

CRIANÇAS PENSANDO AO RESPONDEREM QUESTÕES ABERTAS EM TESTE DE LARGA ESCALA: O QUE APRENDEMOS COM ELAS?

Maria Terezinha Jesus Gaspar – UnB

Eronдина Barbosa da Silva – UCB

Carmyra Oliveira Batista – SEEDF

A elaboração de itens de uma avaliação em larga escala é em geral cercada de aspectos técnicos que a diferenciam da elaboração de questões de uma prova para avaliação da aprendizagem. Dentre esses aspectos, gostaríamos de destacar dois: o primeiro é que um item de avaliação de larga escala deve avaliar uma e apenas uma habilidade; o segundo é que a construção das alternativas, sobretudo as incorretas, deve merecer um cuidado especial. No caso da Prova Brasil, por exemplo, nos anos iniciais, cada item deve ter quatro alternativas, a correta é chamada de gabarito e as incorretas são chamadas de distratores. E esses últimos têm como característica marcante a plausibilidade, ou seja, não podem ser respostas aleatórias, mas erros possíveis dos respondentes. Também não podem ser respostas parciais, chamadas de “meio de caminho”, que poderiam induzir o respondente ao erro.

O elaborador de um item de larga escala, normalmente, infere as possíveis respostas erradas das crianças para construir os descritores. Mas mesmo quem elabora itens há muito tempo, sempre tem dúvidas se os distratores são mesmo plausíveis. A dificuldade é, sobretudo, pensar na diversidade de procedimentos que uma criança poderia realizar e escolher dentre esses os que garantiriam mais plausibilidade.

Foi pensando nisso que a Coordenadoria de Pesquisa em Avaliação do Centro de Seleção e de Promoção de Eventos (CESPE), da Universidade de Brasília (UnB), conduziu um estudo exploratório buscando compreender as estratégias que as crianças utilizam para responderem itens de matemática. Para isso, alguns itens foram pré-testados em dois formatos: i) como item de múltipla escolha, como em geral aparecem nas avaliações em larga escala, e ii) como item de resposta aberta. A criança que respondeu o item de múltipla escolha não teve acesso ao item de resposta aberta e vice-versa.

Assim, esse ensaio tem por objetivo apresentar algumas reflexões sobre as respostas de crianças em um desses itens abertos, de matemática, que foi construído inicialmente como item de múltipla escolha para uma avaliação em larga escala. Esse item sofreu adaptações para torná-lo aberto, de tal forma que oportunizasse as crianças expressarem seus procedimentos de resolução e suas formas de pensar esses procedimentos.

Esse estudo exploratório nasceu do desejo de aplicar itens objetivos, de múltipla escolha e, ao mesmo tempo, itens de respostas abertas, de igual teor, relacionados ao tema grandezas e medidas, para que pudéssemos analisar as respostas das crianças como forma de nos aproximarmos de suas ideias e compreensões e assim, verificarmos a plausibilidade dos distratores propostos pelos elaboradores. Consideramos a experiência em questão relevante tendo em vista que as crianças dos anos iniciais estão cada vez mais cedo sendo “iniciadas” no treino de exames de larga escala e que pouco ou quase nada sabemos sobre o que pensam ao responderem itens de múltipla escolha no formato apresentados nessas avaliações.

Nesse ensaio, desenvolvemos nossas reflexões expondo uma breve discussão a respeito dos exames de larga escala no Brasil, para, em seguida, contextualizarmos o estudo exploratório apresentando um item respondido por 568 crianças, em um pré-teste realizado em Goiás, Espírito Santo, Rio de Janeiro e São Paulo.

Algumas considerações sobre as avaliações em larga escala no Brasil

O Brasil tem tentado, desde o decênio de 1990, criar e incentivar uma cultura avaliativa nos sistemas de ensino do país. Para isso, tem implementado vários exames nacionais que abrangem estudantes da educação básica à educação superior (Provinha Brasil, Saeb, Prova Brasil, Enem, Enade).

