



UNIVERSIDADE DE BRASÍLIA - UNB  
INSTITUTO DE CIÊNCIAS HUMANAS ó IH  
DEPARTAMENTO DE SERVIÇO SOCIAL - SER  
PROGRAMA DE PÓS-GRADUAÇÃO EM POLÍTICA SOCIAL ó PPGPS  
TESE DE DOUTORAMENTO  
ORIENTAÇÃO: PROFª DRª SILVIA CRISTINA YANNOULAS

**AS AVALIAÇÕES EXTERNAS E SEUS EFEITOS SOBRE AS  
POLÍTICAS EDUCACIONAIS: UMA ANÁLISE COMPARADA ENTRE  
A UNIÃO E OS ESTADOS DE MINAS GERAIS E SÃO PAULO**

JOÃO LUIZ HORTA NETO

BRASÍLIA  
2013

JOÃO LUIZ HORTA NETO

**AS AVALIAÇÕES EXTERNAS E SEUS EFEITOS SOBRE AS  
POLÍTICAS EDUCACIONAIS: UMA ANÁLISE COMPARADA ENTRE  
A UNIÃO E OS ESTADOS DE MINAS GERAIS E SÃO PAULO**

*Tese apresentada ao Programa de Pós-Graduação em Política Social da Universidade de Brasília como requisito parcial para obtenção do título de doutor em política social.*

*Área de concentração: Estado, Políticas Sociais e Cidadania.*

*Orientação: Prof<sup>a</sup> Dr<sup>a</sup> Silvia Cristina Yannoulas*

BRASÍLIA  
JULHO DE 2013

Reprodução parcial permitida desde que citada a fonte

HORTA NETO, João Luiz.

As avaliações externas e seus efeitos sobre as políticas educacionais: uma análise comparada entre a União e os Estados de Minas Gerais e São Paulo / João Luiz Horta Neto. ó Brasília, 2013  
358 f.

Tese apresentada ao Programa de Pós-Graduação em Política Social da Universidade de Brasília como requisito para obtenção do título de Doutor em Política Social. Área de concentração: Estado, Políticas Sociais e Cidadania em julho de 2013. Orientadora: Profª Drª Silvia Cristina Yannoulas

1. Avaliação Educacional. Regulação. Prova Brasil. Simave. Saresp. I. As avaliações externas e seus efeitos sobre as políticas educacionais: uma análise comparada entre a União e os Estados de Minas Gerais e São Paulo.

CDU

JOÃO LUIZ HORTA NETO

AS AVALIAÇÕES EXTERNAS E SEUS EFEITOS SOBRE AS POLÍTICAS  
EDUCACIONAIS: UMA ANÁLISE COMPARADA ENTRE A UNIÃO E OS ESTADOS  
DE MINAS GERAIS E SÃO PAULO

*Tese apresentada ao Programa de Pós-Graduação em  
Política Social da Universidade de Brasília como requisito  
para obtenção do título de doutor em política social.  
Área de concentração: Estado, Políticas Sociais e  
Cidadania.*

COMISSÃO EXAMINADORA

---

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Silvia Cristina Yannoulas  
Universidade de Brasília  
Orientadora

---

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Malvina Tania Tuttman  
Universidade Federal do Rio de Janeiro  
Membro Externo

---

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Bernadete Angelina Gatti  
Universidade de São Paulo  
Membro Externo

---

Prof. Dr. Candido Alberto Gomes  
Universidade Católica de Brasília  
Membro Externo

---

Prof. Dr. Reginaldo Guiraldelli  
Universidade de Brasília  
Membro Interno Vinculado ao Programa

---

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Angela Neves  
Universidade de Brasília  
Membro Interno Vinculado ao Programa ó Suplente

A Selma, meu amor

A Isabel e Marcelo, a quem quero  
deixar um mundo mais justo

## AGRADECIMENTOS

Para a realização de um trabalho de fôlego como uma tese, é preciso o apoio e a compreensão das pessoas com quem convivemos na intimidade. Mas não foi só isso que recebi da Selma, minha mulher e o amor da minha vida. Recebi incentivo constante, chamegos nos momentos mais tensos, disponibilidade para discutir comigo, com vivacidade e inteligência que são tão suas, os achados da pesquisa conforme iam surgindo. A ela a minha admiração, meu especial carinho e amor.

Também agradeço a compreensão de minha filha Isabel por ter tido que dividir a atenção do pai com as inúmeras horas na frente do computador.

Um trabalho de pesquisa acadêmica não se constrói sozinho. Muitas pessoas deram contribuições decisivas nessa trajetória.

Primeiramente foi muito importante a confiança, dedicação, generosidade que recebi de minha orientadora Prof<sup>a</sup> Dr<sup>a</sup> Silvia Cristina Yannoulas. Graças a seu apoio nos momentos mais delicados e sua paixão pelo ato de ensinar, esse trabalho transcorreu com tranquilidade.

Ter podido desenvolver esse trabalho junto ao Programa de Pós-graduação em Política Social da Universidade de Brasília (UnB) permitiu ampliar os horizontes desse Engenheiro de formação, pela Unicamp, com mestrado em Educação pela UnB sob orientação da Prof.<sup>a</sup> Marília Fonseca, e educador por vocação. A discussão sobre as lutas pelos direitos sociais e a sua transformação em políticas foram importantes para agregar novos pontos de vista na discussão sobre o direito à educação e sua relação com a avaliação educacional. Para isso, em muito contribuíram os debates ocorridos dentro do Grupo Trabalho, Educação e Discriminação, TEDis, coordenado pela Prof<sup>a</sup> Dr<sup>a</sup> Silvia Cristina Yannoulas. Em especial os momentos de reunião para apresentar os avanços das pesquisas com as colegas de pós-graduação Anabelle Carrilho Costa, Fani Quitéria Nascimento Rehem, Kelma Jaqueline Soares e Natália de Souza Duarte e os três seminários do Projeto "Política Educacional e Pobreza", organizados pelo TEDis, em 2011 e 2012, onde foi possível comentar os achados de cada pesquisa. Outra pessoa ligada ao TEDis, a aluna de graduação Danielle Aranha Farias, auxiliou na coleta e organização de parte dos dados.

Um agradecimento especial ao Prof. Dr. Carlo Pio que generosamente me forneceu importantes sugestões durante a fase de elaboração do projeto.

As contribuições recebidas, durante a qualificação do projeto, dos professores Dr<sup>a</sup> Maria Tereza Serrano Barbosa, Dr. Cândido Alberto Gomes e Dr<sup>a</sup> Marília Fonseca, foram de enorme valia para direcionar a pesquisa, seja pelas valiosas indicações de bibliografia, seja por indicar caminhos alternativos ao trabalho de campo.

Muitas das discussões aqui apresentadas também puderam ser amadurecidas após debates, alguns bastante acalorados, que tive oportunidade de travar com meus colegas pesquisadores do Inep, sobre os rumos e sentidos da avaliação educacional e sobre o direito à educação. Em especial, contribuíram muito Ana Magna Couto, André Vitor dos Santos, Andreia Lunkes Conrado, Alexandre Jaloto, Clara Etiene, Danielle de Oliveira Costa, Estevon Naguno, Guilherme Veiga Rios, Joana Darc Ribeiro, João Galvão Bachetto, Jorge Rondelli da Costa, Lenice Medeiros, Luciana de Oliveira Xavier, Rogério Junqueira Kaiowá e Simone Poch Vieira Palma.

Foi importante o apoio recebido dos dirigentes do Inep, em especial dos diretores da Diretoria de Avaliação da Educação Básica (Daeb) Alexandre André dos Santos e Maria Tereza Serrano Barbosa. Graças a esses apoios pude ter tempo para me dedicar a esse trabalho e oportunidade para participar de eventos onde pude debater os achados da pesquisa.

Muito ajudaram, fornecendo informações e trabalhos que complementaram as pesquisas realizadas: Prof<sup>a</sup> Dr<sup>a</sup> Adriana Bauer, Prof<sup>a</sup> Dr<sup>a</sup> Sandra Zákia de Souza, Renato Carnauba Macedo, Tatiane Moraes, Edna Borges, Inez Barroso. A todos meu muito obrigado.

Também foi fundamental para a conclusão dessa pesquisa o apoio recebido da CAPES, através do Programa Observatório da Educação.

HORTA NETO, João Luiz. **As avaliações externas e seus efeitos sobre as políticas educacionais: uma análise comparada entre a União e os estados de Minas Gerais e São Paulo.** Tese (Doutorado em Política Social) Universidade de Brasília, Brasília, 2013.

## RESUMO

O objeto de estudo deste trabalho é a avaliação educacional, que usa os testes de desempenho cognitivo aplicado aos alunos como seu principal instrumento constituindo-se em uma das ferramentas estruturantes das políticas educacionais pós-burocráticas. Esses testes vêm sendo utilizados por um número cada vez maior de sistemas educacionais, no Brasil e em diversos países, em nível nacional e subnacional, como uma estratégia das reformas educacionais visando o controle da qualidade da educação. Traçou-se como objetivo analisar o desenvolvimento dos testes, aplicados pela União e pelos Estados de Minas Gerais e São Paulo aos alunos do ensino fundamental entre 1990 e 2012, buscando identificar como os resultados obtidos estão sendo utilizados pelas políticas educacionais. Para o desenvolvimento do estudo foi empregado o método comparado, utilizando-se as normas legais que instituíram cada um dos testes, identificando os movimentos de alteração ocorridos com o passar dos anos e os relatórios técnicos utilizados para divulgar seus resultados. Dada a importância que os testes vêm ganhando nas ações desenvolvidas pelos governos nacional e subnacionais, e a associação de seus resultados a um indicador de qualidade educacional, investigou-se, ainda, como essa qualidade educacional estava sendo tratada pelos meios de comunicação escritos disponíveis na Internet. Para a análise dos três casos escolhidos, definiram-se como variáveis a periodicidade de aplicação dos testes, sua abrangência em função da quantidade de alunos, dos anos escolares e áreas do conhecimento testados, a forma de se apresentar os resultados e os tipos de responsabilização, de baixa ou alta consequência, perante os resultados alcançados. Para verificar os efeitos dos resultados sobre as políticas educacionais, foram escolhidas algumas ações de cada um dos três governos. A análise dos achados da pesquisa foram iluminadas a partir da discussão da educação como um direito, dos mecanismos de regulação como elementos de condução da ação política do Estado e das modificações que o conceito de avaliação educacional sofreu até os dias atuais. Concluiu-se que os testes têm sido utilizados não como um instrumento para avaliar os sistemas educacionais e auxiliar no processo de aprendizagem dos alunos, mas sim como um instrumento de regulação do trabalho realizado pela escola e por seus profissionais, e, assim, seus resultados não têm influenciado as políticas educacionais.

Palavras-chave: Avaliação Educacional. Regulação. Prova Brasil. Ideb. Simave. Saesp.



HORTA NETO, João Luiz. **External evaluations and their effects on educational policies: a comparative analysis among the Federal government and the states of Minas Gerais and São Paulo.** Tese (Doutorado em Política Social) Universidade de Brasília, Brasília, 2013.

## ABSTRACT

This study focuses on educational evaluation, which uses cognitive performance tests applied to students as their main instrument constituting one of the tools in structuring post-bureaucratic educational policies. These tests have been used by a growing number of educational systems in Brazil and many other countries at the national and subnational levels, as an educational reform strategy aiming to improve the quality of education. The main goal is to analyze the development of the tests applied to the elementary schools by the Federal Government and the States of Minas Gerais and São Paulo between 1990 and 2012 in order to verify how and whether the results are being used by educational policies. Thus, it was used a comparative method based on the legal rules that established each test, identifying the changes that have occurred over the years, and also the technical reports used to disseminate their results. Given the importance that the tests have been gaining within relevant federal and state governments policies and also the association of its results as an indicator of educational quality, it was also investigated, how the media writings available on the Internet has treated the quality of education. Therefore, three cases were chosen to be analyzed and the following variables will be taken in consideration: the frequency in which the tests are applied to the pupils, its scope depending on the number of students, the school years and areas of knowledge tested, the way in which results are presented and the low or high stakes accountability adopted. To check the effects in terms of policy results, some government's actions were chosen from each of the three cases. The findings were discussed through the reflection that education is a right, that the regulatory mechanisms are elements of good manners for the political action of the State and that the concept of educational evaluation suffered conceptual changes throughout the years so far. To conclude, it was acknowledge that the tests have not been used as a tool to evaluate the educational systems and therefore assist the learning process of the pupils, but rather as an instrument for regulating the school's activities and its staff, and thus their results not have influenced educational policies.

Keywords: Educational Assessment. Regulation. Prova Brasil. Ideb. Simave. Saesp.

