

**DAS AVALIAÇÕES EM LARGA ESCALA ÀS INTERVENÇÕES DIDÁTICAS:
experiências do Observatório da Educação (OBEDUC) - polo UNEMAT**

William Vieira Gonçalves*

Acelmo Jesus de Brito**

RESUMO

Este artigo sintetiza as perspectivas teóricas e estratégicas que influenciaram o estudo e uso da Avaliação em Larga Escala na Educação Básica. Durante quatro anos de envolvimento direto com escolas públicas, entre outras ações, foram desenvolvidas práticas de ensino de matemática orientadas por índices estatísticos correlatos a desempenhos individuais de estudantes. Por questões de espaço inerentes a artigos, apresenta somente um exemplo de elaboração, validação, aplicação, correção, análise e devolutiva dos instrumentos e resultados obtidos. Pretende-se expor e refletir sobre o processo experienciado, justificando a conclusão de que as escolas parceiras ainda estão se adaptando aos descritores e suas implicações curriculares.

Palavras-chave: Ensino de matemática. Descritores. Avaliação em Larga Escala. Observatório da Educação - OBEDUC. Pesquisas Educacionais.

1 INTRODUÇÃO

Com duração de 2011 a 2014, o projeto OBEDUC reuniu integrantes de três instituições universitárias, de forma que cada uma constituiu-se como polo autônomo para decidir quais ações implementar em sua região. No polo da Universidade do Estado de Mato Grosso – *Campus* de Barra do Bugres, vinculado ao Departamento de Matemática, estabelecemos parcerias com 12 escolas públicas de Ensino Fundamental, articulando licenciandos para iniciação à docência junto a

* Mestre em Educação com ênfase a Educação em Ciências pelo Programa de Pós Graduação em Educação da Universidade Federal do Mato Grosso. Coordenador e pesquisador do Grupo de Pesquisa em Educação Matemática – GPEM. Professor concursado em Educação Matemática no Curso de Licenciatura em Matemática da Faculdade de Ciências Exatas e Tecnológicas - FACET da Universidade do Estado de Mato Grosso - UNEMAT, Campus Renê Barbours – Barra do Bugres, Brasil.

** Mestre em Recursos Hídricos pelo Programa de Pós-Graduação em Recursos Hídricos da Universidade Federal do Mato Grosso. Pesquisador do Grupo de Pesquisa em Educação Matemática – GPEM. Professor concursado em Geometria no Curso de Licenciatura em Matemática da Faculdade de Ciências Exatas e Tecnológicas – FACET da Universidade do Estado de Mato Grosso - UNEMAT, Campus Renê Barbours – Barra do Bugres, Brasil.

professores atuantes destas instituições e que aderissem ao projeto. Tais ações foram orientadas por docentes universitários e pós-graduandos. Ressaltamos que o apoio na forma de bolsas de estudo era advindo da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) para todos os integrantes do projeto.

Assumindo o compromisso de melhoria do Índice de Desenvolvimento da Educação Básica – IDEB¹ como um dos principais focos do projeto, tomamos a discussão dos seus resultados como uma possibilidade de planejamento e replanejamento de ações em salas de aulas e atividades extraclasse. No entanto, poucos dos nossos participantes tinham alguma intimidade ou conhecimento acerca das metodologias empregadas em avaliações de larga escala, exigindo a procura por materiais e experiências para aprofundamento teórico e prático.

A princípio, esperávamos considerar os resultados oficiais da Prova Brasil:

Avaliação Nacional do Rendimento Escolar - Anresc (também denominada “Prova Brasil”): trata-se de uma avaliação censitária envolvendo os alunos da 4ª série/5ºano e 8ªsérie/9ºano do Ensino Fundamental das escolas públicas das redes municipais, estaduais e federal, com o objetivo de avaliar a qualidade do ensino ministrado nas escolas públicas. Participam desta avaliação as escolas que possuem, no mínimo, 20 alunos matriculados nas séries/anos avaliados, sendo os resultados disponibilizados por escola e por ente federativo. (INEP, 2007a).

Porém, a declaração do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) indicava dificuldade metodológica, desta forma:

Diferentemente das provas que o professor aplica em sala de aula, a metodologia adotada na construção e aplicação dos testes do Saeb e Prova Brasil é adequada para avaliar redes ou sistemas de ensino, e não alunos individualmente.

Os resultados são produzidos a partir da aferição das habilidades e competências propostas nos currículos para serem desenvolvidas pelos alunos em determinada etapa da educação formal. Como os currículos são muito extensos, um aluno não responde a todas as habilidades neles previstas, em uma única prova. Um conjunto de alunos responde a várias provas. Desta forma, os resultados não refletem a porcentagem de acertos de um aluno respondendo a uma prova, mas a de um conjunto de alunos, respondendo às habilidades do currículo proposto, distribuídas em várias provas diferentes.