A formulação desses exames tem por base, no discurso oficial, a melhoria do ensino brasileiro e envolve um grande número de técnicos e um “arsenal sofisticado de tratamento estatístico” (VIANNA, 2003, p.51) para torná-los fidedignos. Embora saibamos da importância desses exames para a formulação de políticas públicas, já há na literatura científica brasileira uma quantidade razoável de trabalhos que chamam a atenção para a necessidade de a sociedade ler os resultados desses exames não como verdade absoluta sobre a Educação (VIANNA, 2003; ESTEBAM, 20012; FREITAS, 2012), mas como resultados mais gerais e que devem ser combinados aos resultados das avaliações para a aprendizagem na sala de aula e à avaliação institucional.

Essa combinação das três dimensões da avaliação é o que para nós pode ser considerado o texto em que as comunidades escolares, os sistemas de ensino e o Estado brasileiro poderão ler a Educação, em processo, analisando o trabalho pedagógico desenvolvido nas escolas e nas salas de aula para proporem objetivos, metas e ações que dêem sentido cidadão à escolarização em nosso país. Porém, consideramos que muitos educadores, pesquisadores e técnicos da Educação ainda não chegaram a esse nível de compreensão, o que

dirá a sociedade como um todo. Por isso, os exames de larga escala ainda causam leituras equivocadas e colocam em risco a formação do cidadão brasileiro, quando os sistemas de ensino e a sociedade, de modo geral, buscam soluções equivocadas para a Educação como, por exemplo, a meritocracia e a responsabilização de professores e das escolas.

Quanto a isso, Freitas (2009) considera que “é na tensão entre as políticas públicas centrais e as necessidades e projetos locais que se constrói a qualidade das escolas, a partir de indicadores publicizados e assumidos coletivamente, articulados no projeto pedagógico da escola”.

Mesmo considerando que um exame de larga escala não é suficiente para revelar as aprendizagens das crianças brasileiras – tendo em vista que “a aprendizagem é um processo neurológico interno invisível” (LEFRANÇOIS, 2008, p. 6), resultante da experiência¹, sabemos que os exames de larga escala tentam captar algumas habilidades dos estudantes que podem ser compreendidas como ações cognitivas que os levam a escolher alternativas que representem ou indiquem o produto de suas estratégias de resolução.

Trouxemos essa discussão maior sobre o quanto se faz ou se deixa de fazer com os resultados dos exames de larga escala para refletirmos sobre a construção de seus itens. Rodrigues (2008, p.49) considera que “deve-se cuidar, [...], para que o teste realmente meça o que propôs medir, de forma a incluir itens que cubram tão-somente o conteúdo a ser avaliado e que revele os processos usados pelo educando para fazer o teste”. A autora ainda afirma que são dois os pressupostos para a construção de um bom item - “unidimensionalidade” e “independência local”. O primeiro, admite que apenas uma habilidade será medida por item e o segundo, que as respostas (gabarito e distratores) dadas também devem evidenciar a utilização da mesma habilidade requerida.

Os exames de larga escala da educação básica brasileira, como vemos a seguir, também se utilizam desses pressupostos

[...] a partir dos itens do Saeb e da Prova Brasil, **é possível afirmar que um aluno desenvolveu uma certa habilidade**, quando ele é capaz de resolver um problema a partir da utilização/aplicação de um conceito por ele já construído. Por isso, o teste busca apresentar, prioritariamente, situações em que a resolução de problemas seja significativa para o aluno e mobilize seus recursos cognitivos (BRASIL, p. 106) [Grifos nossos].

Esses pressupostos nos leva a indagar se realmente as crianças dos anos iniciais respondem aos itens dos exames a partir das habilidades requeridas ou não? Essa questão,

¹ Experiências relativamente permanentes, mas sem serem oriundas de cansaço, maturação, drogas, lesões ou doenças, conforme explicita o autor.

embora pareça simplista, tem para nós cunho heurístico e epistemológico, pois nos coloca em estado de alerta quanto à natureza e validade dos itens produzidos e impostos às crianças estudantes brasileiras, assim como quanto às estratégias de resolução delas.