## LISTA DE QUADROS

QUADRO 1.1	ó INÍCIO DA APLICAÇÃO DOS TESTES EXTERNOS NA AMÉRICA LATINA.....	26
QUADRO 1.2	ó PAÍSES PARTICIPANTES DAS PESQUISAS CONDUZIDAS PELO IEA.....	27
QUADRO 1.3	ó PAÍSES PARTICIPANTES DO PISA ó 2000 A 2012.....	27
QUADRO 1.16	PROFICIÊNCIA MÉDIA NO PISA E CLASSIFICAÇÃO ENTRE OS 63 PAÍSES PARTICIPANTES: MÉDIA OCDE ó FINLÂNDIA, EUA, NOVA ZELÂNDIA E INGLATERRA ó 2009 .....	73
QUADRO 1.2	ó VARIAÇÃO DA PONTUAÇÃO OBTIDA PELOS ESTUDANTES NAS VERSÕES <i>LONG-TERM</i> E <i>MAIN</i> , DO NAEP, PARA DIVERSOS GRUPOS DE ESTUDANTES ó 1973 A 2011.....	79
QUADRO 1.3	ó ELEMENTOS CHAVES DO MOVIMENTO GLOBAL DE REFORMA EDUCACIONAL (GERM), EM COMPARAÇÃO COM AS POLÍTICAS EDUCACIONAIS DA FINLÂNDIA, DESDE OS ANOS 1990 .....	86
QUADRO 2.1	ó MÉDIA, DESVIO-PADRÃO, FIDEDIGNIDADE E DISCRIMINAÇÃO MÉDIA DAS PROVAS DE PORTUGUÊS, MATEMÁTICA E CIÊNCIAS, APLICADAS A UMA AMOSTRA DE ALUNOS DE 1ª, 3ª, 5ª E 7ª SÉRIES DO 1º GRAU DA REDE OFICIAL DE ENSINO, EM 15 UNIDADES DA FEDERAÇÃO ó 1988 .....	111
QUADRO 2.2	ó MÉDIA DAS PROVAS DE PORTUGUÊS, MATEMÁTICA E CIÊNCIAS, APLICADAS A UMA AMOSTRA DE ALUNOS DE 1ª, 3ª, 5ª E 7ª SÉRIES DO 1º GRAU DA REDE OFICIAL DE ENSINO, EM 25 UNIDADES DA FEDERAÇÃO ó SAEB 1990 .....	115
QUADRO 2.3	ó MÉDIA DAS PROVAS DE PORTUGUÊS, MATEMÁTICA E CIÊNCIAS, APLICADAS A UMA AMOSTRA DE ALUNOS DE 1ª, 3ª, 5ª E 7ª SÉRIES DO 1º GRAU DA REDE OFICIAL DE ENSINO, EM 27 UNIDADES DA FEDERAÇÃO ó SAEB 1993 .....	118
QUADRO 2.4	ó MATRIZ DE ESPECIFICAÇÃO DE ITENS DE LÍNGUA PORTUGUESA E SUA QUANTIDADE PARA AS 4ª E 8ª SÉRIES DO 1º GRAU E PARA A 3ª SÉRIE DO 2º GRAU ó CICLO 1995.....	125
QUADRO 2.5	ó MATRIZ DE ESPECIFICAÇÃO DE ITENS DE MATEMÁTICA E SUA QUANTIDADE PARA AS 4ª E 8ª SÉRIES DO 1º GRAU E PARA A 3ª SÉRIE DO 2º GRAU ó CICLO 1995....	126
QUADRO 2.6	ó MATRIZ DE REFERÊNCIA DE ITENS DE LÍNGUA PORTUGUESA E SUA QUANTIDADE PARA A 4ª SÉRIE DO ENSINO FUNDAMENTAL, EM 1997.....	131
QUADRO 2.7	ó MATRIZ DE REFERÊNCIA DE ITENS DE LÍNGUA PORTUGUESA E SUA QUANTIDADE PARA A 8ª SÉRIE DO ENSINO FUNDAMENTAL, EM 1997.....	131
QUADRO 2.8	ó MATRIZ DE REFERÊNCIA DE ITENS DE LÍNGUA PORTUGUESA E SUA QUANTIDADE PARA A 3ª SÉRIE DO ENSINO MÉDIO, EM 1997 .....	132
QUADRO 2.9	ó MATRIZ DE REFERÊNCIA PARA O SAEB DE MATEMÁTICA E SUA QUANTIDADE PARA AS 4ª E 8ª SÉRIES DO ENSINO FUNDAMENTAL E PARA A 3ª SÉRIE DO ENSINO MÉDIO, EM 1997.....	133
QUADRO 2.10	ó RELAÇÃO ENTRE NÍVEL DE PROFICIÊNCIA NA ESCALA DO SAEB/97 E ETAPA DE ENSINO .....	134
QUADRO 2.11	ó DESCRIÇÃO DOS NÍVEIS DE PROFICIÊNCIA 175 E 250 ó CICLO DE 1997.....	135
QUADRO 2.12	ó COMPARAÇÃO DAS MÉDIAS DE PROFICIÊNCIA EM MATEMÁTICA DAS UNIDADES DA FEDERAÇÃO E DAS REGIÕES COM A MÉDIA NACIONAL, SEGUNDO A SÉRIE ó CICLO DE 1997 .....	138
QUADRO 2.13	ó PERCENTIS DE DISTRIBUIÇÃO, DISPERSÃO E ERRO-PADRÃO DA PROFICIÊNCIA EM MATEMÁTICA PARA A 4ª SÉRIE DO ENSINO FUNDAMENTAL ó BRASIL, REGIÕES E ESTADOS ó CICLO DE 1997 .....	139

QUADRO 2.14	ó POSIÇÃO DAS UNIDADES DA FEDERAÇÃO EM RELAÇÃO À VARIAÇÃO DA MÉDIA DE PROFICIÊNCIA EM MATEMÁTICA ó 1995-1997 .....	141
QUADRO 2.15	ó MATRIZ CURRICULAR DE REFERÊNCIA DO SAEB PARA LÍNGUA PORTUGUESA E QUANTIDADE DE DESCRITORES PARA AS 4ª E 8ª SÉRIES DO ENSINO FUNDAMENTAL E PARA A 3ª SÉRIE DO ENSINO MÉDIO ó 1999 .....	142
QUADRO 2.16	ó MATRIZ CURRICULAR DE REFERÊNCIA DO SAEB PARA MATEMÁTICA E QUANTIDADE DE DESCRITORES PARA AS 4ª E 8ª SÉRIES DO ENSINO FUNDAMENTAL E PARA A 3ª SÉRIE DO ENSINO MÉDIO, EM 1999 .....	143
QUADRO 2.17	ó NÍVEIS DE DESEMPENHO EM MATEMÁTICA ó SAEB/99 .....	145
QUADRO 2.18	ó MATRIZ DE REFERÊNCIA DE LÍNGUA PORTUGUESA, QUANTIDADE DE DESCRITORES PARA AS 4ª E 8ª SÉRIES DO ENSINO FUNDAMENTAL E PARA A 3ª SÉRIE DO ENSINO MÉDIO E GRAU DE PRIORIDADE DOS DESCRITORES, DE P1 AP5, PARA SUA INCLUSÃO NO TESTE, EM 2001 .....	148
QUADRO 2.19	ó MATRIZ DE REFERÊNCIA DE MATEMÁTICA, A QUANTIDADE DE DESCRITORES PARA AS 4ª E 8ª SÉRIES DO ENSINO FUNDAMENTAL E PARA A 3ª SÉRIE DO ENSINO MÉDIO E O GRAU DE PRIORIDADE DOS DESCRITORES, DE P1 AP5, PARA SUA INCLUSÃO NO TESTE, EM 2001 .....	148
QUADRO 2.20	ó MÉDIAS DE DESEMPENHO DO SAEB 2001 EM MATEMÁTICA, AJUSTADAS PELO NÍVEL SOCIOECONÔMICO, PARA AS 4ª E 8ª SÉRIES DO ENSINO FUNDAMENTAL E PARA A 3ª SÉRIE DO ENSINO MÉDIO.....	149
QUADRO 2.21	ó COMPARAÇÃO ENTRE O NÚMERO DE ITENS, BLOCOS, CADERNOS E ITENS RESPONDIDOS E A ÁREA DE CONHECIMENTO DO TESTE RESPONDIDO POR ALUNO NOS CICLOS DE 1995 E 2011 .....	155
QUADRO 2.22	ó TABULAÇÃO DAS MATÉRIAS COLETADAS NO PERÍODO ENTRE 10 DE JANEIRO DE 2009 E 11 DE OUTUBRO DE 2012, UTILIZANDO A FERRAMENTA ãALERTAS DO GOOGLEö, CLASSIFICADAS EM TEMAS GERAIS E ESPECÍFICOS E A FREQUÊNCIA COM QUE FORAM TRATADOS.....	189
QUADRO 2.23	ó TRECHOS DE MATÉRIAS PUBLICADAS EM DIVERSAS MÍDIAS, EM QUE O TEMA CENTRAL É O IDEB, POR LOCALIDADE DA PUBLICAÇÃO .....	193
QUADRO 2.24	ó TABULAÇÃO DAS MATÉRIAS COLETADAS NO PERÍODO ENTRE 10 DE JANEIRO DE 2009 E 11 DE OUTUBRO DE 2012, UTILIZANDO A FERRAMENTA ãALERTAS DO GOOGLEö, QUE TRATA DAS REDES ESTADUAIS E DE SUAS AÇÕES PARA A MELHORIA DO IDEB, MOSTRANDO ESTADO, TAMANHO DA REDE E VARIAÇÃO NO IDEB DO ENSINO FUNDAMENTAL EM 2009 E 2011 .....	198
QUADRO 2.25	ó TABULAÇÃO DAS MATÉRIAS COLETADAS NO PERÍODO ENTRE 10 DE JANEIRO DE 2009 E 11 DE OUTUBRO DE 2012, UTILIZANDO A FERRAMENTA ãALERTAS DO GOOGLEö, QUE TRATA DAS REDES MUNICIPAIS E DAS AÇÕES PARA A MELHORIA DO IDEB, MOSTRANDO ESTADO, MUNICÍPIO, TAMANHO DA REDE E VARIAÇÃO NO IDEB, NO ENSINO FUNDAMENTAL, EM 2009 E 2011 .....	200
QUADRO 2.26	ó TABULAÇÃO DAS MATÉRIAS COLETADAS NO PERÍODO ENTRE 10 DE JANEIRO DE 2009 E 11 DE OUTUBRO DE 2012, UTILIZANDO A FERRAMENTA ãALERTAS DO GOOGLEö, QUE TRATA DAS ESCOLAS E DE SUAS AÇÕES PARA A MELHORIA DO IDEB, MOSTRANDO MUNICÍPIO, TAMANHO DA ESCOLA E EFEITO NO IDEB, NOS ANOS INICIAIS E FINAIS DO ENSINO FUNDAMENTAL.....	204
QUADRO 2.27	ó RESPOSTA À PERGUNTA ESTIMULADA, COM TRÊS OPÇÕES DE RESPOSTA, SOBRE A PERCEPÇÃO DA POPULAÇÃO, REFERENTE A QUEM PODE CONTRIBUIR MAIS PARA GARANTIR A QUALIDADE DA EDUCAÇÃO BÁSICA ó 2006-2010 .....	209
QUADRO 2.28	ó AGREGAÇÃO DAS RESPOSTAS RELACIONADAS AOS GOVERNOS E ÀS ESCOLAS, FORNECIDAS PELA PESQUISA DO IBOPE.....	210