Como cada grupo de alunos representa uma unidade dentro do sistema de ensino, por exemplo, uma escola ou uma rede, tem-se o resultado para cada unidade prevista e não para os alunos individualmente. (INEP, 2007b).

Ou seja, mesmo os resultados delimitados por escola, apresentam problemas para elencar habilidades e competências cognitivas deficitárias em relação a cada turma. Ademais, como tais números não permitem identificar os erros e acertos individuais, tampouco possibilitam refletir o que leva a um ou outro.

¹ Recomenda-se o acesso a <http://portal.inep.gov.br/web/portal-ideb/o-que-e-o-ideb>.

A despeito de a metodologia ser amostral, a análise do INEP tem como parâmetro uma Escala de Desempenho de Matemática^{2, d} derivada de ‘Matrizes de Referência³’:

As matrizes de referência estão subdivididas em tópicos ou temas e estes, em descritores. Cada descritor é uma associação entre conteúdos curriculares e operações mentais desenvolvidas pelos alunos que traduzem certas competências e habilidades. Os descritores, portanto, especificam o que cada habilidade implica e são utilizados como base para a construção dos itens de testes das diferentes disciplinas. Cada descritor dá origem a diferentes itens e, a partir das respostas dadas a eles, verifica-se quais habilidades os alunos efetivamente desenvolveram. (INEP, 2007b).

A partir do conceito de ‘descritores’ acabamos por nos decidir pela construção e aplicação de ‘Simulados da Prova Brasil’ com metodologia adaptada aos nossos objetivos didáticos. Sem pretensões de classificar os estudantes, e sim identificar dificuldades conceituais de matemática, assumimos que tais instrumentos configurar-se-iam em um dos parâmetros articuladores das intervenções didáticas dos nossos graduandos bolsistas junto às escolas participantes do projeto.

2 PRINCÍPIOS ADOTADOS SOBRE AVALIAÇÃO EM LARGA ESCALA

A adoção de indicadores educacionais vem se tornando prática comum e são desejáveis por permitirem o monitoramento do sistema de ensino de uma escola. Sua importância, em termos de diagnóstico e norteamento, ajuda a detectar os alunos com baixo desempenho em termos de rendimento e proficiência, além de ser uma forma de monitorar a evolução temporal do desempenho dos alunos. Entretanto, há vasta literatura que indica seus perigos pedagógicos.

Sousa (2011) afirma que “[...] as avaliações em larga escala, com provas padronizadas, apesar de trazerem informações que usualmente possibilitam a comparabilidade de desempenho dos alunos, não dão conta da amplitude e complexidade do trabalho escolar.” Como exemplo, não denota o desenvolvimento de atitudes.

Ainda em Sousa, encontramos:

[...] dependendo do modo como forem divulgados e explorados os resultados das avaliações em larga escala, corre-se o risco de se intensificar desigualdades

² Disponível em:
http://download.inep.gov.br/educacao_basica/prova_brasil_saeb/escala/2011/escala_desempenho_matematica_fundamental.pdf

³ Disponível em:
http://download.inep.gov.br/educacao_basica/prova_brasil_saeb/menu_do_professor/o_que_cai_nas_provas/MA_T.zip

educacionais, com impactos na segmentação de professores e escolas. Daí deve-se atentar para não utilizar os resultados com fins classificatórios. Não se pode restringir a concepção de qualidade do ensino aos resultados dessas avaliações, correndo-se o risco de não se diferenciar as noções de medida e avaliação, além de fragilizar a perspectiva de uma avaliação formativa. (SOUSA, 2011, p. 311).

Portanto, o princípio vital para nosso trabalho de análise e devolutiva foi considerar os nossos resultados somente para confeccionar sequências didáticas que lidassem com competências prioritariamente estabelecidas com problemas, ressaltando-se também a adoção do diálogo contínuo entre os professores e os graduandos que agiam diretamente nas escolas.

No tocante à necessidade de reconhecer e compreender as várias implicações políticas e sociais destas provas, realizamos encontros periódicos para estudar em grupo o histórico das Avaliações de Larga Escala. Em trabalhos como os de Werle (2011) e Martins (2013), notam-se os impactos que dados estatísticos geraram nas políticas públicas brasileiras. Deve-se, portanto, evitar a mudança de postura educacional voltada apenas para significados pautados pelo pragmatismo e operacionalização. Em outros termos, tais dados não devem ser vistos como absolutos, mas estimular o desenvolvimento de práticas reflexivas e orientadas por objetivos específicos de dimensão cognitiva. Voltaremos a discutir as implicações e críticas que tratam da discussão sobre a qualidade da educação e sua delicada relação com os números. Por agora, apresentamos o bojo metodológico que adotamos.