O estudo exploratório

O estudo exploratório consistiu na aplicação de 4 itens de Matemática, Português e Ciências para o 3º ano, 4º ano, 7º ano e 8º ano do ensino fundamental, em um pré-teste. No caso de Matemática e de Português, os itens foram construídos com base na matriz do SAEB, que é a mesma utilizada na Prova Brasil.

No pré-teste de Matemática do 3º ano cada aluno respondeu a 14 itens de múltipla escolha e 1 item aberto. Nesse ensaio nos propomos a discutir as respostas dadas pelas crianças em um dos itens abertos.

O item de múltipla escolha

O item em formato de múltipla escolha, apresentado às crianças, foi construído a partir do descritor D6: “Estimar a medida de grandezas utilizando unidades de medidas convencionais ou não” que compõe a matriz do SAEB para o 5º ano, conforme mostrado a seguir:

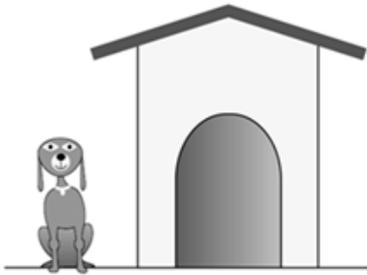
HUGO COMPROU UMA CASINHA PARA O SEU CACHORRO CONFORME MOSTRA A FIGURA ABAIXO. A ALTURA DO CACHORRO, SENTADO, É IGUAL A 50 CENTÍMETROS.	
	
CONTEXTO DO ELABORADOR	
QUAL DEVE SER A ALTURA DA CASINHA?	
A	30 CENTÍMETROS.
B	60 CENTÍMETROS.
C	100 CENTÍMETROS.
D	150 CENTÍMETROS.

Figura 1 – item de múltipla escolha

Como já dissemos, os distratores devem ser plausíveis. Em razão disso, no processo de criação de um item o elaborador deve justificar o gabarito e as respostas incorretas, inferindo os possíveis erros cometidos pelas crianças. A seguir, as justificativas construídas previamente para esse item:

A	ERRADA. Considera que a altura de casa é menos do que a do cachorro
B	ERRADA. Considera que a altura de casa é um pouco maior do que a do cachorro
C	CERTA. Percebe que a altura da casa é o dobro da do cachorro
D	ERRADA. Considera que a altura da casa é 3 vezes maior do que a do cachorro.

Figura 2 – justificativas para o gabarito e distratores do item de múltipla escolha

Após o pré-teste, foi feita a análise estatística dos resultados com base na Teoria Clássica dos Testes (TCT) que estão mostrados na tabela 1 e no gráfico, a seguir:

Análise estatística do item de múltipla escolha											
Item		Índices		Coeficientes Bisserials				Proporções de Respostas			
Descritor	Gabarito	Dificuldade	Bisserial	A	B	C	D	A	B	C	D
D6	C	0,24	0,51	-0,23	0,15	0,51	-0,24	0,26	0,31	0,24	0,11

Fonte: Coordenadoria de Pesquisa em Avaliação (CESPE/UNB)

A tabela 1 nos mostra que o item foi considerado difícil, já que apenas 24% dos alunos optaram pelo gabarito (C), encontrando a resposta esperada. Os demais se distribuíram pelas outras alternativas. Para o distrator (A) foram 26% dos respondentes. O distrator (B) foi escolhido por 31%, percentual maior do que o do gabarito e, por fim, para o distrator (D) foram 11% dos respondentes. Os dados estatísticos mostram, preliminarmente, que o distrator (B) foi mais atrativo do que o gabarito (C), o que exige uma análise pedagógica mais pormenorizada do item.