QUADRO 3.1	ó	CARACTERIZAÇÃO DAS REDES DE ENSINO FUNDAMENTAL REGULAR ESTADUAIS, A PARTIR DA POPULAÇÃO TOTAL, DO NÚMERO DE MUNICÍPIOS, ESCOLAS, ALUNOS MATRICULADOS E PROFESSORAS EM EXERCÍCIO	ó	MG, SP E BRASIL	211
QUADRO 3.2	ó	PREVISÃO DE INSTRUMENTOS E ANOS ESCOLARES A SEREM TESTADOS DENTRO DO PROGRAMA DE AVALIAÇÃO DO SISTEMA ESTADUAL DE ENSINO DE MINAS GERAIS	.....		217
QUADRO 3.3	ó	QUANTITATIVO DOS TRÊS PRIMEIROS CICLOS DO PROGRAMA DE AVALIAÇÃO DO SISTEMA ESTADUAL DE ENSINO DE MINAS GERAIS	ó	1992 A 1995	..... 219
QUADRO 3.4	ó	ORGANIZAÇÃO DOS CADERNOS DE TESTE, A PARTIR DA DIVISÃO DOS 169 ITENS EM BLOCOS E CADERNOS	.....		226
QUADRO 3.5	ó	PADRÕES DE DESEMPENHO NO PROEB PARA OS 5º E 9º ANOS DO ENSINO FUNDAMENTAL	ó	LÍNGUA PORTUGUESA E MATEMÁTICA	..... 233
QUADRO 3.6	ó	NÍVEIS DE DESEMPENHO NO PROALFA PARA OS 2º, 3º E 4º ANO DO ENSINO FUNDAMENTAL	EM	LÍNGUA PORTUGUESA	..... 236
QUADRO 3.7	ó	ESCALA DE PROFICIÊNCIA DO CICLO DE ALFABETIZAÇÃO PARA O 3º ANO DO ENSINO FUNDAMENTAL (EM AZUL OS NÍVEIS DE COMPETÊNCIA INDICADOS POR ESPECIALISTAS E EM VERDE OS NÍVEIS DE DESEMPENHO DEFINIDOS PELA SECRETARIA)	.....		237
QUADRO 3.8	ó	SÍNTESE DO SIMAVE ENTRE 2000 E 2011: ANOS ESCOLARES, ÁREAS DE CONHECIMENTO E NÚMERO DE ALUNOS TESTADOS	.....		245
QUADRO 3.9	ó	PLACA INFORMATIVA CONTENDO O IDEB DE 2011 DE UMA ESCOLADA REDE ESTADUAL DE MINAS GERAIS	.....		251
QUADRO 3.10	ó	SÍNTESE DO SARESP ENTRE 1996 E 2011: ANOS ESCOLARES, ÁREAS DE CONHECIMENTO E NÚMERO DE ALUNOS TESTADOS	.....		261
QUADRO 3.11	ó	EXPECTATIVAS DE APRENDIZAGEM DEFINIDAS A PARTIR DOS NÍVEIS DE PROFICIÊNCIA DO SARESP	.....		266
QUADRO 3.12	ó	MATRIZ CURRICULAR BÁSICA PARA OS ENSINOS FUNDAMENTAL E MÉDIO	....		276
QUADRO 4.1	ó	COMPARATIVO ENTRE TESTES APLICADOS E TAMANHO DA REDE DE ESCOLAS DE ENSINO FUNDAMENTAL	ó	UNIÃO, MINAS GERAIS E SÃO PAULO, 1990 A 2012	..... 285
QUADRO B1	ó	ROTATIVIDADE DE PROFESSORAS DOS ANOS INICIAIS DO ENSINO FUNDAMENTAL NA REDE ESTADUAL	ó	2007 A 2012	..... 335
QUADRO B2	ó	ROTATIVIDADE DE PROFESSORAS DOS ANOS INICIAIS DO ENSINO FUNDAMENTAL NA REDE MUNICIPAL	ó	2007 A 2012	..... 336
QUADRO B3	ó	ROTATIVIDADE DE PROFESSORAS DOS ANOS INICIAIS DO ENSINO FUNDAMENTAL NA REDE PRIVADA	ó	2007 A 2012	..... 337
QUADRO B4	ó	ROTATIVIDADE DE PROFESSORAS DOS ANOS INICIAIS DO ENSINO FUNDAMENTAL NAS TRÊS REDES	ó	2007 A 2012	..... 338
QUADRO C1	ó	EXPECTATIVAS DE APRENDIZAGEM PARA A ÁREA DE LÍNGUA PORTUGUESA, PARA O 5º ANO DO ENSINO FUNDAMENTAL	.....		344
QUADRO C2	ó	EXPECTATIVAS DE APRENDIZAGEM PARA A ÁREA DE MATEMÁTICA, PARA O 5º ANO DO ENSINO FUNDAMENTAL	.....		345
QUADRO C3	ó	SÍNTESE COM O NÚMERO DE HABILIDADES PARA AS ÁREAS DE CONHECIMENTO TESTADAS PELO SARESP, POR ANO ESCOLAR	.....		347
QUADRO C4	ó	EXEMPLO DE HABILIDADES DO SARESP PARA O 5º ANO DO ENSINO FUNDAMENTAL, AGRUPADOS POR SITUAÇÃO E COMPETÊNCIAS DE ÁREA PARA LÍNGUA PORTUGUESA	.....		348
QUADRO C5	ó	EXEMPLO DE HABILIDADES DO SARESP PARA O 5º ANO DO ENSINO FUNDAMENTAL, AGRUPADOS POR TEMAS E COMPETÊNCIAS DE ÁREA PARA MATEMÁTICA	.....		349

QUADRO C6 ó EXPECTATIVAS DE APRENDIZAGEM DEFINIDAS A PARTIR DOS NÍVEIS DE PROFICIÊNCIA DO SARESP.....	350
QUADRO C8 ó DESCRIÇÃO DA ESCALA DE MATEMÁTICA PARA OS PONTOS 225 E 300.....	356

## LISTA DE GRÁFICOS

GRÁFICO 1.1	ó QUANTIDADE DE ESTADOS, POR ANO, QUE CONTAM COM TESTES AVALIATIVOS PRÓPRIOS ó 1992 A 2012 .....	23
GRÁFICO 2.1	ó DESEMPENHO NAS 1ª, 3ª, 5ª E 7ª SÉRIES DO 1º GRAU EM MATEMÁTICA: CICLO DE 1990 CALIBRADO E 1993 .....	119
GRÁFICO 2.2	ó DESEMPENHO NAS 1ª, 3ª, 5ª E 7ª SÉRIES DO 1º GRAU EM PORTUGUÊS: CICLO DE 1990 CALIBRADO E 1993 .....	119
GRÁFICO 2.3	ó DESEMPENHO DE PORTUGUÊS NAS 5ª E 7ª SÉRIES DO 1º GRAU EM CIÊNCIAS: CICLO DE 1990 CALIBRADO E 1993 .....	119
GRÁFICO 2.4	ó 4ª SÉRIE DO 1º GRAU, MATEMÁTICA: PERCENTUAL DE ALUNOS QUE SE ENCONTRAM ACIMA DOS NÍVEIS DE PROFICIÊNCIA 150, 225 E 300 ó REGIÕES E BRASIL ó 1995.....	127
GRÁFICO 2.5	ó 8ª SÉRIE DO 1º GRAU, MATEMÁTICA: PERCENTUAL DE ALUNOS QUE SE ENCONTRAM ACIMA DOS NÍVEIS DE PROFICIÊNCIA 150, 225 E 300 ó REGIÕES E BRASIL ó 1995.....	127
GRÁFICO 2.6	ó 3ª SÉRIE DO 2º GRAU, MATEMÁTICA: PERCENTUAL DE ALUNOS QUE SE ENCONTRA ACIMA DOS NÍVEIS DE PROFICIÊNCIA 150, 225, 300 E 375 ó REGIÕES E BRASIL ó 1995.....	127
GRÁFICO 2.7	ó PERCENTAGEM DE ALUNOS ACIMA DO NÍVEL 175 NA ESCALA DE MATEMÁTICA BRASIL E REGIÕES ó CICLO DE 1997 .....	137
GRÁFICO 2.8	ó PERCENTAGEM DE ALUNOS ACIMA DO NÍVEL 250 NA ESCALA DE MATEMÁTICA BRASIL E REGIÕES ó CICLO DE 1997 .....	137
GRÁFICO 2.9	ó PERCENTAGEM DE ALUNOS ACIMA DO NÍVEL 325 NA ESCALA DE MATEMÁTICA BRASIL E REGIÕES ó CICLO DE 1997 .....	137
GRÁFICO 2.10	ó PROFICIÊNCIA MÉDIA POR SÉRIE EM MATEMÁTICA. BRASIL E REGIÕES ó 1995-1997.....	140
GRÁFICO 2.11	ó EVOLUÇÃO DA PROFICIÊNCIA MÉDIA EM LÍNGUA PORTUGUESA NO SAEB/PROVA BRASIL ó ESCOLAS PÚBLICAS ó 1995 A 2011 .....	156
GRÁFICO 2.12	ó EVOLUÇÃO DA PROFICIÊNCIA MÉDIA EM MATEMÁTICA NO SAEB/PROVA BRASIL ó ESCOLAS PÚBLICAS ó 1995 A 2011 .....	156
GRÁFICO 3.1	ó PROFICIÊNCIA MÉDIA DE LÍNGUA PORTUGUESA NO PROEB PARA OS 5º E 9º ANOS DO ENSINO FUNDAMENTAL DA REDE ESTADUAL ó 2000 A 2011 .....	228
GRÁFICO 3.2	ó PROFICIÊNCIA MÉDIA DE MATEMÁTICA NO PROEB PARA OS 5º E 9º ANOS DO ENSINO FUNDAMENTAL DA REDE ESTADUAL ó 2000 A 2011 .....	229
GRÁFICO 3.3	ó PROFICIÊNCIA EM LÍNGUA PORTUGUESA PARA OS 5º E 9º ANOS DA REDE ESTADUAL DE ENSINO FUNDAMENTAL NO SAEB E NO PROEB ENTRE 2007 E 2011 .....	230
GRÁFICO 3.4	ó PROFICIÊNCIA MÉDIA EM MATEMÁTICA PARA OS 5º E 9º ANOS DA REDE ESTADUAL DE ENSINO FUNDAMENTAL NO SAEB E NO PROEB ENTRE 2007 E 2011 .....	230
GRÁFICO 3.5	ó PERCENTUAL DE ALUNOS DISTRIBUÍDOS POR PADRÃO DE DESEMPENHO PARA OS ANOS 5º E 9º DO ENSINO FUNDAMENTAL E PARA A 3ª SÉRIE DO ENSINO MÉDIO DA REDE ESTADUAL NA ÁREA DE LÍNGUA PORTUGUESA NO PROEB, ENTRE 2008 E 2010.....	234
GRÁFICO 3.6	ó PERCENTUAL DE ALUNOS DISTRIBUÍDOS POR PADRÃO DE DESEMPENHO PARA OS ANOS 5º E 9º DO ENSINO FUNDAMENTAL E PARA A 3ª SÉRIE DO ENSINO MÉDIO DA REDE ESTADUAL NA ÁREA DE MATEMÁTICA NO PROEB, ENTRE 2008 E 2010.....	234
GRÁFICO 3.7	ó PROFICIÊNCIA MÉDIA NO PROALFA PARA OS ALUNOS DO 2º, DO 3º E DO 4º ANO BD DA REDE ESTADUAL DO ENSINO FUNDAMENTAL, ENTRE 2005 E 2011 .....	239

GRÁFICO 3.8	ó GANHO DE PROFICIÊNCIA MÉDIA NA PASSAGEM DO 2º PARA O 3º ANO DA REDE ESTADUAL DO ENSINO FUNDAMENTAL, MEDIDO PELO PROALFA .....	240
GRÁFICO 3.9	ó GANHO DE PROFICIÊNCIA MÉDIA NA PASSAGEM DO 2º PARA O 3º E DO 3º PARA O 4º ANO DO ENSINO FUNDAMENTAL DA REDE ESTADUAL, MEDIDO PELO PROALFA .....	241
GRÁFICO 3.10	ó PROFICIÊNCIA MÉDIA NO SARESP EM LEITURA PARA OS 5º, 7º E 9º ANOS DO ENSINO FUNDAMENTAL E PARA A 3ª SÉRIE DO ENSINO MÉDIO ó 2007 A 2011.	267
GRÁFICO 3.11	ó PROFICIÊNCIA MÉDIA NO SARESP EM MATEMÁTICA PARA OS 5º, 7º E 9º ANOS	267
GRÁFICO 3.12	ó DISTANCIAMENTO DAS MÉDIAS DE PROFICIÊNCIA AFERIDAS NO SARESP 2011, EM RELAÇÃO À EXPECTATIVA DO NÍVEL ADEQUADO PARA OS ANOS/SÉRIES TESTADOS EM LEITURA .....	269
GRÁFICO 3.13	ó DISTANCIAMENTO DAS MÉDIAS DE PROFICIÊNCIA AFERIDAS NO SARESP 2011, EM RELAÇÃO À EXPECTATIVA DO NÍVEL ADEQUADO PARA OS ANOS/SÉRIES TESTADOS EM MATEMÁTICA .....	269
GRÁFICO 3.14	ó PROFICIÊNCIA NO SARESP EM GEOGRAFIA PARA OS 7º E 9º ANOS DO ENSINO FUNDAMENTAL E PARA A 3ª SÉRIE DO ENSINO MÉDIO ó 2009 A 2011.....	270
GRÁFICO 3.15	ó PROFICIÊNCIA NO SARESP EM HISTÓRIA PARA OS 7º E 9º ANOS DO ENSINO FUNDAMENTAL E PARA A 3ª SÉRIE DO ENSINO MÉDIO ó 2009 A 2011.....	270
GRÁFICO 3.16	ó PROFICIÊNCIA NO SARESP EM CIÊNCIAS E CIÊNCIAS DA NATUREZA EM 2008 E 2010 PARA OS 7º E 9º ANOS DO ENSINO FUNDAMENTAL E PARA A 3ª SÉRIE DO ENSINO MÉDIO .....	271