3 A METODOLOGIA DESENVOLVIDA

A metodologia utilizada tanto no processo de coleta dos dados quanto nos procedimentos adotados durante o desenvolvimento e a aplicação dos simulados se alicerça nos pressupostos de uma metodologia mista. Para Creswell (2007) “[...] procedimentos de métodos mistos representam métodos pré-determinados e emergentes, questões abertas e fechadas, formas múltiplas de dados contemplando todas as possibilidades, análise estatística textual.”

Os envolvidos nas aplicações e devolutivas dos simulados foram os professores de matemática das escolas parceiras, os bolsistas, os pesquisadores e colaboradores do projeto. E, em geral, sempre participaram diretamente como fonte de dados, os alunos de faixa etária de 10 e 16 anos.

Como norte de nossas investigações e análises, utilizamos dois recursos: o primeiro está pautado na classificação através do erro, que consiste na investigação direcionada por meio das contagens dos erros de cada um dos alunos participantes; o segundo está pautado na investigação dos acertos e a implicação disso no processo de construção dos indicadores que

permitem verificar o desempenho individual com base no desempenho coletivo, indicando possibilidade de intervenções pontuais.

Os simulados foram construídos tendo como base questões disponibilizadas pelo INEP em outras edições da Prova Brasil, configuradas por questões fechadas de múltipla escolha, em que cada questão apresentava quatro alternativas com apenas uma verdadeira. Os mesmos foram aplicados em todos os 5º e 9º anos de todas as escolas parceiras.

Quanto à parte quantitativa da metodologia, esta se encontra alicerçada na estatística descritiva e na estatística inferencial.

As contribuições da estatística descritiva deram-se através dos percentuais de acerto e erro que ocorreram quando visualizamos a quantidade de questões que cada aluno acertou e/ou errou, bem como a quantidade de alunos que acertaram e/ou erraram cada questão, as médias de acerto e erro por questão e por aluno, e também os desvios dos mesmos.

Já as contribuições da estatística inferencial realizaram-se por meio dos gráficos de controle, segundo Montgomery (2009, p. 300):

[...] um gráfico de controle para o número de itens não conforme (np), se baseia nos seguintes parâmetros:

LSC – linha superior de controle

LM – linha média

LIC – linha inferior de controle

$$LSC = n.p + 3\sqrt{n.p(1-p)} \quad (1)$$

$$LM = n.p \quad (2)$$

$$LIC = n.p - 3\sqrt{n.p(1-p)} \quad (3)$$

Tais que em nosso contexto “ p ” define o percentual de acerto de questões da turma e “ n ” representa o número de questões contidas na prova.

Ainda para Montgomery (2009), existe a necessidade de verificar algumas pré-suposições estatísticas antes da construção dos gráficos de controle, declaradas a seguir:

- Os dados têm que ser independentes, que de acordo com Schwaab e Pinto (2007), sendo A e B dois eventos contido no espaço amostral U , a probabilidade e ocorrência do evento A , denominado $P(A)$, não pode alterar a probabilidade de ocorrência do evento B , denominado $P(B)$.

- Os dados têm que ter um bom ajuste à função de distribuição Normal, e para verificar esta pré-suposição, utilizamos o teste de *Kolmogorov-Smirnov* com um nível de significância igual a 5%.

De acordo com Shapiro e Francia (1972), o teste *Kolmogorov-Smirnov (K-S)* é um teste baseado na diferença máxima entre as funções de probabilidades acumuladas, empíricas

e teóricas, de variáveis contínuas. Sendo $F(x)$ a função de distribuição acumulada, com parâmetros especificados, para qual se quer verificar a aderência dos dados.

As hipóteses a serem testadas são:

H_0 : os dados provêm de $F(x)$ (há aderência)

H_1 : os dados não provêm de $F(x)$ (não há aderência)

E obtendo os valores teóricos de $F(x)$ para cada valor x_i . Se a amostra provém de uma população que segue a distribuição teórica, $F(x)$ a distribuição observada $S(x)$ não deve se afastar muito de $F(x)$.

As provas foram aplicadas em dias normais de aulas, sendo os aplicadores, o professor titular da turma e um bolsista acadêmico do projeto, que sempre seguiram as mesmas indicações da aplicação da Prova Brasil. A correção dos gabaritos e a tabulação dos resultados foram realizadas pelos bolsistas que atuaram na aplicação da prova.