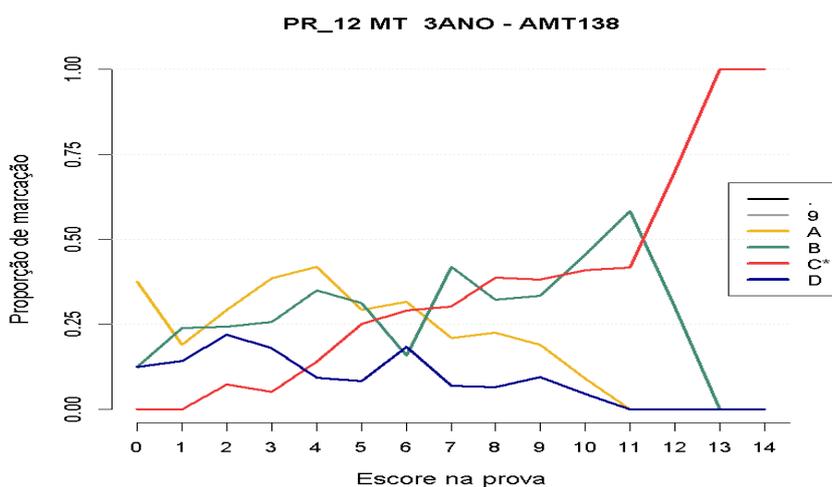
De acordo com a TCT, a dificuldade de um item é medida a partir do percentual de respondentes que escolhem a alternativa correta, como mostra a tabela 2 a seguir:

Percentual de respondentes do item de múltipla escolha	
Dificuldade	Percentual (P) de respondentes que escolheram a opção correta.
Fácil	$P \geq 60$
Médio	$40 \leq P < 60$
Difícil	$P < 40$

Fonte: Coordenadoria de Pesquisa em Avaliação (CESPE/UNB)

Apesar da dificuldade, conforme mostrado na tabela, o coeficiente bisserial da alternativa correta neste item foi de 0,51 e, portanto, maior do que 0,30. Isso significa que,

neste item, a correlação entre a proporção de alunos que foi bem no pré-teste como um todo e escolheram a alternativa correta foi maior do que a proporção de alunos que tiveram baixo desempenho no pré-teste como um todo e escolheram a alternativa correta. De acordo com a TCT um item é discriminativo se o coeficiente bisserial do gabarito é maior do que 0,30 e os coeficientes bisseriais dos distratores são negativos. No item analisado, o distrator (B) teve coeficiente bisserial positivo. Isso significa que mesmo alunos de alta proficiência no pré-teste escolheram essa alternativa errada. Outra forma de observar este fato é analisando o gráfico a seguir que relaciona o desempenho do aluno no teste como um todo e a proporção dos alunos que marcaram cada uma das alternativas do item. Por exemplo, mais de 50% dos alunos que acertaram 11 itens do teste, marcaram a alternativa B o que significa que mais da metade dos alunos que responderam corretamente a 11 dos 14 itens do teste escolheram a alternativa B ao invés do gabarito C.



Fonte: Coordenadoria de Pesquisa em Avaliação (CESPE/UNB)

É importante dizer que análise de um item não se encerra em dados estatísticos, esses apenas servem de ponto de partida para a necessária análise pedagógica. Assim, de posse desses dados é necessário questionar, por exemplo: por que o item se revelou tão difícil? Por que alunos de bom desempenho na prova como um todo foram atraídos para um distrator em particular? Os distratores são mesmo plausíveis? O item se refere mesmo ao descritor que pretende avaliar? O item é mesmo unidimensional? Que habilidade o item está avaliando de fato? Há algum problema estrutural com o item? O contexto, o comando e as alternativas estão bem redigidos e são adequados? A linguagem é clara e objetiva? A figura está adequada e é necessária para responder o item? E mais, se o item não fosse de múltipla escolha, que respostas as crianças dariam?

Foram análises pedagógicas como essa que nos levaram a considerar a possibilidade de realizar o estudo exploratório, ou seja, de apresentar itens no formato aberto, a fim de buscar aproximações e/ou distanciamentos entre as respostas dadas pelas crianças e as alternativas criadas pelos elaboradores de itens de múltipla escolha. A intenção era analisar as estratégias dos alunos para responder esse item e compará-las ao que estávamos chamando de alternativas plausíveis.

