem os pontos de corte entre os Níveis de Desempenho. Em outras palavras, dentre todos os itens utilizados na avaliação, ocorre a distinção daqueles que recebem uma marca representando a separação de níveis, o que atribui ao método a denominação Método do Marcador.

Para que o Método do Marcador seja aplicado, a primeira providência a ser tomada é a formação de uma equipe capaz de desenvolvê-lo, constituída tanto por especialistas das Competências avaliadas, como por especialistas da área de estatística. Cabe à parte pedagógica da equipe a determinação do número de Níveis de Desempenho, a seleção de seus nomes, a definição dos pontos de corte e a descrição pedagógica dos níveis. É responsabilidade dos estatísticos determinar, com a aplicação da TRI (Teoria da Resposta ao Item), o ponto de alocação dos itens na escala.

É importante, no momento inicial, criar oportunidade para que todos da equipe conheçam o Método do Marcador, ou revejam as etapas desse procedimento. Também é necessário que todos conheçam os objetivos da avaliação e as características do teste.

Quanto à quantidade de níveis, observa-se que, na prática, não são usados muitos, prevalecendo uma organização em torno de três a cinco níveis¹⁰. Quatro parece ser um número ideal, porque assim se trabalha um nível negativo, que revela insuficiência de Competência em um determinado domínio, e três níveis positivos, desde aquele que reflete a Competência mínima em um domínio, até aquele que expressa um grau de Competência mais elevado.

Para se estabelecer os Níveis de Desempenho, é fundamental preparar o catálogo de itens, a partir do qual será feita a análise para a indicação dos pontos de corte. A fidelidade da organização desse material é decisiva para o sucesso da implementação do método. É nessa etapa que os especialistas em estatística contribuem muito com o processo, pois a construção do catálogo depende do mapa dos itens¹¹ feito previamente, tendo sido utilizados os dados estatísticos analisados com a TRI.

O catálogo compreende todos os itens utilizados na avaliação, ordenados pelo grau de dificuldade, do mais fácil ao mais difícil. Cada item deve ocupar uma página do catálogo.

Nele, pode haver itens abertos e fechados, sendo que os fechados aparecem somente uma vez e são associados a apenas um escore. Os abertos aparecem mais vezes no catálogo, dependendo das categorias que as respostas desses itens assumem, ou seja, respostas avaliadas de forma qualitativa, das

10 Para alguns exemplos, ver Jornet e Backhoff (2008, p. 7).

11 Fazer o mapa dos itens de um teste é fazer a alocação dos elementos na escala, o que, normalmente, é definido pelo valor da Probabilidade de Resposta Correta (PRC) de cada item, igual a 65% de acerto.

mais elementares às mais complexas, sendo que há uma página no catálogo para cada resposta correspondente a cada escore.

Existe chance de ocorrer um intervalo considerável (gaps) entre o nível de dificuldade de um item e de outro que lhe sucede. Nesse caso, pode-se optar por inserir no catálogo itens provenientes do banco de itens que definem a Competência, os quais não foram utilizados nos testes. Dessa maneira, diminuem-se os intervalos de dificuldade, o que possibilita que os níveis sejam estabelecidos de forma mais eficaz e que, posteriormente, sua descrição e interpretação sejam melhores.

Mais uma decisão deve ocorrer nessa fase inicial. Ela consiste na definição de qual item do catálogo considerar como o de ponto de corte, se aquele que está na última página de um nível, ou o que ocupa a primeira página do nível seguinte.

Neste momento, com as decisões tomadas e de posse do catálogo de itens, os integrantes da equipe pedagógica fazem seus primeiros julgamentos a fim de demarcar os Níveis de Desempenho. A conclusão dessa tarefa compreende várias etapas, demandando dos participantes da equipe dinâmicas diversificadas, como, por exemplo, a alternância entre momentos de trabalho individual e coletivo.

Durante o trabalho individual, cada integrante da equipe faz seus próprios julgamentos para a demarcação dos níveis e, nos momentos de trabalho coletivo, socializa suas ideias e empreende a comparação dos seus pontos de corte com os pontos de corte dos demais especialistas participantes. Cada momento de trabalho coletivo é seguido por outro individual, ou em pequenos grupos, quando os especialistas refletem sobre suas decisões e sobre a possibilidade de mantê-las ou não.