Os resultados inicialmente eram apresentados ao professor titular da turma, como indicador de suas ações, bem como ao grupo do projeto atuante na referida escola como orientações das ações do grupo e, posteriormente, aos membros do projeto atuantes na linha de Avaliação em Larga Escala.

Partindo dos resultados obtidos, dava-se início ao processo de análise exploratória do rendimento, e com base no formato de tabulação usado, houve possibilidade de observar a quantidade de questões que cada aluno acertou, e também a quantidade de alunos que erraram e acertaram cada questão disponibilizada.

A seguir mostramos um exemplo do instrumento na Figura 1.

Figura 1

		E. E. Daniel Martins Moura - 9º Ano A																																						
Resposta Certa	Acertos	C	A	C	C	D	B	C	B	B	C	B	C	B	B	A	A	B	A	C	B	C	B	A	D	B	D	C	C	C	C	A	B	D	D	A	D	C		
Aluno		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37		
1																																								
2																																								
3																																								
4																																								
5																																								
6																																								
7																																								
8																																								
9																																								
10																																								
11																																								
12																																								
13																																								
14																																								
15																																								
16																																								
17																																								
18																																								
19																																								
20																																								
21																																								
22																																								
23																																								
24																																								
25																																								
26																																								
27																																								
28																																								
29																																								

Fonte: Os autores.

Antes de retomarmos a apresentação de exemplos dos resultados, torna-se necessário salientar que os resultados representam a análise de um estudo de caso específico para um determinado grupo de alunos. Portanto, esse mesmo simulado, quando aplicado em outra turma da mesma rede e até da mesma escola, poderia naturalmente apresentar um resultado bastante distinto do que foi obtido por esse estudo.

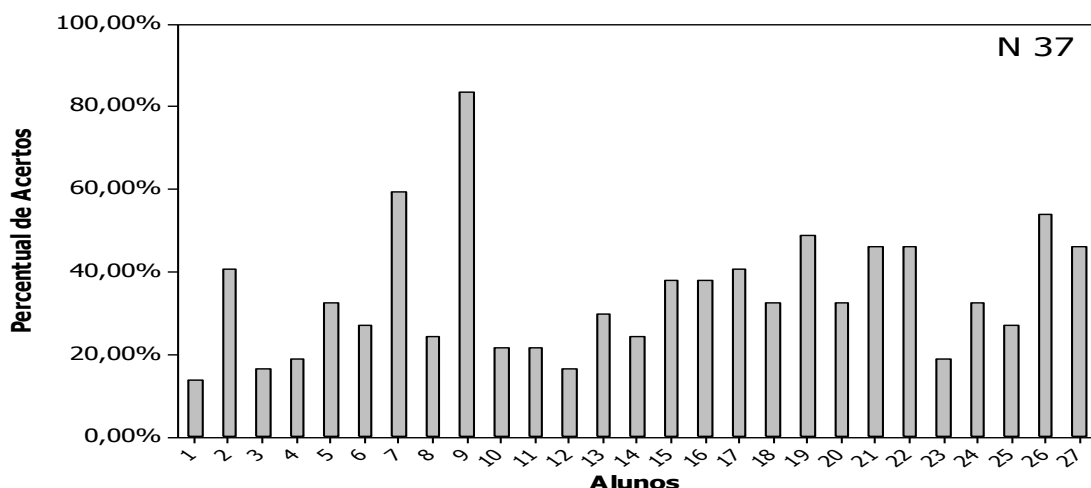
Buscamos ligar cada questão a um descritor com o intuito de permitir identificar quais pontos seriam de maior interesse ao formular atividades de intervenção posteriores.

Repetimos, nossas análises tomaram dois rumos distintos: o primeiro com caráter descritivo com base em percentuais de acerto e erro; o segundo, de caráter mais inferencial, tomou como base uma ferramenta de Controle de Qualidade.

Ainda reiteramos que aqui os alunos não serão identificados nas análises e, ao invés de seus nomes, faremos uma analogia a um número correspondente que diverge do número que os mesmos são relacionados nos diários dos professores.

Nossas primeiras análises baseavam-se no percentual de questões que cada aluno acertou. A Figura 2 mostra o gráfico que relaciona os alunos com os seus percentuais de acertos na prova.