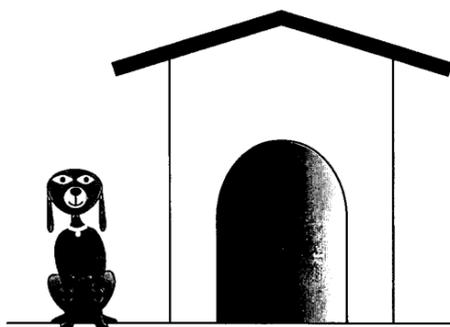
O item no formato aberto

O mesmo item foi apresentado a 568 crianças no pré-teste, mas no formato aberto, conforme mostrado a seguir:

NESTA QUESTÃO, FAÇA O QUE SE PEDE.



HUGO COMPROU UMA CASINHA PARA O SEU CACHORRO, CONFORME MOSTRA A FIGURA ABAIXO. A ALTURA DO CACHORRO, SENTADO, É IGUAL A 50 CENTÍMETROS.



QUAL DEVE SER A ALTURA DA CASINHA?

ESCREVA COMO VOCÊ PENSOU PARA RESPONDER A ESTA QUESTÃO.

Figura 3 – item aberto

Dos 568 respondentes do item aberto, 112 crianças ou 19,7% do total, deixaram a questão em branco; 2 alunos, que representa 0,3% do total, responderam “30 cm”, que corresponde ao distrator (A) do item no formato de múltipla escolha; 17 crianças, ou 3% do total, responderam “60 cm”, que corresponde ao distrator (B); 78 crianças, ou 13,7%, encontraram “100 cm”, que corresponde ao gabarito (C) e 3 crianças, ou 0,5%, responderam “150 cm”, que corresponde ao distrator (D). Mas além dessas respostas esperadas e tidas

como plausíveis pelos elaboradores, muitas outras respostas foram dadas e estão apresentadas na tabela 3, a seguir:

Outras respostas dadas ao item aberto	
Resposta (R)	Quantidade de estudantes
$R \geq 200$	3
$101 \leq R \leq 149$	12
90	13
80	19
70	19
60	17
$51 \leq R \leq 59$	22
50	32
40	2
30	1
$32 \leq R \leq 35$	2
$10 \leq R \leq 20$	26
$2 \leq R \leq 9$	18
$1,2 \leq R \leq 1,9$	2

FONTE: Coordenadoria de Pesquisa em Avaliação (CESPE/UNB)

Ao analisarmos esses dados evidenciamos que as opções erradas (80 cm, 60 cm e 70 cm) são mais plausíveis do que as opções apresentadas no item de múltipla escolha pelos elaboradores. Isto nos faz indagar: Como reformular o item e adaptar as opções plausíveis, a partir da ótica das crianças, de modo que o gabarito (100 cm) não atraia estudantes que tiverem baixo desempenho na prova? Quais seriam as justificativas plausíveis para tais opções?

As respostas abertas nos revelaram ações cognitivas muito interessantes. As crianças que responderam corretamente a questão expressaram seus pensamentos dentro de quatro categorias:

(1) *Utilizando a noção de metade*: as crianças observam a altura do cachorro e consideram que esta é a metade da altura da casinha, chegando à resposta 1 metro ou 100 centímetros, por meio de uma comparação, conforme mostram as figuras 4 e 5, a seguir. É importante ressaltar ainda que a resposta dada na figura 1 evidencia que o aluno já construiu a habilidade de estabelecer relação entre centímetro e metro. Ele demonstra saber que 100 centímetros é igual a 1 metro.

QUAL DEVE SER A ALTURA DA CASINHA?

1 METRO

ESCREVA COMO VOCÊ PENSOU PARA RESPONDER A ESTA QUESTÃO.

SEJA METADE DO CACHORRO ESTA MO
MEIADA CASINHA ENTÃO É 1 METRO

ESCREVA COMO VOCÊ PENSOU PARA RESPONDER A ESTA QUESTÃO.

Eu acho que a casinha tem que ter
100 centímetros no quadrado. O cachorrinho tem 50
centímetros. ele tem a metade do tamanho
do lar.

Figuras 4 e 5 – respostas utilizando a noção de metade

(2) *Utilizando a noção de dobro:* por meio de uma comparação, a criança dobra a altura do cachorro e estima a altura da casinha, conforme mostra a figura 6.