Quanto mais retroalimentação¹² os membros da equipe tiverem, mais produtivo será o trabalho e, quanto mais cedo os momentos coletivos ocorrerem, mais impacto eles causarão no julgamento dos participantes da equipe.

Para auxiliar nos seus julgamentos, os membros da equipe também levam em conta os dados estatísticos de cada item, como o valor dos seus parâmetros e o valor da Probabilidade de Resposta Correta.

Na última etapa de retroalimentação, cada especialista deve informar em qual item considera mais apropriado colocar

¹² Na descrição do método feita por Cizek e Bunch (2007), a retroalimentação (feedback) ocorre três vezes durante todo o processo.

a marca que separa um nível do outro. Caso não seja unânime o consenso sobre um ponto de corte, a definição do item para esse fim é obtida considerando-se a média dos valores das PRC dos itens indicados pelos especialistas.

Por fim, após a estipulação das marcas, é necessário fazer a transição da marcação do catálogo de itens para a escala. Isso é feito de forma bastante simples, pois a marca na escala será colocada exatamente onde está alocado o item selecionado para ser o ponto de corte.

Estabelecidos os níveis, o próximo passo é rever os itens neles alocados, a fim de fazer o levantamento dos aspectos envolvidos em cada um, para que, ao final, um conjunto desses aspectos possa caracterizar e descrever pedagogicamente cada um dos níveis.

Como os dados estatísticos permitem conhecer a proficiência dos alunos, é possível alocá-los nos níveis cuja especificação pedagógica permite apontar o conhecimento que os estudantes detêm.

O Método do Marcador, que de certa forma é uma simplificação do Método de Angoff¹³, vem ganhando popularidade por algumas razões. A primeira é que o procedimento pode ser usado em avaliações mais complexas, que utilizam itens abertos e fechados. A segunda é que, na perspectiva da equipe pedagógica que estabelece os Níveis de Desempenho, o método apresenta uma tarefa relativamente simples, uma vez que os integrantes da equipe podem estar familiarizados com o método em um nível conceitual. A terceira é que o método tem vantagens de uma perspectiva psicométrica, porque está baseado nas análises da TRI. Embora alguns aspectos computacionais sejam matematicamente complexos, a maioria deles ocorre preliminarmente ao trabalho de estabelecer os Níveis de Desempenho e é realizada pelos estatísticos envolvidos na avaliação. Assim, a tarefa da equipe que implementará o método ficará mais centrada nas questões pedagógicas associadas à definição dos níveis e a sua descrição, e essa especificidade ajuda a reduzir não só o potencial de erros, como também o tempo requerido para o trabalho.

Uma experiência positiva com o emprego do Método do Marcador é encontrada atualmente no México, que utiliza tal método para estabelecer Níveis de Desempenho na avaliação

13 Esse método é descrito por Cizek e Bunch, (2007, p. 81-95).

do Excale, Exámenes de la Calidad y el Logro Educativos, realizada pelo Instituto Nacional para a Avaliação da Educação (Inee)¹⁵.

15 A sigla significa Instituto Nacional para la Evaluación de la Educación.

A INTERPRETAÇÃO DA ESCALA DE PROFICIÊNCIA DO PROJETO GERES SEGUNDO O MÉTODO DO MARCADOR

Primeiramente, com a utilização da TRI, os três parâmetros de cada um dos itens utilizados nas cinco ondas do Projeto Geres foram estimados conjuntamente. Isso foi viabilizado usando-se o programa BILOGMG 3.0. Em seguida, construiu-se o Mapa de Itens, alocando-se cada um dos itens ao ponto da escala correspondente à PRC de 65%. De posse dessa ordenação, construiu-se o catálogo dos itens, peça fundamental para a aplicação do Método do Marcador.

O catálogo foi composto por 134 itens, sendo que cada um deles foi considerado apenas uma vez, mesmo que fosse comum na onda de aplicação do teste¹⁶, ou entre ondas. Isso é justificado porque, no Geres, é usada uma escala única e cada item tem somente um ponto de alocação nessa escala.

16 Em cada onda foram aplicados dois tipos distintos de testes, um numa versão mais fácil e outro numa versão mais difícil. Esses dois tipos de teste continham alguns itens comuns.

Estipulou-se a utilização de quatro níveis, denominados Insuficiente, Básico, Proficiente e Avançado. Esses adjetivos sugerem uma interpretação pedagógica dos níveis que, no entanto, precisa ser concretizada por meio de descrição e da exemplificação com a linguagem dos conteúdos, habilidades, contextos e processos cognitivos demonstrada através de itens utilizados nos testes.