Figura 2 – Gráfico com os percentuais de acertos dos alunos

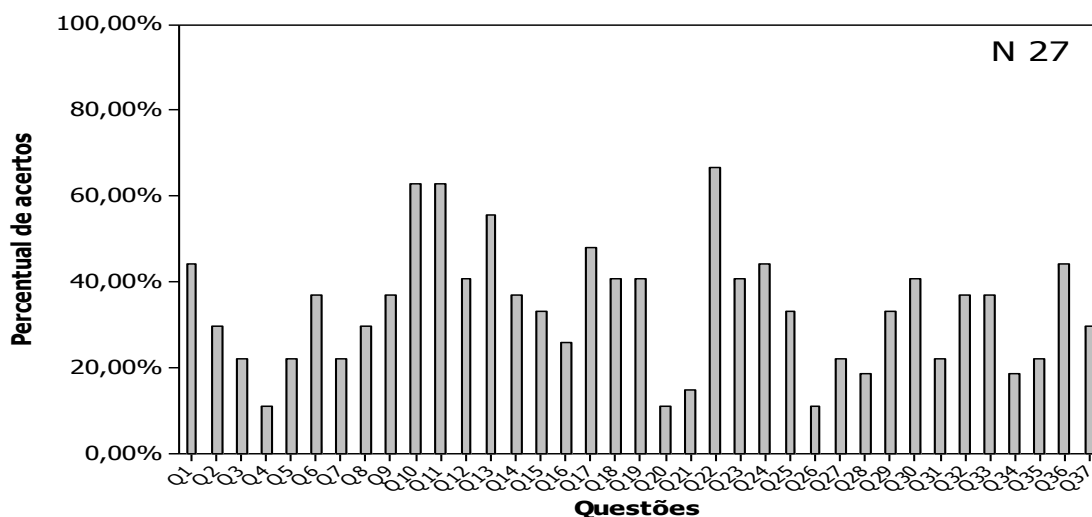


Fonte: Os autores.

Através de gráficos como o apresentado na Figura 2, buscamos identificar quais estudantes demandavam maior atenção e direcionávamos atividades específicas para lidar com os descritores em que tais alunos evidenciavam maior necessidade. Para isso, retornávamos ao instrumento de tabulação que mostramos na Figura 1.

Abaixo, na Figura 3, mostramos um gráfico que relaciona o percentual de acerto em cada uma das questões disponibilizadas.

Figura 3 - Gráfico com os percentuais de acertos das questões



Fonte: Os autores.

Através deste gráfico, buscávamos elencar atividades que poderiam ser aplicadas a todos os estudantes da turma. O principal objetivo era estratificar as prioridades de

competências e habilidades que precisávamos retrabalhar com as turmas. Para isso, revia-se e buscava-se refletir sobre as ações praticadas anteriormente com tais turmas.

Visando tecer um olhar holístico dos processos que levaram às respostas de cada item, os dados eram dispostos na Tabela 1, que se apresenta com dupla entrada para as alternativas possíveis de A até D. Salientamos as correspondências das legendas da tabela em que:

QT – quantidade de alunos que assinalaram tal alternativa como correta;

P – percentual de QT em relação à quantidade de alunos.

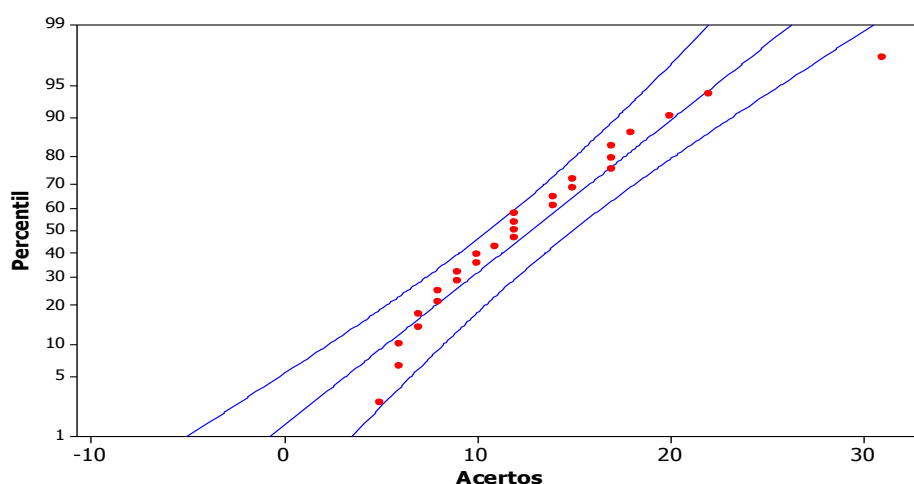
Tabela 1 – Discriminação das alternativas tidas como corretas pelos alunos.