100

ESCREVA COMO VOCÊ PENSOU PARA RESPONDER A ESTA QUESTÃO.

SE COLOCAR UM CACHORRO NOSE
MARRÃO DO

Figura 6 – resposta utilizando a noção de dobro

(3) *Utilizando a noção de completar:* como mostra a figura 7, a criança obtém a resposta por meio da observação da altura da casinha comparada com a altura do cachorro e considera o que falta para chegar a cem.

QUAL DEVE SER A ALTURA DA CASINHA?

3cm

ESCREVA COMO VOCÊ PENSOU PARA RESPONDER A ESTA QUESTÃO.

Eu coloquei uma régua que eu tinha ali.
A casinha do cachorro é ali que era cinquenta
centímetros e mais?

Figura 7 – resposta utilizando a noção de completar

(4) *Utilizando a noção de escala:* utilizando um raciocínio bastante elaborado para uma criança de 3º ano, essa resposta, apresentada na figura 8, mostra que a criança utilizou a régua para medir a altura tanto da casa quanto do cachorro e, após achar os resultados, 3cm para a altura do cachorro e 6 cm para a altura da casa, associou a primeira à medida 50 cm e deduziu a medida solicitada.

QUAL DEVE SER A ALTURA DA CASINHA?

100 centímetros

ESCREVA COMO VOCÊ PENSOU PARA RESPONDER A ESTA QUESTÃO.

Usei a unidade de medida a "cachorro" mede 3
centímetros e a casinha 60 centímetros

Figura 8 – resposta utilizando a noção de escala

Como vemos, as crianças que acertaram a questão se utilizaram de outras habilidades além da estimativa como as noções de escala e de completar, o que nos leva a considerar que o item não avaliou apenas a habilidade prevista no descritor D6.

Outras categorias emergiram das respostas das crianças que não conseguiram acertar o item e que mereceram nossa atenção:

(1) *Confunde medida com contagem*, isto é, faz pequenos riscos, ou numera de baixo para cima na lateral da casa, à esquerda do cachorro para dar um valor à altura, como mostram a seguir as figuras 9 e 10.



ESCREVA COMO VOCÊ PENSOU PARA RESPONDER A ESTA QUESTÃO.

A CASINHA TEM 30 CENTÍMETROS
O CACHORRO TEM 15 CENTÍMETROS



ESCREVA COMO VOCÊ PENSOU PARA RESPONDER A ESTA QUESTÃO.

ASIM MES

Figuras 9 e 10 – tentativa de respostas utilizando contagem

(2) *Tenta representar a régua, em escala*, no desenho. A partir da cabeça do cachorro, que demarca 50 centímetros, marca pontos na mesma lateral da casinha contando de cinco em cinco e acha 70 centímetros, conforme a figura 11.



QUAL DEVE SER A ALTURA DA CASINHA?

ESCREVA COMO VOCÊ PENSOU PARA RESPONDER A ESTA QUESTÃO.

Eu fiz esta questão colocando as centímetros da calharia e da casinha de 50.
 Eu coloquei as centímetros da calharia que foram 50 centímetros e fui colocando até chegar nos 70 centímetros que foram os da casinha.

Figura 11 – tentativa de resposta utilizando a representação de uma régua

(3) *Usa várias unidades, convencional e não-convencional, para estimar a medida.* Como mostra a figura 12, a criança tenta achar a resposta utilizando a régua. Entra em conflito e busca uma justificativa utilizando medidas não-convencionais, mas esse raciocínio não deixa explícito como chegou a 80 centímetros.

QUAL DEVE SER A ALTURA DA CASINHA? 80 centímetros

ESCREVA COMO VOCÊ PENSOU PARA RESPONDER A ESTA QUESTÃO.

Eu pensei que eu devia medir com a régua, mas eu vi que não deu muita conta aí eu me dei com o dedo mais também não deu muita conta aí que eu me dei com a mão e acho que deu aí eu acho que deu 80 centímetros.