## LISTA DE SIGLAS E ABREVIATURAS

ANA	Avaliação Nacional de Alfabetização
ANPED	Associação Nacional de Pós-Graduação e Pesquisa
APRECE	Associação de Municípios e Prefeitos do Ceará
AYP	Adequate Yearly Progress
CAEd	Centro de Políticas Públicas e Avaliação da Educação
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CBA	Ciclo Básico de Alfabetização
CBC	Currículo Básico Comum
CEB	Câmara de Educação Básica
CCSSO	Council of Chief State School Officers
CEALE	Centro de Alfabetização, Leitura e Escrita
CENP	Coordenadoria de Normas Pedagógicas
CETPP	Centro de Estudos de Testes e Pesquisas Psicológicas
CNE	Conselho Nacional de Educação
CONSED	Conselho Nacional dos Secretários de Educação
ECIEL	Programa de Estudos Conjuntos de Integração Econômica Latinoamericana
EDURURAL	Programa de Educação Básica para o Nordeste Brasileiro
ENEM	Exame Nacional do Ensino Médio
EUA	Estados Unidos da América
GERES/2005	Estudo Longitudinal da Geração Escolar 2005
GERM	Global Educational Reform Movement
FCC	Fundação Carlos Chagas
FNDE	Fundo Nacional de Desenvolvimento da Educação
GERAES	Gestão Estratégica dos Recursos e Ações do Estado
IBGE	Instituto Brasileiro de Geografia e Estatística
IDEB	Índice de Desenvolvimento da Educação Básica
IDESP	Índice de Desenvolvimento da Educação do Estado de São Paulo
IEA	International Association for the Evaluation of Educational Achievement
IFES	Instituições Federais de Educação Superior
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
ICMS	Imposto sobre Circulação de Mercadorias e Serviços



IPI	Imposto sobre Produtos Industrializados
IPVA	Imposto sobre Veículos Automotores
IRRE	Institute for Research and Reform in Education
ITCD	Imposto sobre Transmissão Causa Mortis e Doação de Quaisquer Bens ou Direitos
ITR	Imposto Territorial Rural
LDB	Lei de Diretrizes e Bases da Educação Nacional
LLECE	Laboratório Latino Americano de Evaluación de la Calidad de la Educación
MEC	Ministério da Educação
NAEP	National Assessment of Educational Progress
NCLB	No Child Left Behind
NGA	National Governors Association
OAB	Ordem dos Advogados do Brasil
OECD	Organization for Economic Co-Operation and Development
OREALC	Oficina Regional de Educación para América Latina y El Caribe
PAAE	Programa de Avaliação da Aprendizagem Escolar
PAIC	Programa de Alfabetização na Idade Certa
PAIE	Programa de Apoio a Inovações Educacionais
PAR	Plano de Ações Articuladas
PARFOR	Plano Nacional de Formação de Professores da Educação Básica
PDE	Plano de Desenvolvimento da Educação
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PL	Projeto de Lei
PMDI	Plano Mineiro de Desenvolvimento Integrado
PNAIC	Plano Nacional de Alfabetização na Idade Certa
PPA	Plano Plurianual
PQE	Programa Qualidade das Escolas
PROALFA	Programa de Avaliação da Alfabetização
PROEB	Programa de Avaliação da Rede Pública de Educação Básica
PROFA	Programa de Formação de Professores Alfabetizadores
PSDB	Partido da Social Democracia Brasileira
PT	Partido dos Trabalhadores
SAEB	Sistema de Avaliação da Educação Básica

SAEP	Sistema de Avaliação do Ensino Público de 1º Grau
SIAPE	Sistema de Ação Pedagógica
SIMAVE	Sistema de Mineiro de Avaliação da Educação Pública
SARESP	Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo
TAAS	Texas Assessment of Academic Skills
TCT	Teoria Clássica do Testes
TIMMS	Trends in International Mathematics and Science Study
TRI	Teoria de Resposta ao Item
UAB	Universidade Aberta do Brasil
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNDIME	União dos Dirigentes Municipais de Educação
UNICEF	Fundo das Nações Unidas para a Infância
VUNESP	Fundação para o Vestibular da Universidade Estadual Paulista
WEI	World Education Indicators

## SUMÁRIO

<b>INTRODUÇÃO .....</b>	<b>21</b>
<b>CAPÍTULO 1</b>	
<b>TEMAS QUE AFETAM A AVALIAÇÃO EDUCACIONAL.....</b>	<b>36</b>
EDUCAÇÃO COMO DIREITO E O DIREITO À APRENDIZAGEM .....	36
REGULAÇÃO .....	40
A AVALIAÇÃO EDUCACIONAL E A AVALIAÇÃO EXTERNA DA EDUCAÇÃO.....	44
A EMERGÊNCIA DA AVALIAÇÃO EXTERNA.....	49
<i>A trajetória da avaliação: da época da criação dos sistemas educacionais nacionais até a época da profissionalização.....</i>	<i>49</i>
<i>A fase da prestação de contas e da responsabilização (accountability).....</i>	<i>56</i>
<i>As políticas com consequências (high-stakes) .....</i>	<i>64</i>
Medidas de valor agregado .....	64
Padrões Estaduais Básicos Comuns (Common Core State Standards) .....	66
Rankings e ampliação dos testes.....	68
Soluções para as escolas de baixo desempenho .....	69
Formação de professoras e sua retenção nas escolas.....	73
Burlando o teste: ensinando para o teste e as fraudes.....	75
Quem ganha com as políticas high-stakes? .....	77
<i>O accountability é o único caminho? .....</i>	<i>81</i>
<b>PARA SE COMPREENDER OS INSTRUMENTOS UTILIZADOS.....</b>	<b>89</b>
<i>A Teoria da Resposta ao Item (TRI) .....</i>	<i>91</i>
<i>A pré-testagem dos itens.....</i>	<i>92</i>
<i>A montagem dos testes.....</i>	<i>93</i>
<i>A escala de proficiência e sua interpretação pedagógica.....</i>	<i>94</i>
<b>CAPÍTULO 2</b>	
<b>O GOVERNO FEDERAL E A ORGANIZAÇÃO DOS TESTES AVALIATIVOS.....</b>	<b>97</b>
<b>A EVOLUÇÃO DOS TESTES AVALIATIVOS PRODUZIDOS PELA UNIÃO.....</b>	<b>97</b>
<i>O desenvolvimento da expertise em avaliação educacional.....</i>	<i>98</i>
<i>O surgimento da avaliação educacional na legislação.....</i>	<i>100</i>
<i>Da criação do Saeb até a Prova Brasil.....</i>	<i>107</i>
Primeira fase: de 1988 a 1993 .....	107
Ciclo de 1990 .....	113
Ciclo de 1993 .....	116
Segunda fase: de 1995 a 2003 .....	122
Ciclo de 1995 .....	122
Ciclo de 1997 .....	129
Ciclo de 1999 .....	141
Ciclo 2001.....	147
Ciclo de 2003 .....	150
3. Terceira fase: a partir de 2005 .....	152
4. Mudanças na composição dos testes entre 1995 e 2011 .....	153
<i>A evolução do desempenho do sistema educacional.....</i>	<i>155</i>
<i>Provinha Brasil.....</i>	<i>159</i>
<b>O ÍNDICE DE DESENVOLVIMENTO DA EDUCAÇÃO BÁSICA (IDEB) .....</b>	<b>164</b>
<b>AS AÇÕES DO GOVERNO FEDERAL QUE UTILIZAM OS RESULTADOS DAS AVALIAÇÕES .....</b>	<b>172</b>
<i>Plano de Desenvolvimento da Educação (PDE).....</i>	<i>173</i>
<i>O Pacto Nacional de Alfabetização na Idade Certa (Pnaic).....</i>	<i>181</i>
<b>A DISCUSSÃO DA QUALIDADE EDUCACIONAL NA MÍDIA ELETRÔNICA ESCRITA.....</b>	<b>187</b>

### **CAPÍTULO 3**

<b>OS TESTES AVALIATIVOS NOS ESTADOS DE MINAS GERAIS E SÃO PAULO .....</b>	<b>211</b>
<b>O ESTADO DE MINAS GERAIS.....</b>	<b>212</b>
<i>As primeiras pesquisas para avaliar o ensino, a consolidação e o abandono do Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais – 1988 a 1996.....</i>	<i>212</i>
<i>A modificação do Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais e a implantação do Simave.....</i>	<i>221</i>
<i>Os instrumentos do Simave.....</i>	<i>225</i>
Programa de Avaliação da Rede Pública de Educação Básica (Proeb).....	225
A expressão dos resultados do Proeb.....	228
Programa de Avaliação da Alfabetização.....	235
A expressão dos resultados do Proalfa.....	235
Programa de Avaliação da Aprendizagem Escolar.....	242
As aplicações dos diferentes tipos de testes que compõem o Simave.....	244
As ações que se baseiam nos resultados das avaliações educacionais.....	246
Acordo de Resultados.....	246
Programa de Intervenção Pedagógica.....	248
Divulgação dos resultados do Ideb 2011 nas escolas.....	250
<b>O ESTADO DE SÃO PAULO.....</b>	<b>253</b>
<i>O Programa de Avaliação da Rede Estadual Paulista.....</i>	<i>254</i>
A criação do Saresp e seu desenvolvimento.....	256
Os níveis de proficiência do Saresp e as expectativas de aprendizagem.....	264
Os resultados do Saresp.....	266
As ações e os programas que utilizam os resultados do Saresp à luz da legislação.....	271
O Idesp, o estabelecimento de metas e o cálculo para a bonificação por resultados.....	277

### **CAPÍTULO 4**

<b>ANÁLISE COMPARADA DOS CASOS.....</b>	<b>281</b>
---	------------

<b>CONSIDERAÇÕES FINAIS.....</b>	<b>291</b>
----------------------------------	------------

<b>REFERÊNCIAS.....</b>	<b>301</b>
-------------------------	------------

#### **APÊNDICE A**

<b>O PROCESSO DE CONSTRUÇÃO DA INTERPRETAÇÃO DAS ESCALAS DO SAEB E OS PROBLEMAS SURTIDOS.....</b>	<b>329</b>
---	------------

#### **APÊNDICE B**

<b>ROTATIVIDADE DE PROFESSORAS NO ENSINO FUNDAMENTAL.....</b>	<b>335</b>
---	------------

#### **APÊNDICE C**

<b>OS POSSÍVEIS IMPACTOS DO INSTRUMENTAL UTILIZADO NAS AVALIAÇÕES EXTERNAS SOBRE O CURRÍCULO E A APRENDIZAGEM.....</b>	<b>342</b>
AS MATRIZES DE REFERÊNCIA DO SARESP E SUA RELAÇÃO COM OS CURRÍCULOS.....	342
OS NÍVEIS DE PROFICIÊNCIA DO SARESP E SEUS IMPACTOS SOBRE A APRENDIZAGEM.....	350

## INTRODUÇÃO

O processo educacional formal é complexo e envolve não somente a participação da aluna, do aluno e da professora,<sup>1</sup> mas também da família, da comunidade escolar, do conjunto da sociedade e do Estado. Esses temas são objeto de estudo em todo o mundo.

A avaliação da aprendizagem não é algo novo dentro do processo educacional; está instituída há muito tempo, está presente no dia a dia das professoras, das alunas e dos alunos e, nos últimos cem anos, seus conceitos e objetivos vêm mudando de acordo com as transformações políticas e econômicas da sociedade. As professoras têm a seu dispor uma vasta produção sobre diferentes processos avaliativos que oferecem uma base sólida de conhecimentos e de técnicas que podem ser aplicados em diferentes situações, fornecendo alternativas para aprimorar o trabalho em sala de aula e sendo capazes de aprimorar o aprendizado de alunas e alunos. Além disso, em diversas situações, e por diversas razões, as professoras também passam por avaliações ao longo de sua carreira. Para os estudantes, a avaliação deveria ser um momento de reflexão sobre seu processo de aprendizagem, de forma a garantir que passos futuros possam ser dados com segurança, ou então sobre a necessidade de aprofundar os estudos realizados até aquele momento.