Os quatro níveis exigem a escolha de três itens que definam os respectivos pontos de corte. Idealmente, o Método do Marcador requer que essa escolha seja feita em etapas sucessivas do trabalho conjunto de uma equipe de especialistas. Nesse trabalho, entretanto, a implementação do método foi feito por apenas uma pessoa durante sua pesquisa de mestrado. Diante disso, o passo inicial desse processo foi identificar, através do procedimento de criação de conglomerados, disponível no programa SPSS-13, pontos de corte iniciais. Em seguida, foram identificados, no catálogo, os três itens localizados imediatamente antes e os três situados imediatamente depois de cada item indicado para o ponto de corte.

Esses sete itens formaram um grupo para análise. Para a escolha do item definidor do ponto de corte foram considerados, além da análise pedagógica, os dados psicométricos fornecidos pelos três parâmetros dos itens e as características cognitivas predominantes para cada nível. O processo para o primeiro ponto de corte é explicado, a seguir, em detalhes.

O item M010074GE foi indicado pelo programa SPSS-13 para ser o ponto de corte entre os níveis Insuficiente e Básico, dessa forma, para ser o primeiro ponto de corte da escala. Então, procedeu-se à análise desse item juntamente com os três itens que o precederam, e com os três que o sucederam no catálogo. Em outras palavras, foram considerados, na análise, os itens que estavam alocados em pontos imediatamente inferiores e imediatamente superiores na escala de proficiência. Esses sete itens constituíram o conjunto a ser analisado e são apresentados um pouco mais à frente.

A primeira reflexão empreendida teve como referência a identificação da onda em que cada um dos sete itens foi aplicado. Quatro deles foram utilizados na primeira onda e três, na quinta onda. Esse fato chama atenção porque é esperado que itens da primeira onda, e até mesmo da segunda ou da terceira onda, estejam localizados nos pontos mais baixos da escala de proficiência, mas não se espera que isso ocorra com itens da quinta onda. Essa ocorrência sugere um artefato estatístico ocasionado pelo uso de tais itens exclusivamente na quinta onda¹⁷. Em decorrência desse fato, opta-se pela seleção de um item que foi aplicado na primeira onda para assumir a posição de ponto de corte.

Assim, passou-se a considerar apenas os quatro itens da primeira onda para prosseguir com a reflexão. O passo seguinte consistiu em fazer a análise dos parâmetros de cada item.

O parâmetro *a* se refere ao grau de discriminação do item e, quanto maior ele for, haverá mais chances de distinguir bem os alunos que conseguem resolvê-lo daqueles que não são capazes de fazê-lo.

O parâmetro *b* expressa a dificuldade do item, mas nesse trabalho não foi utilizado como referência. O que foi considerado como o indicador do grau de dificuldade do item foi o valor da Probabilidade de Resposta Correta.

O parâmetro *c* demonstra o acerto ao acaso.

17 O planejamento dos testes do Geres colocou apenas um item comum da quinta onda com as ondas anteriores. Isso ocasionou uma alocação de itens da quinta onda em posições inesperadas. Por exemplo, quatro itens da quinta onda foram alocados a pontos muito baixos da escala de medida, na mesma região onde estão alocados os itens da primeira onda. Isso é consequência apenas do fato de que esses itens foram muito fáceis para os alunos da quinta onda, não indicando que tais itens sejam semelhantes aos que lhe são próximos na escala. Idealmente, esses itens deveriam ter aparecido em outras ondas, permitindo assim que o processo de alocação encontrasse posições mais razoáveis. Diante do ocorrido, optou-se por não utilizar os itens da quinta onda para assumir a posição de ponto de corte.

Dentre os quatro itens analisados, o item M010084GE demonstra não ter o maior parâmetro a , mas, no entanto, apresenta o maior valor da PRC. Isso significa que ele é considerado o item mais difícil.

Esse mesmo item apresenta o segundo maior índice de acerto ao acaso (parâmetro c) entre todos os itens analisados, o que sugere que os alunos tenham encontrado maiores dificuldades em resolvê-lo, considerando o conjunto de itens em questão.

Assim, até esse ponto da análise, o item M010084GE tem dois indicadores a seu favor, a maior PRC e o parâmetro c .