Figura 12 – tentativa de resposta utilizando unidades: convencional e não-convencional

(4) *Analisa a imagem desconsiderando os dados.* Conforme mostra a figura 13, a criança demonstra que o senso de medida ainda não foi explorado de maneira suficiente e que ela se apega apenas a percepção visual para formular sua resposta.

ESCREVA COMO VOCÊ PENSOU PARA RESPONDER A ESTA QUESTÃO.

ESTA CASA.. ESTA MUITO GRANDE PARA ESTE CACHORRO E E DEVE CO PRO OUTRA

Figura 13 – tentativa de resposta apoiada somente na imagem

Encontramos também outras respostas em que as crianças contam histórias relacionadas ou não ao contexto da questão. Em algumas, aparecem ainda ideias que mostram a confusão entre altura e peso, como por exemplo, “a casa é grande e pesa muito”.

Por fim, encontramos muitas respostas incompreensíveis, o que nos fez evidenciar que a alfabetização das crianças está em processo. A dificuldade de expressar suas ideias por intermédio do uso da escrita revela obstáculos que podem estar ligados à compreensão do item como texto: leitura do contexto, do comando, do gabarito e dos distratores, para responderem, sejam itens de múltipla escolha ou aberto. Isto é, consideramos que, devido a idade, as crianças do 3º ano não são trabalhadas o suficiente e o necessário para desenvolverem a leitura objetiva, a inferencial e a avaliativa (ZAMBONI, BORTONI, BORTONI-RICARDO, 2008, p. 224-226), que consideramos fundamentais para a resolução de itens dos exames de larga escala.

A leitura objetiva é a decodificação do texto; a leitura inferencial se refere à produção de sentido que o leitor dá as palavras do texto e que o leva à compreensão e a leitura avaliativa é aquela em que o leitor toma partido e se posiciona perante o texto. No caso da resolução de itens de larga escala, entendemos que essas três dimensões da leitura se inter-relacionam e são elas que possibilitam a conexão entre a experiência matemática da criança e a sua produção de estratégias para a resolução dos itens.

O que aprendemos com as respostas das crianças?

As crianças, de maneira geral, apresentaram dificuldade em identificar qual é a pergunta a ser respondida para resolver o problema assim como identificarem os dados que permitiriam responder a pergunta. Também evidenciamos problemas na construção do conceito de centímetro e metro e, por isso, levantamos a hipótese de que esse conteúdo pode estar sendo trabalhado de maneira superficial ou sem a produção de sentido para as crianças, porque elas apresentaram estratégias em que estava envolvido o uso de outros instrumentos de medida sem considerar as devidas transformações (dedo, mão, etc).

A análise de outras respostas que não a correta nos levou a constatar que as crianças estão confundindo contagem com medida. Ainda não compreendem que medir é comparar grandezas do mesmo tipo escolhendo uma delas como unidade. Fixada a unidade, a medida de uma grandeza pode ser um número natural, racional ou irracional positivo, ou seja, um número real positivo e, o conjunto dos números reais positivos não é enumerável. Logo, medir não é contar.

Verificamos também a dificuldade de as crianças apresentarem seus raciocínios em forma de um pequeno texto e esse fato nos deixa apreensivas com relação a pouca idade das crianças envolvidas com a resolução de testes de larga escala, tendo em vista que o 3º ano atende crianças de 8 anos e que mesmo as crianças em defasagem idade/série precisam estar envolvidas em atividades pedagógicas mais expressivas e livres para que desenvolvam a habilidade de expressão escrita relativa a narrativa de suas ideias e a habilidade de medir. Portanto, precisam vivenciar a avaliação formativa, aquela que compreende práticas/procedimentos/instrumentos que permitem ao professor e estudantes e demais membros da comunidade escolar acompanharem as aprendizagens em processo no cotidiano da escolarização.