Questões	A		B		C		D		Questão Correta
	QT	P	QT	P	QT	P	QT	P	
Q1	9	33,3%	2	7,4%	12	44,4%	3	11,1%	C
Q2	8	29,6%	14	51,9%	4	14,8%	1	3,7%	A
Q3	4	14,8%	11	40,7%	6	22,2%	6	22,2%	C
Q4	3	11,1%	7	25,9%	3	11,1%	12	44,4%	C
Q5	8	29,6%	10	37,0%	3	11,1%	6	22,2%	D
Q6	5	18,5%	10	37,0%	7	25,9%	5	18,5%	B
Q7	6	22,2%	11	40,7%	6	22,2%	4	14,8%	C
Q8	7	25,9%	8	29,6%	8	29,6%	3	11,1%	B
Q9	3	11,1%	10	37,0%	6	22,2%	8	29,6%	B
Q10	1	3,7%	5	18,5%	17	63,0%	3	11,1%	C
.
.
.
Q25	1	3,7%	9	33,3%	12	44,4%	5	18,5%	B
Q26	9	33,3%	3	11,1%	12	44,4%	3	11,1%	D
Q27	5	18,5%	11	40,7%	6	22,2%	5	18,5%	C
Q28	10	37,0%	5	18,5%	5	18,5%	6	22,2%	C
Q29	3	11,1%	10	37,0%	9	33,3%	4	14,8%	C
Q30	1	3,7%	8	29,6%	11	40,7%	7	25,9%	C
Q31	6	22,2%	9	33,3%	9	33,3%	3	11,1%	A
Q32	4	14,8%	10	37,0%	7	25,9%	6	22,2%	B
Q33	7	25,9%	5	18,5%	5	18,5%	10	37,0%	D
Q34	3	11,1%	7	25,9%	11	40,7%	5	18,5%	D
Q35	6	22,2%	3	11,1%	14	51,9%	3	11,1%	A
Q36	4	14,8%	7	25,9%	3	11,1%	12	44,4%	D
Q37	5	18,5%	7	25,9%	8	29,6%	6	22,2%	C

Fonte: os autores.

Na Figura 4 e Tabela 2, a seguir, exibimos os resultados dos testes de validação que a teoria adotada exigia.

Figura 3 - Aderência dos dados a função de distribuição normal



Fonte: os autores.

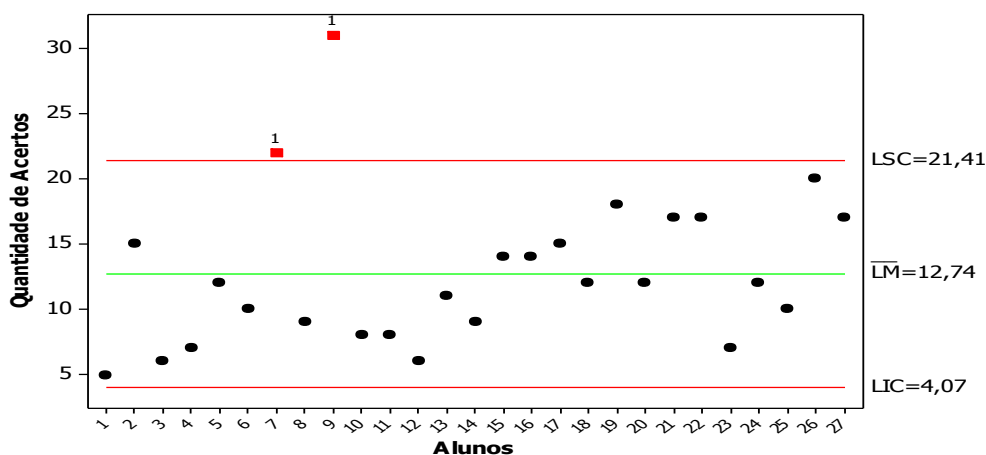
Tabela 2 – Teste de aderência *Kolmogorov - Smirnov*, a função de distribuição Normal

Variável	Média	Desvio Padrão	Nº de Observações	Estatística K-S	P-Valor
Quantidade de acerto por aluno	12,74	5,808	27	0,143	0,15

Fonte: os autores.

Finalmente, na Figura 5, podemos ver um exemplo do panorama geral do desempenho da turma, através de um gráfico de dispersão que mostra a alocação pontual dos estudantes e quantos estão nos extratos. Distto, pode-se inferir sobre o grau de atenção e o cronograma da ordem de atendimento às turmas.

Figura 4 - Gráfico de controle para o número de acertos



Fonte: os autores.