Outra consideração plausível a ser feita se refere ao contexto, pois quanto mais distante das vivências do aluno o item estiver inserido, mais difícil ele se torna. Dos quatro itens, o M010084GE é o único que está inserido em uma situação de uso pública, enquanto os outros três têm contexto pessoal. Esse fato constitui mais um indício que aponta para a escolha do item M010084GE como indicador do ponto de corte.

Observando-se o processo cognitivo, constata-se que todos esses quatro itens demandam o processo de reprodução¹⁸.

Apesar de os indicadores apontarem para o item M010084GE como o melhor para assumir a função de ponto de corte, ainda restava fazer a análise pedagógica.

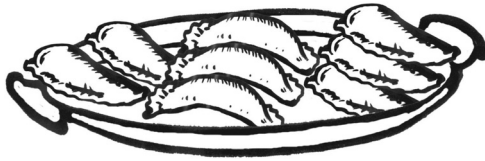
A seguir, é apresentada uma demonstração sintética da reflexão pedagógica que levou à conclusão da escolha do item M010084GE como o ponto de corte entre os níveis Insuficiente e Básico. Nos quadros abaixo se encontram os itens, apresentados em ordem crescente dos valores das suas respectivas PRC. Os itens da quinta onda também estão nessa sequência, mas, como se decidiu a princípio não usá-los para o ponto de corte, a sua análise pedagógica não foi privilegiada nesse momento.

18 O processo cognitivo de reprodução é usado na resolução de itens simples para testar a identificação de conceitos e conhecimentos e a aplicação de procedimentos rotineiros. Na maioria das vezes, a solução está apoiada em dados e fatos memorizados e na execução de ações repetidas.

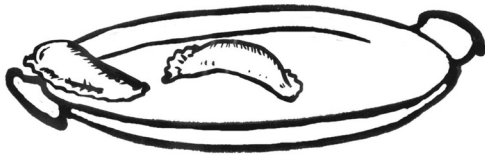
M010086GE - ONDA 1

Os amigos de Rita foram brincar na casa dela. A mãe de Rita fritou 8 pastéis para as crianças.

Veja a bandeja com esses pastéis!



As crianças comeram alguns pastéis e sobraram 2 na bandeja. Veja!



Faça um X no número que mostra a quantidade de pastéis que as crianças comeram.

3

4

5

6

a = 0,99; b = -1,59; c = 0,13;

PRC (65%) = -1,36

Este item apresenta um problema que insere uma ação subtrativa. O problema pode ser resolvido pela subtração, ou pela adição que parte do resto e, por meio do complemento, encontra-se a parte faltosa. A ilustração, usada como suporte para o item, pode induzir a resolução pela contagem.

O contexto é pessoal e o processo cognitivo é o de reprodução.

MO40012GE - ONDA 5

Uma fábrica produz 72 bichos de pelúcia e 25 bichos de borracha em um dia.

Quantos bichos essa fábrica produz em um dia?

- A) 100 bichos.
- B) 97 bichos.
- C) 99 bichos.
- D) 112 bichos.

a = 1,16; b = -1,62; c = 0,04;

PRC (65%) = -1,35

MO10014GE - ONDA 1 E ONDA 2

Ana faz aniversário no dia vinte e oito de outubro.



Faça um X na folha do calendário onde está escrito o dia do aniversário de Ana.

a = 0,95 b = -1,39 c = 0,18

PRC (65%) = -1,21

A tarefa requerida pelo item é simples, pois solicita do aluno a identificação de um número de dois algarismos.

O contexto é pessoal e o processo cognitivo é o de reprodução.

M010074GE- ONDA 1

Veja quatro crianças segurando uma placa com sua idade!



Faça um X na criança que tem a menor idade

$a = 1,02$; $b = -1,34$; $c = 0,11$;

PRC (65%) = -1,09

Para a resolução deste item, o aluno deve identificar os números e depois estabelecer uma comparação entre eles a fim de encontrar o menor. Como o campo numérico utilizado é reduzido, a dificuldade do item não é grande.

O contexto é pessoal e o processo cognitivo é o de reprodução.

M040013GE - ONDA 5

Luiza foi ao supermercado com R\$75,00 e gastou R\$40,00.
Quanto recebeu de troco?

- A) R\$25,00
- B) R\$30,00
- C) R\$20,00
- D) R\$35,00

$a = 0,88$; $b = -1,24$; $c = 0,11$;

PRC (65%) = -0,95

MO40065GE - ONDA 5

Roberto fez uma tabela de altura de jogadores de basquete.