Acreditamos que a dificuldade apresentada pelas crianças ao responderem este item provavelmente ocorreu devido à necessidade de aplicar o conceito de escala para responder à pergunta. Essa constatação nos faz pensar, sem querermos generalizar, que itens desse tipo podem não estar de acordo com o descritor D6: “Estimar a medida de grandezas utilizando unidades de medidas convencionais ou não”, pois envolvem outras habilidades e, portanto, não são unidimensionais, o que nos mobiliza a olhar com mais atenção os itens das avaliações de larga escala. Essa possibilidade merece ser analisada com atenção porque provavelmente muitas crianças que acertam os gabaritos de determinadas questões em itens de exames de larga escala podem igualmente estar construindo estratégias de resolução em desacordo com o descritor indicado. Além disso, nos coloca em suspensão quanto à construção dos distratores e sua plausibilidade.

Em suma e mais importante ainda, a análise exposta neste ensaio nos leva, mais uma vez, a reafirmar o cuidado e a responsabilidade dos educadores, gestores e técnicos de todos os níveis e modalidades da Educação e de outros órgãos estatais envolvidos com as avaliações de larga escala, pois, considerá-las de forma absoluta como expressão das aprendizagens das crianças sem relacioná-las com as avaliações da aprendizagem propostas pelas escolas e com a avaliação institucional, que engloba toda a organização do trabalho pedagógico escolar, constitui-se em uma limitação e um desconhecimento do que significa a formação humana.

A análise nos leva a considerar, por fim, que todo processo avaliativo, seja ele de larga escala ou não, precisa ser avaliado. Mesmo a elaboração de itens de larga escala, que segue rigorosas recomendações técnicas, necessita de permanente avaliação. E é, portanto, papel dos que estão direta ou indiretamente ligados aos exames exercer essa função crítica. E nesse sentido, a pesquisa científica é, sem dúvida nenhuma, uma grande aliada. Assim, a

apropriação do funcionamento das avaliações de larga escola, bem como das pesquisas produzidas sobre esses exames podem fazer com que gestores, técnicos, educadores, estudantes e comunidade em geral se engajem no acompanhamento e compreensão da educação pública brasileira não apenas pelo olhar externo dos exames, mas justamente pela composição crítica das informações advindas de todos os processos avaliativos.

Referências

[BRASIL]. Ministério da Educação. Plano de Desenvolvimento da Educação: **Prova Brasil**: ensino fundamental: matrizes de referência, tópicos e descritores. Brasília: MEC, SEB, Inep, 2008.

ESTEBAM, Maria Teresa. **Diferença, alteridade e aprendizagem: desafios infantis ao saber docente**. XVI ENDIPE - Encontro Nacional de Didática e Práticas de Ensino - UNICAMP - Campinas – 2012.

FREITAS, Luiz Carlos de et al. **Avaliação Educacional**: caminhando pela contramão. 2. ed. Petrópolis, RJ: Vozes, 2009.

Disponível em: <<http://www.scielo.br/pdf/es/v26n92/v26n92a10.pdf>>.

Acessado em 2 mar 2013.

_____. **Educação, avaliação qualitativa e inovação**. Texto preparado para o MEC/INEP para o Ciclo de Colóquios sobre Educação, Avaliação e Inovação. Novembro de 2012. *mimeo*.

RODRIGUES, Margarida M. M. Proposta de análise de itens das provas do Saeb sob a perspectiva pedagógica e psicométrica. **Estudos em Avaliação Educacional**, v 17, n. 34, maio/ago, 2006.

Disponível em < <http://educa.fcc.org.br/pdf/eae/v17n34/v17n34a03.pdf>>

Acessado em 20 mar 2013.

SORDI, Mara R. (2012). **Implicações ético-epistemológicas da negociação nos processos de avaliação institucional participativa**. Educação e Sociedade, v. 33, n. 119, 485-512.

VIANNA, Heraldo M. Avaliações nacionais em larga escala: análises e propostas. **Estudos em Avaliação Educacional**, n. 27, jan/jun, 2003.

ZAMBONI, Lilian Márcia; BORTONE, Maria Elizabeth; BORTONI-RICARDO, Stella Maris. **Alfabetização e Linguagem**: leitura e escrita – estratégias de apoio e prática de leitura, interpretação e produção de textos no 1º e 2º ciclos. Centro de Formação Continuada de Professores (CEFORM), Universidade de Brasília (UnB), Brasília, 2008.