Todavia, a avaliação não tem ficado restrita ao interior da escola. A partir dos anos 1980, começa a ganhar força uma forma de avaliação conhecida como *avaliação educacional externa*: a partir de diversos instrumentos, em que o mais destacado é o teste padronizado aplicado às alunas e aos alunos, procura-se obter informações sobre o desempenho das escolas e dos sistemas de ensino, muitas vezes com o objetivo de subsidiar as políticas educacionais. Os testes padronizados, compostos por vários itens, têm sua elaboração e aplicação realizada por uma instituição externa à escola, não contando, normalmente, com a participação das professoras, das equipes gestoras das escolas e dos sistemas de ensino, em nenhuma de suas fases. Além dos testes, normalmente utilizam-se questionários aplicados às alunas e aos alunos, às professoras e à equipe gestora da escola, com o objetivo de obter informações que expliquem os diferentes desempenhos no interior do sistema escolar e de propor ações e políticas educacionais pertinentes. É importante destacar que essas explicações não estabelecem uma relação de causalidade entre o desempenho da aluna e do aluno e o contexto que o produziu, mas podem indicar caminhos que permitam aos governos selecionar entre as

---

<sup>1</sup> Como a maior parte da categoria profissional docente na educação básica brasileira é formada por mulheres, com cerca de 86% do total (INEP, 2007), este trabalho utilizará a flexão professora para se referir ao conjunto de professoras e professores.

diversas ações e políticas educacionais aquelas que tenham mais possibilidades de sucesso para promover melhorias no sistema educacional.

Avaliar é um conceito em si mesmo polissêmico, algumas vezes confundido com os instrumentos usados para medir resultados educacionais, geralmente por meio de exames ou testes. Avaliação educacional, neste trabalho, significa ir além das medições ou da apresentação de resultados e envolve a definição de políticas educacionais e de estratégias governamentais, levando ao aperfeiçoamento institucional e com capacidade de influir inclusive no processo de ensino-aprendizagem (HORTA NETO, 2010).

Um teste interno, instrumento utilizado para medir a aprendizagem, conduzido por uma professora e envolvendo alunas e alunos, tem características diferentes de um teste externo. Aquele, aplicado pela professora, é um evento único e independente, com ênfase em um conjunto de objetivos cognitivos que incidem sobre o futuro imediato da aluna e do aluno e que, normalmente, é interpretado independentemente de seu contexto socioeconômico ou educacional. Já o teste externo faz parte de um conjunto de instrumentos utilizados em um levantamento periódico de informações, com o objetivo de captar a evolução do quadro educacional, que tem como principais usuários, mas não únicos e exclusivos, os gestores dos sistemas educacionais.

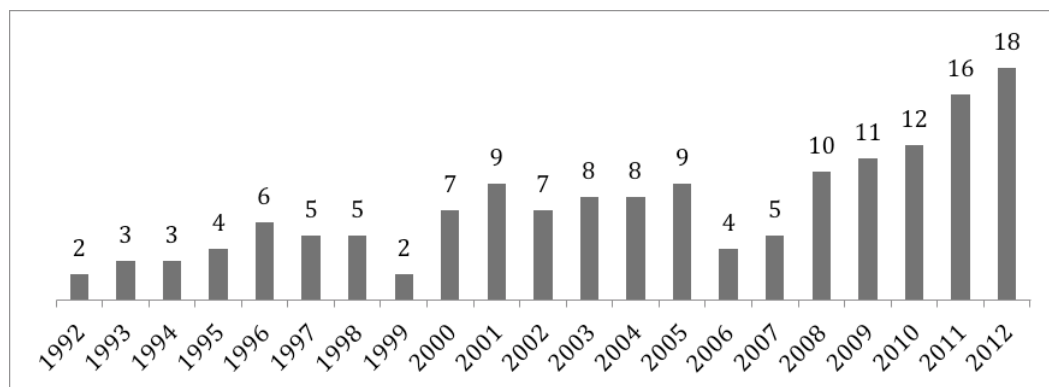
Os testes fornecem dados que precisam ser trabalhados, de forma a se transformarem em informações que possibilitem compreender o processo como um todo e sugerir alternativas para as políticas. Assim, é necessária uma análise técnica e pedagógica sobre os itens que se utilizou para identificar processos que precisam ser aprimorados. Os resultados obtidos devem ensejar ações que possam efetivamente modificar os processos no sentido de garantir o direito à educação em sua plenitude, garantindo o acesso e a permanência na escola e a aprendizagem.

Pesquisas educacionais desenvolvidas no Brasil e em outras latitudes vêm demonstrando que o desempenho das alunas e dos alunos na educação básica depende de um amplo conjunto de fatores que envolvem a participação da família, os hábitos de estudo, o envolvimento da comunidade escolar no processo educacional, a formação das professoras, as condições socioeconômicas das alunas e dos alunos, entre outros. Esses fatores, quando estudados em conjunto com os resultados dos testes, apresentam uma imagem mais nítida, apesar de ainda limitada, do estágio atual do processo educacional e fornecem, no caso dos testes externos, informações para o desenvolvimento de políticas educacionais. No entanto, apesar de existirem múltiplos fatores que influenciam o desempenho das alunas e dos alunos,

não invalida o fato de que o desempenho em si é capaz de indicar problemas que devem ser mais bem estudados e corrigidos.

Discutir o desempenho dos sistemas educacionais a partir de testes aplicados exige clareza sobre todo o processo de sua realização e aplicação e qual seu sentido. Essa não é uma discussão trivial, pois discute a função social da escola e as modificações pelas quais ela vem passando ao longo do tempo e o papel do Estado nesse processo. Discute também as razões que estão levando cada vez mais os governos a se interessar pelas avaliações educacionais (HORTA NETO, 2010a). Por que o uso dos testes externos e não outro instrumento? Esse será um dos temas tratados neste trabalho.

No Brasil, nos últimos anos, tem-se ampliado o número de governos municipais e estaduais e de gestores das redes privadas de ensino que utilizam testes externos cada vez mais sofisticados para medir o desempenho dos estudantes. Para se ter uma ideia desse aumento, construiu-se o Gráfico 1.1, que retrata o avanço do número de Estados que contam com avaliações próprias durante o período em análise.



**Gráfico 1.1 – Quantidade de Estados, por ano, que contam com testes avaliativos próprios – 1992 a 2012**

Fonte: BROOKE; CUNHA; FALEIROS, 2011

2012- elaboração do autor.

Conforme se observa no Gráfico 1.1, existe inconstância no número de Estados que fazem uso dos testes com o passar do tempo, fruto, principalmente, de mudanças de governo, que priorizam ou não seu uso. De toda forma, é a partir de 2008 que cresce o número de Estados a utilizar os testes.

Em 1992, os Estados de Minas Gerais e Ceará começam a aplicação dos testes avaliativos. Nos dois anos seguintes, 1993 e 1994, além dos dois Estados, desenvolve-se o teste em Mato Grosso do Sul e, em 1995, em São Paulo (BROOKE; CUNHA; FALEIROS, 2011).

Em 2012, conforme levantamento desta pesquisa, eram os seguintes os Estados e os respectivos programas: Acre ó Sistema Estadual de Avaliação da Aprendizagem Escolar (Seape); Alagoas ó Sistema de Avaliação Educacional de Alagoas (Saveal); Amazonas ó Sistema de Avaliação do Desempenho Educacional do Amazonas (Sadeam); Bahia ó Sistema de Avaliação Baiano de Educação (Sabe); Ceará ó Sistema Permanente de Avaliação da Educação Básica do Ceará (Spaace); Espírito Santo ó Programa de Avaliação da Educação Básica do Espírito Santo (Paebes); Goiás ó Sistema de Avaliação Educacional do Estado de Goiás (Saego); Minas Gerais ó Sistema Mineiro de Avaliação da Educação Pública (Simave); Mato Grosso do Sul ó Sistema de Avaliação da Rede Pública de Mato Grosso do Sul (Saems); Paraíba ó Sistema de Avaliação da Educação da Paraíba, Avaliando Idepb; Paraná ó Sistema de Avaliação da Educação Básica do Paraná (Saep); Pernambuco ó Sistema de Avaliação Educacional de Pernambuco (Saepe); Piauí ó Sistema de Avaliação Educacional do Piauí (Saepi); Rio de Janeiro ó Sistema de Avaliação da Educação do Estado do Rio de Janeiro (Saerj); Rio Grande do Sul ó Sistema Estadual de Avaliação Participativa (Seap); Rondônia ó Sistema de Avaliação Educacional de Rondônia (Saero); São Paulo ó Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (Saresp); Tocantins ó Sistema de Avaliação da Educação do Tocantins (Salto).

Esses Estados, com exceção do Espírito Santo, que usa o termo *programa*, adotam a expressão *sistemas de avaliação*. Apesar de serem chamados de sistemas, utilizam apenas testes aplicados aos alunos, diferenciados por ano escolar testado, sendo o Estado do Rio Grande do Sul o único que utiliza uma série de instrumentos que fornecem algumas dezenas de indicadores. Além disso, mesmo se referindo a uma avaliação, concentram suas atividades na aplicação de instrumentos de medida, desenvolvidos, na maioria das vezes, por entidades externas à Secretaria de Educação. Portanto, a discussão é necessária, posto que a atividade de avaliar vai além de simplesmente medir e divulgar os resultados. Por essa razão, neste trabalho será frequente o uso da expressão *teste avaliativo* com referência ao instrumento de medida utilizado em avaliações educacionais.

Com os desempenhos obtidos pelas escolas em mão, alguns governos começaram a utilizar estratégias de premiação com o objetivo declarado de estimular resultados escolares cada vez melhores. Exemplos disso são os programas Escola Nota 10, no Ceará; Prêmio Escola de Valor, no Amazonas; Prova Rio, no Rio de Janeiro; e Índice de Qualidade Educacional, no Ceará. Baseados também nos resultados das avaliações, passam a ser instituídos programas de incentivos e pagamento de bônus para a equipe escolar, quando a escola alcança metas fixadas, normalmente de forma unilateral, pelo órgão gestor do sistema.



São exemplos dessas políticas os programas criados nos Estados do Amazonas, Ceará, Pernambuco, Rio de Janeiro, Minas Gerais, Espírito Santo, São Paulo e alguns municípios, entre outros. Outras estratégias conhecidas são a distribuição de prêmios e o sorteio de brindes aos alunos.

Desde 1988, ano dos primeiros estudos para a aplicação do Sistema de Avaliação da Educação Básica (Saeb), utilizado desde 1990, vem aumentando a quantidade de testes aplicados e o número de alunas e de alunos testados, tanto em nível nacional como subnacional (HORTA NETO, 2006). Depois do Saeb, criaram-se a Prova Brasil, a Provinha Brasil e a Avaliação Nacional de Alfabetização, ANA, um novo teste ligado ao Pacto Nacional pela Alfabetização na Idade Certa (Pnaic), cujas propostas serão discutidas no trabalho. Outros exames vão sendo criados e aprimorados: o Exame Nacional dos Estudantes do Ensino Médio (Enem) e o Exame Nacional de Certificação de Competências da Educação de Jovens e Adultos (Encceja), voltados para o ensino médio, e o Exame Nacional de Desempenho do Estudante (Enade), voltado para a educação superior. Como se percebe, é uma quantidade significativa de instrumentos que o governo federal tem a sua disposição, gerando grande volume de dados sobre o desempenho dos alunos, todos produzidos pelo Instituto Nacional de Estudos e Pesquisas Educacionais (Inep), autarquia ligada ao Ministério da Educação (MEC). Além dos instrumentos listados, existem outros para medir a evolução dos cursos de pós-graduação mediante a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes).

Em outras partes do mundo, os testes vêm sendo utilizados com frequência, sendo desenvolvidos pelos governos nacionais ou por organizações internacionais ou multilaterais. Nos EUA, os testes externos vêm sendo usados desde os anos 1960, o mesmo ocorrendo, nessa época, em alguns países da Europa. Na Oceania, a partir dos anos 1980. Também começam a ser implantados com bastante força na América Latina no fim dos anos 1980, conforme indicam os dados do Quadro 1.1.

### Quadro 1.1 – Início da aplicação dos testes externos na América Latina

Ano de aplicação do primeiro teste avaliativo externo	País
1975	Cuba
1980	Colômbia
1981	Panamá
1982	Chile
1988	Brasil e Costa Rica
1990	Honduras
1991	República Dominicana
1992	Guatemala
1993	El Salvador e Argentina
1995	México
1996	Bolívia, Equador, Paraguai, Nicarágua, Uruguai e Peru.
1998	Venezuela

Fonte: FERRER, 2006, com elaboração do autor.

Além de criar testes avaliativos externos próprios, diversos países vêm participando de alguns estudos internacionais. Entre os principais estudos destacam-se:

6 A International Association for the Evaluation of Educational Achievement (IEA), entidade civil com sede na Holanda, fundada há 60 anos, que tem 62 institutos de pesquisas educacionais governamentais e universitários como membros efetivos. Outros 20 países participam dos estudos desenvolvidos pela IEA. Os dois maiores estudos são o Progress in International Reading Literacy Study (Pirls), que mede o desempenho de alunas e de alunos da 4ª série em Leitura, e o Trends in International Mathematics and Science Study (Timms), que mede o desempenho de alunas e de alunos das fases inicial e final do ensino fundamental em Matemática e Ciências. No Quadro 1.2, apresenta-se o número total de países participantes em cada estudo até 2011, assim como o nome dos países latino-americanos.