Observe

JOGADOR	ALTURA
Anderson	1,92 m
Ratto	1,85 m
Pipoka	2,04 m
Warren	2,04 m
Caio	1,95 m

Qual é o jogador mais baixo desse time de basquete?

- A) Anderson.
- B) Ratto.
- C) Pipoka.
- D) Caio.

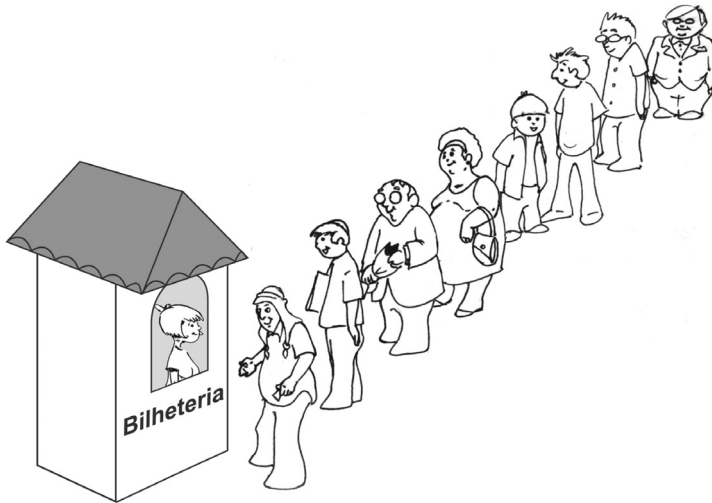
a = 0,83; b = -1,26; c = 0,04;

PRC (65%) = -0,87

M010084GE - ONDA 1 E ONDA 2

O cinema já está quase lotado.

Restam 5 ingressos para vender. Veja quantas pessoas estão na fila da bilheteria!



Faça um X no número que mostra quantas pessoas da fila ficarão do lado de fora do cinema

3

4

5

6

a = 0,79, b = -1,14, c = 0,14;

PRC (65%) = -0,85

A resolução deste item depende do entendimento da ação operatória envolvida. Para a sua resolução, é preciso que o aluno conte as pessoas que estão na fila e subtraia desse número a quantidade de ingressos que ainda restam para a venda. esse raciocínio é, de certa maneira, complexo para a criança, porque envolve uma subtração de valores de diferentes naturezas, ou seja, subtrair 5 ingressos de 8 pessoas, o que na realidade nada mais é do que a representação de ingressos por pessoas.

O contexto é público e o processo cognitivo é o de reprodução.

Diante do que foi exposto, ou seja, na reflexão pedagógica empreendida para a determinação do ponto de corte, constata-se que o item M010086GE pode ser resolvido pela contagem, num procedimento bastante elementar. O item M010014GE exige do aluno apenas a identificação numérica, ao passo que o item M010074GE requer, além da identificação de números, a comparação entre eles. Já o item M010084GE envolve um processo cognitivo mais elaborado, embora a operação de subtração implicada seja bem simples. É a demanda desse processo cognitivo que faz com que o último item citado se destaque dos demais, sugerindo uma mudança de patamar cognitivo, que aqui se configura no início de um novo Nível de Desempenho.

Para o segundo e o terceiro pontos de corte, foi realizada uma análise semelhante à descrita para o primeiro ponto.

Com os três pontos de corte determinados ao fim desse procedimento, os níveis foram delimitados, e os 134 itens, integrantes do catálogo utilizado, foram alocados a um dos quatro níveis. O quadro 1 mostra o número de itens em cada nível.

QUADRO 1 – Quantidade de itens em cada nível de desempenho

INSUFICIENTE	BÁSICO	PROFICIENTE	AVANÇADO
19 itens	32 itens	53 itens	30 itens

Após alocar todos os itens usados nos testes do Geres nos quatro níveis criados para a interpretação dos resultados, a etapa seguinte consistiu em descrevê-los.

Para isso, todos os itens foram submetidos a uma análise pedagógica, a fim de extrair os aspectos, neles envolvidos, das dimensões da Competência Matemática. Esses aspectos caracterizam os quatro níveis, e essa caracterização permite apontar o que sabe e pode fazer um aluno que tem sua proficiência alocada em um determinado nível.