4 CONSIDERAÇÕES FINAIS

Gatti (2014) em uma das mais indicadas leituras sobre pesquisas educacionais da atualidade, em nossa opinião, chama a atenção para a necessidade de se relacionar e avaliar os impactos destes estudos em políticas públicas e vice versa. A relevante pesquisadora adverte:

É preciso considerar, no entanto, que são escassas, senão inexistentes, sínteses de conhecimento acumulado sobre temas básicos em educação que sejam acessíveis em sua linguagem e porte, a interlocutores diversos: gestores de diferentes níveis da rede educacional, mídias educacionais ou mídia em geral, associações de pais, de professores etc. Não há grupos de suporte que ofereçam contribuição dessa natureza aos organismos que tomam decisões quanto à gestão pública e pedagógica em educação (o que também ocorre em outras áreas), bem como a outros setores de interesse. Em muitos casos trabalha-se a partir de opiniões formadas no senso comum (mesmo que seja um senso comum “científico”), ou com base em modismos que adquirem forte influência; algumas vezes observa-se que se toma por base um só trabalho de pesquisa e, é claro, a leitura que dele se faz. Com esse cenário a contribuição das pesquisas na área da educação acaba empobrecida. (GATTI, 2014, p. 28).

Buscando não sermos ingênuos, não temos a pretensão de indicar um método que possa ser generalizado a situações similares a de nossa pesquisa. O principal resultado que sugerimos, seriam estes instrumentos geradores de dados, que podem ser reutilizados e certificadamente lidos pelos professores e gestores. O que acabamos por vivenciar foi resultante da diversidade de ideias e concepções de todos os atores desta história. Podemos sim, afirmar que o processo deu-se de forma a permitir posturas investigativas, laureadas pelo compromisso de professores e pesquisadores com objetivos em comum, além de reconhecer e desenvolver estratégias para lidar com as dificuldades prementes de aprendizagem.

Em relação à formação inicial de professores, pode-se dizer que nossa experiência oportunizou espaços e discussões caracterizadas por momentos extremamente ricos. Vejamos:

Pesquisas mostram que programas dirigidos especificamente para qualificar a formação inicial de professores nas instituições de ensino superior (IES), como o programa de fomento da Capes-Pibid (Programa de Iniciação à Docência) ou o Bolsa Alfabetização do Estado de São Paulo, trazem resultados positivos, estimulando estudantes a escolherem a docência e nela permanecerem, e valorizando esses cursos no interior das instituições que os acolhem e implementam. (GATTI, 2014, p. 33).

Claro, não passamos incólumes às tentativas de aplicação de atividades de intervenção nas escolas parceiras. Existiram casos em que as relações interpessoais impediam experimentações potencialmente relevantes. Destas situações apreendemos possibilidades de diálogo que permitiram aos envolvidos, desde os de iniciação à docência até os mais experientes, situar-se no complexo campo das relações profissionais. Estas, provavelmente

contribuíram para a formação de um sentido mais amplo do que significa trabalhar em equipe e com posicionamentos divergentes sobre ações educativas. Em nossa avaliação, o projeto precisa de continuidade, pois pode ser um destes programas que fortalece e diversifica as iniciativas governamentais de formação para o magistério.

A confirmação da adequação e efetividade diagnóstica dos nossos instrumentos exigem mais dados para uma análise macroscópica. No entanto, podemos confirmar sua utilidade mais direta para identificar os pontos nevrálgicos dos planejamentos iniciais. Várias das ações em parceria com as escolas pautaram-se nos relatórios estatísticos, preferencialmente nos estágios iniciais, servindo como catalisadores dos planos de ação que os graduandos levavam aos professores titulares das turmas. Tais aproximações estabeleceram confiança mútua, desencadeando espirais reflexivas sobre as atividades desenvolvidas nestas instituições de ensino.

Mesmo estas constatações anteriores podem ser obliteradas pelos aspectos políticos e pedagógicos que permeiam as avaliações em larga escala. Não são raras as advertências sobre a inversão das responsabilidades dos aparelhos estatais para com as escolas. Autores como Sousa e Arcas (2010) e Bautheney (2014) advertem sobre os fenômenos equivocados de se definir a qualidade da educação pelos desempenhos dos estudantes em provas padronizadas.

Responsabilizar integralmente os professores, gestores e ou familiares por resultados abaixo das metas, não desobriga o Estado da garantia legal de propiciar condições para melhorar a qualidade dos sistemas de ensino. Não obstante, constatam-se verdadeiras mobilizações procurando treinar alunos para estas avaliações, procurando maquiar o real estado do sistema e dificultando a melhoria do processo.

Limitando-se ao escopo que as avaliações em larga escala permitem, afirmamos que a dificuldade para ler e compreender as ‘matrizes de referências’ personifica-se nos ‘descritores’. Os professores já em ação não receberam formação apropriada para lidar com tais indicações. Aliás, mesmo a academia precisa atualizar-se neste sentido.