**Quadro 1.2 – Países participantes das pesquisas conduzidas pelo IEA**

Pesquisa	Ano	Total de países participantes	Países latino-americanos
Pirls	2001	41	Argentina, Colômbia
	2006	50	Nenhum
	2011	49	Colômbia
Timms	1995	45	Argentina, Colômbia e México.
	1999	38	Chile
	2001	49	Argentina e Chile
	2006	63	Colômbia
	2011	64	Chile

Fonte: Elaboração do autor, com base em informações disponíveis em <www.iea.nl>.

ó O Programme for International Student Assessment (Pisa), patrocinado pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE), do qual participam, além dos países-membros, alguns países associados. O primeiro estudo foi realizado em 2000 e vem sendo repetido a cada três anos, examinando o desempenho das alunas e dos alunos de 15 anos de idade, época em que as alunas e os alunos devem estar concluindo a educação obrigatória. Cada estudo concentra-se em uma área do conhecimento. No ano 2000, o foco foi Leitura; em 2003, Matemática; em 2006, Ciências. A partir daí, o ciclo se repete, como foi dito, trienalmente. A participação de países latino-americanos, com destaque para o Brasil e o México, que participaram de todas as edições, pode ser observada no Quadro 1.3.

**Quadro 1.3 – Países participantes do PISA – 2000 a 2012**

Ano	Total de países participantes	Países latino-americanos
2000	43	Argentina, Brasil, Chile, México e Peru.
2003	41	Brasil, México e Uruguai.
2006	57	Argentina, Brasil, Chile, Colômbia, México e Uruguai.
2009	75 <sup>(*)</sup>	Argentina, Brasil, Chile, Colômbia, México, Panamá, Peru, Venezuela e Uruguai.
2012	64	Argentina, Brasil, Chile, Colômbia, Costa Rica, México, Peru e Uruguai

Fonte: Elaboração do autor, com base em informações disponíveis em <www.pisa.oecd.org>.

Nota: (\*) Em 2009, participaram 65 países; em 2010, o teste foi aplicado em 10 outros países.

ó O Laboratório Latino Americano de Evaluación de la Calidad de la Educación (LLECE), programa coordenado pela Oficina Regional de Educación para América Latina y El Caribe (Orealc), entidade ligada à Unesco, criada em 1963, com sede no Chile, que tem como objetivo apoiar os 41 países-membros da região na definição de estratégias para o desenvolvimento de suas políticas educativas. Do LLECE participaram 18 países, entre eles Brasil, Argentina, Chile e México. O primeiro estudo foi realizado em 1997, o segundo em 2008. O terceiro estudo, conhecido como Terce, será aplicado em 2013, testando, assim como fizeram os dois primeiros, Leitura, Escrita e Matemática em estudantes do terceiro e do sexto ano de escolaridade e Ciências em estudantes do sexto ano. O destaque é a participação de Cuba, o único teste internacional em que o país participa.

Também alguns estudos com os dados estatísticos dos sistemas de diversos países vêm sendo produzidos. Dada a sua importância, dois deles se destacam:

ó *World Education Indicators* (WEI), estudo anual que apresenta dados educacionais produzidos pelo Unesco Institute of Statistics, órgão com sede no Canadá. Em 2008, o órgão produziu um longo estudo denominado Survey of Primary Schools, que envolveu 11 países, entre eles Brasil, Chile e Argentina, com o objetivo de entender e monitorar os fatores que influem na qualidade e na equidade da educação primária (UNESCO, 2008).

ó *Education at a Glance*, estudos coordenados pela OCDE, publicados trienalmente desde 2001, que fornecem uma visão geral sobre os sistemas educativos dos países que participaram do Pisa. Tratam de temas como investimentos em educação e mecanismos de acesso e permanência das alunas e dos alunos nas escolas, todos relacionados aos seus resultados obtidos no Pisa.

Nessa profusão de testes, pode acontecer de uma mesma escola testar seus alunos anualmente mais de três vezes, entre pré-testes, utilizados para avaliar a qualidade dos itens que compõem os testes, e testes municipais, estaduais, nacional e internacionais, como o Pisa e o Terce. Além disso, cria-se, também, um mercado da avaliação educacional que está em

franco desenvolvimento, oferecendo livros<sup>2</sup> e consultorias com o objetivo de preparar as escolas para a participação nesses testes.

No Brasil, passados mais de 20 anos do início dos testes para medir o desempenho dos sistemas educacionais, é importante refletir sobre os efeitos que o processo de massificação da sua utilização teve sobre as políticas educacionais. Discussão sobre a utilização ou não dos dados gerados pelos testes avaliativos é tão antiga quanto o são as abordagens avaliativas (DANTAS, 2009) e não é exclusividade do caso brasileiro. Pesquisas em diferentes países buscam respostas para essa questão.

Nesse contexto, esta pesquisa procurará responder à seguinte questão: quais os efeitos dos testes avaliativos aplicados entre 1990 e 2012 sobre as políticas educacionais nacionais e subnacionais voltadas para o ensino fundamental?

Outras indagações também são importantes para ajudar a iluminar o caminho da pesquisa. Que modificações sofrem os testes avaliativos com as mudanças de governo? Quais as semelhanças e as diferenças entre as ações que fazem uso dos resultados dos testes avaliativos envolvendo o governo nacional e os governos subnacionais? Quais as características principais de cada teste? Como a mídia eletrônica escrita vem tratando a avaliação educacional? Quais as semelhanças e as diferenças entre as avaliações educacionais no Brasil e em outros países?

Justifica-se o período utilizado, entre 1990 e 2012, para pesquisar os efeitos dos testes avaliativos, devido a 1990 ter sido o ano da primeira aplicação do Saeb e por 2012 refletir a época em que os testes passam a ganhar centralidade nos governos nacional e subnacionais. Foram escolhidas as políticas implementadas a partir de 2010, porque esse período de vinte anos, desde a primeira aplicação, parece ser um período razoável de amadurecimento dos processos ligados aos testes avaliativos, capazes de gerar influências sobre as políticas educacionais.

*A tese defendida neste trabalho* é capaz de responder às questões formuladas é a de que os testes avaliativos tiveram influência marginal sobre as políticas educacionais voltadas para o ensino fundamental, pois se tornaram um fim em si mesmo e não elementos centrais para a formulação de políticas educacionais. Outrossim, os resultados dos testes são usados pelo Estado como instrumento de regulação do trabalho realizado pela escola e seus profissionais.

---

<sup>2</sup> Um exemplo é o livro *Prova Brasil Saeb – Língua Portuguesa* anunciando que traz atividades para debate, leitura e interpretação de textos muito interessantes para crianças de 4ª a 5ª série. O objetivo é auxiliar no desenvolvimento das capacidades avaliadas na Prova Brasil, Saeb e em diversas avaliações sistêmicas Estaduais e Municipais de todo o Brasil. Disponível em: <<http://www.didako.com.br>>.

*O objetivo geral* da pesquisa é analisar o desenvolvimento dos testes avaliativos e como seus resultados são utilizados pelas políticas educacionais.

Foram definidos os seguintes objetivos específicos:

É Compreender o processo que colocou os testes avaliativos como ferramentas centrais na formulação de políticas.

É Identificar experiências internacionais em que os testes avaliativos tenham ou não centralidade.

É Compreender os fatores que influenciaram o surgimento dos testes avaliativos do governo federal e dos governos estaduais.

É Desvelar os pressupostos conceituais que embasaram a construção do Saeb e dos testes avaliativos subnacionais pesquisados, comparando suas similaridades e diferenças.

É Identificar as políticas educacionais formuladas pela União e por alguns Estados selecionados a partir de 2010.

Para o desenvolvimento da pesquisa, optou-se pelo método comparado. Esse método é utilizado por diversas áreas, acabando por se tornar, em cada uma, um campo de estudos próprio. Exemplos disso são a Educação Comparada, a Sociologia Comparada e a Política Comparada.

A pesquisa comparada é especialmente adequada para abordar problemas determinados por causas múltiplas e conjunturais, em que diferentes condições se combinam em diversificadas, e algumas vezes contraditórias, formas, para produzir o mesmo resultado ou um resultado similar (RAGIN, 1994). Contudo, deve-se levar em consideração que toda análise deve ser pautada pela objetividade, afastando-se de qualquer subjetividade (GARRIDO, 1996).

Os métodos comparados preocupam-se com a diversidade, entendida como o exame de semelhanças ou diferenças em um número moderado de casos. A questão sobre quantos casos utilizar é tratada por diversos autores, que afirmam que o máximo deve variar entre cinquenta (RAGIN, 1994) e dez (LIJPHART, 1971), mas, como não podia deixar de ser, no mínimo são necessários dois casos (GARRIDO, 1996). Dessa forma, surge outra pergunta com relação à quantidade adequada de casos para garantir validade aos achados da pesquisa. Essa pergunta está relacionada aos métodos quantitativos, que, para minimizar os erros estatísticos na análise da amostra, procuram utilizar o maior número possível de casos.

Segundo Peters (1998), a pergunta que sempre deve ser feita é o que a pesquisa ganha adicionando outra dúzia de casos ou mesmo um único caso à amostra, desprendendo-se do paradigma da pesquisa quantitativa. O autor sugere que se determine, primeiramente, quantos casos tomarão parte da análise e, em seguida, se selecione do universo inicial quais serão os casos analisados, adicionando-se novos casos na medida, se isso for necessário para aprofundar a pesquisa.

No caso desta pesquisa, optou-se por trabalhar com três casos. Primeiro, por conta da limitação de tempo, já que se iria buscar a maior quantidade de fontes documentais possíveis para caracterizar o desenvolvimento dos testes. Segundo, porque o trabalho de localizar esses documentos seria penoso, dada a profusão de pesquisas e instrumentos normativos que seriam necessários para analisar o objeto de estudo e pelo longo período analisado. Por fim, tão importante quanto os demais, é que se houvesse necessidade de deslocamento para pesquisar documentos que não estivessem disponíveis *on-line*, os gastos com viagem seriam um fator limitador.

O passo seguinte seria escolher os três casos, sabendo-se que, de uma forma geral, os casos devem ser escolhidos entre aqueles comparáveis e que eles devem também possuir fronteiras bem delimitadas espacial e temporalmente e serem coerentes entre si e com o que se quer pesquisar.

Para orientar essa escolha, Peters (1998) apresenta uma tipologia de estudos comparados. O primeiro tipo é aquele em que os casos são de um mesmo país e se utilizam as unidades subnacionais como unidades de análise; tem como principal vantagem manter constante ou minimizar um bom número de fatores que poderiam, de alguma forma, confundir a análise. Outro tipo refere-se à análise de processos e de instituições similares em um número limitado de países. Nesse caso, seleciona-se uma pequena quantidade de exemplos que parecem similares, e de algum modo importantes, de um processo ou de uma instituição, e esses exemplos são utilizados para iluminar a natureza tanto do processo como da instituição em si, ou então iluminar as políticas dos países em que eles ocorrem. Um terceiro tipo utiliza-se de tipologias ou de outras formas de esquemas de classificação, tanto para países como para unidades subnacionais, usando-as para comparar grupos de países ou para revelar algo sobre o fenômeno interno. O quarto tipo envolve a análise estatística ou descritiva dos dados de um subconjunto de países, normalmente selecionados por regiões ou por graus de desenvolvimento, testando algumas hipóteses sobre a relação de variáveis naquele conjunto de países. Por fim, o quinto tipo parte da análise estatística de vários países,

buscando desenvolver padrões e/ou testar relações para o fenômeno estudado, devendo-se tomar cuidado para não correlacionar variáveis que apresentem dificuldades de comparação.

Com base nessa discussão, escolheu-se comparar casos dentro de um mesmo país. Assim a escolha recairia sobre aqueles que, no Brasil, já aplicam testes avaliativos: o governo federal e os 18 estados.

O primeiro caso escolhido foi o do governo federal, porque o Saeb foi a primeira iniciativa estruturada de avaliação educacional com o objetivo de coletar informações sobre os sistemas educacionais, e também porque o Saeb está na origem de todos os sistemas de avaliação (BROOKE; CUNHA; FALEIROS, 2011). Em alguns Estados, como no Ceará, por exemplo, foram realizados estudos específicos, no final dos anos 1970 e começo dos anos 1980, que também utilizaram testes de desempenho aplicado aos alunos, mas seu objetivo foi avaliar programas específicos, não o sistema educacional (SOUZA; OLIVEIRA, 2007).