CONSIDERAÇÕES FINAIS

Entende-se que o uso de Níveis de Desempenho possibilita viabilizar uma comunicação mais acessível com os vários

públicos. Revela o que um aluno localizado em certo nível consegue desempenhar, tendo como referência determinada Competência.

A partir das informações normativas, os pais podem inferir como está a aprendizagem de seu filho. O próprio aluno, diante da alocação de sua proficiência em um nível, pode se conscientizar de sua aprendizagem e se envolver mais com o seu processo de formação, desde que a descrição pedagógica do nível consiga comunicar com clareza o que ele demonstra saber e ser capaz de fazer.

Isso representa uma grande contribuição do procedimento de Estabelecer Níveis de Desempenho, mas o maior ganho que esse procedimento oferece é a informação pedagógica, que pode gerar ações de intervenção que se concretizem na melhoria do ensino brasileiro. De posse dessas informações, os educadores podem repensar e replanejar seus projetos de ensino, almejando um avanço na qualidade do ensino e da aprendizagem. Além disso, aqueles que pensam e executam as políticas públicas de educação no Brasil ficam abundantes de dados concretos que permitam a escolha de decisões acertadas e que tenham o foco ajustado na melhoria da educação do nosso país.

Para o Estabelecimento de Níveis de Desempenho, escolheu-se o Método do Marcador, dentre os vários métodos existentes na literatura, devido a sua simplicidade e por sua eficiência já comprovada. Esse método baseia-se, inicialmente, no estabelecimento de um número de níveis e na escolha de etiquetas com sentido pedagógico para cada nível. O material essencial para a implementação do método é o catálogo de itens, cuja organização é viabilizada pelo Mapa de Itens feito previamente e que consiste, basicamente, em uma reorganização dos itens, ordenados de acordo com sua alocação na escala de medida, em um catálogo no qual cada página contenha um item.

A seguir, usando o catálogo de itens, um painel de especialistas escolhe o item cuja alocação será usada como ponto de corte para cada nível. Neste trabalho, os níveis foram denominados Insuficiente, Básico, Proficiente e Avançado. Novamente, a escolha de quatro níveis reflete a influência de avaliações de outros países.

A ideia latente é que, em uma situação ideal, os alunos incorporam a Competência em três níveis. Há aqueles que, ao fim do período, dominam a competência no nível considerado adequado por aqueles que construíram o currículo. Há outros que vão além dessa expectativa e devem ser classificados no nível Avançado. No entanto, para qualquer competência cognitiva, deve-se considerar que alguns alunos têm domínio apenas no nível básico, o que lhes permite funcionar socialmente, sem, entretanto, reconhecer essa competência como essencial para sua escolha profissional ou forma de participação social. Além desses três níveis positivos, há alunos para os quais o domínio da competência é insuficiente para sua inserção social em qualquer definição, por isso seu nível é denominado Insuficiente.

Como implementação prática desse método, utilizaram-se os dados coletados, no âmbito do Projeto Geres, para a medida da Competência Matemática dos alunos das séries iniciais do Ensino Fundamental. Importante ressaltar que, nesta pesquisa, a análise do projeto Geres foi realizada apenas ao final da coleta dos dados. Assim, a análise aqui apresentada não substitui a que fora realizada de cada onda, e que pode ser encontrada nos documentos daquele projeto.

REFERÊNCIAS BIBLIOGRÁFICAS

CIZEK, Gregory J.; BUNCH, Michael B. *Standard setting: a guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage, 2007.

KANE, Michael. Validating the performance standards associated with passing scores. *Review of Educational Research*, v. 64, n. 3, p. 425-461, 1994.

JORNET, Jesús M.; BACKHOFF, E. Modelo para la determinación de Niveles de Logro y Puntos de Corte de los Exámenes de la Calidad y el Logro Educativos (Excale). *Colección Cuadernos de Investigación*, México: INEE, n. 30, mar. 2008.

RUIZ-PRIMO, M. A.; JORNET, J. M.; BACKHOFF, E. Acerca de la validez de los Exámenes de la Calidad y el Logro Educativos (Excale). *Colección Cuadernos de Investigación*, n. 20. México: INEE, 2006.

FABIANA ESMÉRIA DE CASTRO ALVES UBRIACO
Analista de pesquisa do CAEd – Centro de Políticas Públicas
e Avaliação da Educação
fabianaubriaco@gmail.com

Recebido em: DEZEMBRO 2011
Aprovado para publicação em: MAIO 2012