Os descritores não estão enquadrados por tópicos de conteúdo, comuns aos materiais e livros didáticos que são oferecidos e utilizados. Isto distancia a prática da teoria e impede a transição adequada para a nova semântica dos currículos escolares. Conclusivamente, as etapas finais desta edição do projeto OBEDUC-UNEMAT procuram delinear estudos e formações para lidar com este impasse.

OF THE LARGE-SCALE ASSESSMENTS TO THE DIDACTICAL INTERVENTIONS:

ABSTRACT

This article synthesizes the theoretical and strategic perspectives that influenced the study and use of Large-Scale Assessment in Basic Education. During four years of direct involvement with public schools, among other actions, math teaching practices were developed guided by statistical indices related to individual student performance. For reasons of space inherent to articles, presents just a example of elaboration, validation, enforcement, correction, analysis and feedback of the instruments and the results obtained. It is intended expose and reflect on the process experienced. Justifying the conclusion that the partner schools are still adapting to the descriptors and their curricular implications.

Keywords: Teaching Math. Descriptors. Large-scale Assessment. Observatory of the Education - OBEDUC. Educational Research.

REFERÊNCIAS

BAUTHENEY, K. C. S. F. Incongruências no discurso sobre qualidade da educação brasileira. **Estudos em Avaliação Educacional**, v. 25, n. 57, p. 138–162, 2014. Disponível em: <http://educa.fcc.org.br/scielo.php?script=sci_arttext&pid=S0103-68312014000100007&lng=pt&nrm=iso&tlng=pt>. Acesso em: 27 ago. 2014.

CRESWELL, J. W. **Projeto de pesquisa: métodos qualitativo, quantitativo e misto**. 2. ed. Porto Alegre: Bookman, 2007.

GATTI, B. A. Formação inicial de professores para a educação básica: pesquisas e políticas educacionais. **Estudos em Avaliação Educacional**, v. 25, n. 57, p. 24–54, 2014. Disponível em: <http://educa.fcc.org.br/scielo.php?script=sci_arttext&pid=S0103-68312014000100003&lng=pt&nrm=iso&tlng=pt>. Acesso em: 27 ago. 2014.

INEP. Aneb e Anresc (Prova Brasil). Disponível em: <<http://portal.inep.gov.br/web/saeb/aneb-e-anresc>>. Acesso em: 10 jul. 2014a.

_____; INEP. Perguntas Frequentes. Disponível em: <<http://portal.inep.gov.br/web/saeb/perguntas-frequentes>>. Acesso em: 10 ago. 2014b.

MARTINS, A. M. O campo das políticas públicas de educação: uma revisão da literatura. **Estudos em Avaliação Educacional**, v. 24, n. 56, p. 252–272, 2013. Disponível em: <http://educa.fcc.org.br/scielo.php?script=sci_arttext&pid=S0103-68312013000300012&lng=pt&nrm=iso&tlng=pt>. Acesso em: 27 ago. 2014.

MONTGOMERY, D. C. **Introduction to Statistical Quality Control**. Jeferrson City: John Wiley & Sons Inc., 2009.

SCHWAAB, M.; PINTO, J. C. **Fundamentos de Estatística e Estimação de Parâmetros**. Rio de Janeiro: Ed.E-papers, 2007.

SHAPIRO, S. S.; FRANCIA, R. S. Na Approximate Analysis of Variance Test for Normality. **Jounal of The American Statistical Association**, v. 67, p. 215–216, 1972.

SOUSA, S. Z. Avaliações em larga escala e os desafios à qualidade educacional. **Roteiro**, v. 36, n. 2, p. 309–314, 2011. Disponível em: <<http://editora.unoesc.edu.br/index.php/roteiro/article/view/1184>>. Acesso em: 27 ago. 2014.

SOUSA, S. Z.; ARCAS, P. H. Implicações da Avaliação em Larga Escala no Currículo: revelações de escolas estaduais de São Paulo. **Educação: Teoria e Prática**, v. 20, n. 35, p. 181–199, 2010. Disponível em: <<http://www.periodicos.rc.biblioteca.unesp.br/index.php/educacao/article/view/4091>>. Acesso em: 27 ago. 2014.

WERLE, F. O. C. Políticas de avaliação em larga escala na educação básica: do controle de resultados à intervenção nos processos de operacionalização do ensino. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 19, n. 73, p. 769–792, 2011. Fundação CESGRANRIO. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-40362011000500003&lng=pt&nrm=iso&tlng=pt>. Acesso em: 27 ago. 2014.