Como um dos objetivos foi verificar de que forma a coleta sistemática de dados estava influenciando o desenvolvimento de políticas educacionais que utilizavam os resultados dos testes em sua formulação, foram utilizados três critérios de seleção. O primeiro era o de antiguidade, e nesse caso se encontravam os Estados de Minas Gerais, Ceará, Mato Grosso do Sul e São Paulo. O segundo critério era o de abrangência, ou seja, que estivesse sendo testado o maior número de alunos e de anos escolares possíveis, isso porque se sabia que alguns sistemas de ensino, como no caso do Distrito Federal, utilizavam a justificativa de que o fato de não terem dados mais detalhados sobre o desempenho dos alunos impossibilitaria utilizar seus resultados para a definição de políticas (HORTA NETO, 2006). Com base nesse critério, eliminou-se o Estado de Mato Grosso do Sul, cujos testes foram aplicados a amostras de populações ora da 5ª série do ensino fundamental, ora da 8ª série, ora dos alunos do curso de Magistério, ou aplicados a alunos da 3ª série do ensino médio de uma única escola (BROOKE; CUNHA; FALEIROS, 2011). Apesar de o Estado do Ceará ter começado a utilização dos testes em 1992 juntamente com Minas Gerais, sua aplicação foi irregular. De início, os testes foram aplicados a uma amostra de escolas entre 1992 e 1994 e, nos anos seguintes, foram aplicados de forma irregular (SOUZA; OLIVEIRA, 2007). Por sua vez, São Paulo, apesar de ter começado a aplicar seus testes três anos depois, em 1995, o fez anualmente desde 1996, com apenas dois intervalos em 1999 e 2006.

Assim, os dois outros casos estudados, junto com o governo federal, foram Minas Gerais e São Paulo. Além de obedecer ao critério estipulado, os dois Estados respondiam, em 2012, por cerca de 40% da matrícula de alunos do ensino fundamental e também por cerca de



40% das professoras em exercício nessa etapa do ensino, portanto causam impacto importante sobre o restante do País.

Para levantar dados sobre os três casos, delimitaram-se algumas estratégias comuns a todos. A primeira delas foi reconstruir a trajetória histórica que marca o desenvolvimento dos testes, identificando suas principais influências e, a partir daí, seus princípios norteadores, comparando-os às prescrições legais, pois nem sempre o descrito na norma consegue ser aplicado na prática. A segunda estratégia foi acompanhar as diversas aplicações dos testes, para verificar as mudanças que estavam ocorrendo em seus princípios iniciais, procurando identificar, também, as causas dessas mudanças e, nesse processo, levantar os resultados obtidos nos testes, comparando-os no âmbito de cada Estado e, quando possível, com o resultado do Saeb e da Prova Brasil. A outra estratégia comum foi identificar, entre as ações desenvolvidas no interior dos órgãos gestores, aquelas que pudessem ter alguma relação com os resultados dos testes. Estabeleceu-se, também, que as bases para a obtenção dos dados seriam as informações constantes dos relatórios técnicos que divulgam os resultados dos testes e as normas legais emanadas pelos órgãos gestores e pelos legislativos da União e dos dois Estados.

Escolhidos os três casos a serem analisados, definiram-se quatro critérios para a comparação: 1) a periodicidade da avaliação; 2) sua abrangência em função da quantidade de alunos, dos anos escolares e das áreas do conhecimento testados; 3) a forma de apresentar os resultados; e 4) os tipos de responsabilização, de alta ou baixa consequência, perante os resultados alcançados.

Para verificar os efeitos dos resultados sobre as políticas educacionais adotadas a partir de 2010, foram utilizados também alguns critérios para a seleção das políticas que seriam analisadas, estabelecendo-se que seriam no mínimo duas por Ente da Federação. A quantidade e o tipo de beneficiários e o tipo de impacto que poderiam causar foram os critérios utilizados.

Assim, no caso do governo federal, escolheu-se o Plano de Desenvolvimento da Educação (PDE). Esse é um programa guarda-chuva que abriga uma série de ações e programas, alguns novos e outros muito antigos. Para sua operacionalização, serve-se do Plano de Metas Compromisso Todos pela Educação, que utiliza, entre seus indicadores, o resultado dos testes. Dentro do PDE, foi selecionado o Programa Nacional de Alfabetização na Idade Certa (Pnaic), um programa em que os testes têm um papel de destaque.

No caso de Minas Gerais, são analisadas três ações. A primeira é o Acordo de Resultados, um instrumento gerencial por meio do qual o governo do Estado acompanha o

cumprimento de metas preestabelecidas, inclusive metas para serem cumpridas por cada escola. A segunda é o Programa de Intervenção Pedagógica (PIP), que prevê ações nas escolas, para elevar o desempenho dos alunos nos testes. A terceira ação é relativa à divulgação do Índice de Desenvolvimento da Educação (Ideb) na porta das escolas estaduais.

Quanto a São Paulo, são discutidas duas ações. A primeira é o Índice de Desenvolvimento da Educação do Estado de São Paulo (Idesp), como instrumento para acompanhar o cumprimento de metas por escolas, baseada nos resultados dos testes, e a bonificação por resultados. A segunda ação é o estabelecimento da Matriz Curricular Básica.

O tema desse trabalho e a questão colocada me são muito caros e a eles venho dedicando boa parte do meu tempo desde minha chegada ao Inep, em 2004. Primeiro, moveu-me a curiosidade sobre como funcionava o Saeb, o que me levou a estudá-lo no curso de especialização. Em seguida, quis compreender como o sistema foi construído e de que forma ele estava influenciando a gestão de um sistema público de ensino, e aprofundi o estudo no mestrado. Para mim, que trabalho no Inep e que tenho uma preocupação muito grande com a função social dessa importante instituição de mais de 75 anos de existência, ter verificado que o Saeb era considerado um instrumento ilustre, mas desconhecido, foi um choque. Isso porque minha dissertação demonstrou que os gestores o consideravam importante, mas seus resultados não eram utilizados, pois pouco se entendia sobre seus instrumentos e o significado de seus resultados. Durante meu trabalho como servidor da Coordenação-Geral de Concepções Pedagógicas, da Diretoria de Avaliação da Educação Básica (Daeb) do Instituto, comecei a observar uma crescente utilização das avaliações, com o aumento de Estados utilizando sistemas próprios e o Inep ampliando a abrangência dos instrumentos que havia desenvolvido. Disso surgiu a inquietação em saber, depois de tanto tempo de aplicação e da quantidade de dados e informações gerados pelos testes, quais seus efeitos sobre as políticas educacionais, tema desta tese.

O desenvolvimento desta pesquisa contou com dois apoios fundamentais. Primeiramente, do Inep, que me concedeu licença para finalizá-la e financiou minha participação em diversos eventos onde apresentei parte dos resultados, na medida em que foram sendo obtidos (HORTA NETO, 2011, 2012 e 2013; HORTA NETO; YANNOULAS, 2012), e também no Ciclo de Debates 25 Anos de Avaliação de Sistemas Educacionais no Brasil, promovido pela Fundação Carlos Chagas (FCC), em 2012. O outro apoio veio do Programa Observatório da Educação (Obeduc), no contexto do Projeto Política Educacional e Pobreza, Edital Capes/Inep nº 038/2010.

O trabalho está estruturado em quatro capítulos. O primeiro trata dos temas ligados à avaliação educacional, utilizados para iluminar a pesquisa e a discussão de seus achados. Nele são discutidos o direito à educação e sua relação com o direito à aprendizagem; os processos de regulação e suas implicações nas ações dos governos; a emergência da avaliação educacional; o processo de *accountability* e seu reflexo nas ações de governos dos Estados Unidos e da Finlândia; e os instrumentos utilizados na montagem e na interpretação dos resultados dos testes avaliativos.

O segundo capítulo é dedicado à discussão dos testes avaliativos direcionados ao ensino fundamental, organizados pelo governo federal, e os principais programas federais que utilizam os resultados desses testes. Para isso, faz-se uma descrição da evolução dos testes avaliativos no Brasil, enfatizando o desenvolvimento da *expertise* nacional na área de avaliação educacional e como a legislação educacional vem dando suporte a esse processo de avaliação. Discute-se, também, a evolução do Sistema de Avaliação da Educação Básica (Saeb), por meio dos relatórios técnicos publicados e utilizados para divulgar seus resultados; o desenvolvimento da Provinha Brasil e seus resultados; os limites do Índice de Desenvolvimento da Educação Básica (Ideb) como um indicador de qualidade educacional; e alguns programas federais, para se compreender como os resultados dos testes avaliativos os influenciaram. Por fim, discutem-se as visões que são divulgadas sobre os resultados dos testes avaliativos, mediante matérias escritas disponibilizadas na internet, entre 2009 e 2012, em artigos de jornais, revistas, *sites* e *blogs*. Com relação à discussão sobre a divulgação dos resultados pela mídia, ele foi parcialmente publicado como capítulo do livro *Política educacional e pobreza: múltiplas abordagens para uma relação multideterminada*, organizado por Sílvia Yannoulas (2013).

No terceiro capítulo são apresentados e discutidos os testes avaliativos e algumas políticas desenvolvidas pelas redes estaduais dos Estados de Minas Gerais e São Paulo que utilizam seus resultados.

O quarto capítulo dedica-se a comparar os três casos pesquisados, do governo federal e dos Estados de Minas Gerais e São Paulo, e a seguir apresentam-se as conclusões da pesquisa.

## CAPÍTULO 1

### TEMAS QUE AFETAM A AVALIAÇÃO EDUCACIONAL

Neste capítulo são discutidos alguns temas que são essenciais para compreender o movimento que possibilitou à avaliação educacional ter atingido a centralidade que tem hoje. A educação, tema deste trabalho, é problematizada, principalmente a relação entre direito à educação e direito à aprendizagem, este último tema de alguns documentos do MEC. Os processos de regulação, com a contribuição do perito, e suas implicações nas ações dos governos também são discutidos como pano de fundo para o debate sobre a avaliação em suas diferentes fases até a última classificada como a fase do *accountability*. Nessa fase, discutem-se algumas ações desenvolvidas nos Estados Unidos, no Brasil, na Finlândia e em outros países, com relação à responsabilização e à prestação de contas. Por fim, apresenta-se uma breve explicação sobre os instrumentos que compõem os testes avaliativos, as técnicas adotadas para sua construção e a forma normalmente utilizada para explicar pedagogicamente os resultados numéricos obtidos como resultados.

#### ***Educação como direito e o direito à aprendizagem***

A educação, como um direito social, está presente na Constituição de 1988, por ser inerente à cidadania e aos direitos humanos. A educação básica, em que a educação infantil é sua raiz, o ensino fundamental seu tronco e o ensino médio seu acabamento. A educação foi declarada como direito do cidadão e dever do Estado. Como nos ensina Cury, esse direito vem sendo construído ao longo dos anos:

Da instrução própria das primeiras letras no Império, reservada apenas aos cidadãos, ao ensino primário de quatro anos nos Estados da Velha República, do ensino primário obrigatório e gratuito na Constituição de 1934, à sua extensão para oito anos em 1967, derrubando a barreira dos exames de admissão, chegamos ao direito público subjetivo e ao novo conceito ora analisado [*direito social*] (2008, p. 295).

Portanto, como todo direito social, a trajetória da educação foi longa e envolveu várias e distintas lutas, apesar de ter se generalizado, desde o século 19, em diversos países. Por

outro lado, foi um avanço, pois a educação passou a ser considerada um direito público subjetivo, ou seja, que deve haver escola para todos, significando que qualquer residente no Brasil tem o direito de requerer ao Estado a prestação educacional, sob pena de ser responsabilizada a autoridade competente para oferecê-la (VIEIRA, 2001, p. 9).

A educação, além de estar vinculada a um direito, é também uma obrigação, pois ninguém pode deixar de frequentar a escola entre os 4 e 17 anos de idade, de acordo com o artigo 208 da Constituição. No entanto, não é possível haver garantia do direito à educação sem que seja assegurado que sua oferta atenderá às necessidades dos beneficiados, e isso implica o papel ativo e responsável do Estado, formulando políticas e obrigando-se a oferecer ensino em igualdade de condições para todos. Aliás, a própria Constituição prevê que essa igualdade de condições se materialize, garantindo-se um padrão de qualidade, conforme o artigo 206, inciso VII.

Com relação à qualidade, dado à polissemia do termo, pode-se observar que com o passar do tempo, no Brasil, vários foram os significados que lhe foram sendo atribuídos. Oliveira e Araújo (2005) estabelecem três momentos. No primeiro, a qualidade estava associada à ampliação da oferta de vagas, acarretando, a partir dos anos 1940, a construção de prédios escolares. Com isso, ao mesmo tempo que se ampliou a oportunidade de escolarização para as populações mais pobres, marginalizadas da escola, os sistemas de ensino e as escolas não se prepararam para recebê-las, pois, dada a diferença de experiências culturais com os outros grupos que já frequentavam a escola, essas populações necessitavam de atenção diferenciada. Isso acabou gerando grandes obstáculos para o prosseguimento dos estudos delas, refletindo-se na queda do fluxo escolar. Para reforçar esse argumento, os autores afirmam que, nos anos 1920, 60% da população eram de analfabetos. A partir do final de 1970 e durante os anos 1980, a qualidade começou a ser entendida como permanência com sucesso na escola. Como a taxa de repetência no final dos anos 1980 era de 48%, uma série de iniciativas começou a ser desenvolvida: implantação dos ciclos e de programas de aceleração da aprendizagem ou mesmo campanhas contra a reprovação, sem estarem ligadas a nenhum programa de melhoria da aprendizagem. Nem todas essas ações foram implementadas com o devido cuidado e, como consequência, a melhora do fluxo não foi necessariamente acompanhada pela melhora da aprendizagem. Surge, então, o terceiro momento, a partir dos anos 1990, em que a qualidade passa a ser entendida como melhora do desempenho cognitivo dos alunos, medida pela aplicação de testes avaliativos, momento em que o Sistema Nacional de Avaliação da Educação Básica (Saeb) entra em operação.

Pela descrição oferecida, percebe-se que nos dois primeiros momentos havia ações concretas do poder público no sentido de atender ao conceito de qualidade em um contexto histórico determinado. Mesmo sendo criticáveis e de abrangência limitada, as ações foram sendo tomadas envolvendo a construção de escolas e de programas para a melhoria do fluxo, buscando formas de contribuir para a garantia do direito à educação. No terceiro momento, o centro das atenções tem sido no sentido de aumentar a quantidade de testes avaliativos, como se, a partir deles, toda a comunidade escolar se mobilizasse por si e nenhuma outra ação do poder público fosse necessária para garantir o direito à educação, a não ser utilizar os resultados obtidos para fixar metas de desempenho e expô-las ao escrutínio público. No contexto das reformas educacionais pós-burocráticas, a avaliação perde seu sentido como instrumento da prática pedagógica e apequena-se.

Como o desempenho ganha a centralidade, o próprio direito à educação, envolvendo a garantia de acesso e permanência para todos e ensino de qualidade uniforme também para todos, traz em seu bojo o potencial de emancipar o indivíduo e de nivelar as desigualdades dos que frequentam a escola e acaba também se apequenando. Nesses novos tempos, o lema é o direito à aprendizagem, expressão primeiramente utilizada pelo Unicef em duas publicações (FUNDO DAS NAÇÕES UNIDAS PARA A INFÂNCIA *et al*, 2008; MINISTÉRIO DA EDUCAÇÃO *et al*, 2007) e depois pelo próprio MEC.

O caminho trilhado pelo MEC para conceituar o direito à aprendizagem começa quando começam a ser debatidas as expectativas de aprendizagem.

A Câmara de Educação Básica (CEB), do Conselho Nacional de Educação (CNE), apresentou e foram homologadas duas Diretrizes Curriculares que, entre outras coisas, trataram das expectativas de aprendizagem. A primeira delas fixou as *Diretrizes Curriculares para o Ensino Fundamental de Nove Anos*<sup>3</sup> e, em seu artigo 49, determina que o MEC deveria encaminhar ao CNE proposta de expectativas de aprendizagem dos conhecimentos escolares, que devem ser atingidas pelos alunos em diferentes estágios do ensino fundamental. A segunda fixou as *Diretrizes Curriculares Nacionais para o Ensino Médio*<sup>4</sup> e, apesar de não fazer a mesma determinação ao MEC, feita na Diretriz anterior, em seu artigo 17, inciso VI, determina que os sistemas de ensino deverão desenvolver sistemas próprios de avaliação e utilizar os sistemas nacionais, coordenados pelo MEC, para acompanhar os resultados, tendo como referência as expectativas de aprendizagem. O interessante é que nas *Diretrizes*

---

<sup>3</sup> Resolução CNE/CEB nº 7, de 14/12/2010, baseada no Parecer CNE/SEB nº 11, de 7/7/2010.

<sup>4</sup> Resolução CNE/SEB nº 2, de 30/1/2012, baseada no Parecer CNE/SEB nº 5, de 4/5/2011.

*Curriculares Nacionais Gerais*<sup>5</sup> para a Educação Básica, nível que congrega as duas etapas mencionadas nas diretrizes anteriores, não trata dessas expectativas de aprendizagem.

Assim, apesar de as expectativas de aprendizagem estarem no corpo das resoluções, havia dúvida sobre o que seriam essas expectativas, em qual conceito se apoiariam e como fixá-las. Como trazia um conceito novo, era de se esperar que ele estivesse delineado nos pareceres que embasaram as diretrizes, mas isso não aconteceu e a dúvida permanece.

O fato é que, em dezembro de 2012, o MEC lança documento sobre os direitos de aprendizagem para o ciclo de alfabetização (MEC, 2012d). Na introdução desse documento, afirma-se que sua elaboração atende à determinação de encaminhar ao CNE a proposta de expectativas de aprendizagem, conforme as *Diretrizes Curriculares para o Ensino Fundamental de Nove Anos*. Como o documento não explicita as razões do encaminhamento dos direitos de aprendizagem em lugar das expectativas de aprendizagem, como determinam as diretrizes, depreende-se que o MEC considerou a palavra "expectativa" como tendo o mesmo significado da palavra "direito", que, na realidade, têm sentidos muito distintos, pois a primeira trata de algo desejável e a outra de algo que é exigido. O documento trata ainda a "aprendizagem" como direito humano (MEC, 2012d, p. 8) e inova ao interpretar o artigo 32, inciso I, da Lei de Diretrizes e Bases da Educação (LDB), que determina que "o ensino fundamental obrigatório [...] terá por objetivo a formação básica do cidadão, mediante o desenvolvimento da capacidade de aprender, tendo como meios básicos o pleno domínio da leitura, da escrita e do cálculo", como sendo, entre outros direitos, o "prioritário" (MEC, 2012d, p. 25). Dessa forma, algo que no espírito da Lei se relaciona com a formação básica do cidadão, em íntima relação com o direito à educação, transmuta-se para o direito à aprendizagem da leitura, da escrita e do cálculo, prioritariamente.

Ao que parece, os dirigentes do MEC procuraram interpretar à sua maneira a LDB, de forma a justificar, pelo menos em parte, o aumento dos testes, pois se, dentro da lógica utilizada, a aprendizagem da leitura, da escrita e do cálculo é um direito, os testes para medir a aprendizagem seriam os instrumentos para garantir o cumprimento desses direitos.

Na realidade, a aprendizagem não pode ser compreendida fora do direito à educação, assim como o é o ensino de qualidade. No entanto, reduzir um direito tão amplo e fundamental à construção da cidadania a algo restrito é um equívoco conceitual e político. Mesmo que o objetivo fosse usar uma expressão forte para dar visibilidade a um problema que merece ser atacado, isso não justificaria a interpretação reducionista dada à lei.

---

<sup>5</sup> Resolução CNE/CEB nº 4, de 13/7/2012, baseada no Parecer CNE/SEB nº 7, de 7/4/2010.

Para compreender a centralidade dada à expressão *direito à aprendizagem* e à utilização dos testes avaliativos para medir o alcance desse direito, é importante discutir os processos de regulação que vêm sendo impostos aos sistemas educacionais.

## **Regulação**

Maroy (2011a) refere-se à polissemia do termo *regulação* e o situa em dois momentos. No primeiro deles, durante o século XIX, o termo era associado apenas à função de equilíbrio interior de um sistema mecânico ou vivo, envolvendo o ajuste de vários movimentos ou atos e seus efeitos entre si. É adotado pela sociologia, quando discute os mecanismos de regulação e de integração, que contribuem para a coerência e a coesão da sociedade. Sob essa ótica, o destaque é dado às leis e aos princípios hierárquicos.

Em outro momento, entre os anos 1970-1980, muito bem retratado por Yannoulas e Oliveira (2013), o termo foi reconceituado para levar em conta elementos novos ligados ao papel do Estado e à condução da ação política, em razão, segundo Maroy (2011a), das crises financeiras, das críticas neoliberais, da globalização e da individualização da sociedade. Nesse processo, novas formas de intervenção são gestadas, envolvendo a descentralização, a desconcentração, a contratualização, as parcerias público-privadas e a avaliação. Assim, surgem novos atores e públicos e mobilizam-se atores do mundo empresarial ou da sociedade civil envolvidos na produção, na gestão e até mesmo na regulação das atividades coletivas.

Portanto, nesse novo contexto, não é possível mais entender a regulação apenas como regulamentação, pois,

Desde o momento em que as políticas públicas são pensadas em termos de ação pública e não como obra linear de uma vontade política da administração, mas como resultantes de conflitos, de arranjos, de compromissos, onde intervém uma diversidade de atores (público ou privados), a análise da ação pública adota a noção de regulação como um conceito que permite pensar e nomear as novas formas de intervenção e de papéis do Estado em um contexto de governança (MAROY, 2011a, p. 689).

Dessa forma, a regulação passa a ser algo mais sutil que a regulamentação, pois implica o desenvolvimento de sofisticados mecanismos, simbólicos ou efetivos, que superam os simples regramentos institucionais. Significa a criação de culturas institucionais



diferenciadas, na qual a avaliação educacional passa a ocupar papel central, mas com quase nenhum impacto nos processos de aprendizagem.

Segundo Barroso (2005), a regulação do sistema educativo é a expressão de um conjunto de fatores. Primeiro, não se pode perder de vista que a regulação é parte constitutiva desse sistema, como de qualquer outro sistema, e que, para se concretizar, não depende apenas da produção de regras, mas do ã(re) ajustamento da diversidade de açõesö (p. 733) dos atores (com suas posições, seus interesses e estratégias) a essas regras. Assim, como um sistema social complexo, a regulação do sistema educativo não é um processo único, automático e previsível, mas sim uma composição de múltiplas regulações, que envolve as professoras, as famílias, a comunicação social e outros agentes sociais, e não apenas o controle direto daqueles que se quer regular a partir da aplicação de uma norma. O autor lembra, também, que ãnã são os homens que são regulados e estruturados, mas os jogos que lhes são oferecidosö (CROZIER; FRIEDBERG, *apud* BARROSO, 2005, p. 730), concordando, nesse aspecto, com Maroy, em relação às sutilezas dos mecanismos de regulação.

Barroso destaca que é possível identificar três dimensões do processo de regulação dos sistemas sociais. A primeira seria a *regulação institucional*, em que as autoridades públicas intervêm para introduzir regras e constrangimentos no mercado ou na ação social. A segunda seria a *regulação autônoma*, por meio da qual os vários atores, normalmente localizados no escalão inferior do sistema, não somente definem as regras para o funcionamento do sistema, mas também ajustam este às suas diferentes estratégias e ações. Por fim, a terceira seria a *regulação conjunta*, que implica a interação entre as duas outras formas de regulação, com o objetivo de produzir regras comuns.

Nesse quadro, a regulação institucional na área educacional vem passando por diversos estágios. Com base em uma pesquisa sobre como evoluem os modos de regulação de cinco países da Europa,<sup>6</sup> Maroy (2011) afirma que nesses países, todos eles Estados unitários, em diferentes graus, convivem algumas tendências comuns. A primeira tendência, presente em todos os países, é a manutenção ou ampliação da autonomia das escolas. A segunda é a busca de um equilíbrio entre a centralização e a descentralização. Assim, os países com sistemas mais centralizados tendem a descentralizar as decisões para as instâncias intermediárias ou locais, e aqueles mais descentralizados, principalmente com relação aos objetivos do currículo, tendem a centralizar suas ações. A terceira é a promoção da escolha da família, conhecida também como *parent choice*, em que a família escolhe a escola que as

---

<sup>6</sup> Os países envolvidos são Inglaterra, Bélgica (comunidade francesa), França, Hungria e Portugal.

crianças e os jovens devem estudar e o Estado financia essa escolha. A quarta é diversificar a oferta escolar, permitindo uma diversidade de estabelecimentos, de acordo com os interesses das famílias, como, por exemplo, escolas com horários adaptados ou escolas especializadas em determinado domínio.

A quinta tendência é o aumento da regulação do controle do trabalho de ensino, em que se diminui a autonomia profissional das professoras, sujeitando-as a regramentos como, por exemplo, os códigos de boa prática, em que são divulgados modelos de sucesso que devem ser seguidos. A sexta e última tendência é o crescimento da avaliação externa, variando a sofisticação técnica de seus instrumentos, sua utilização como instrumento de acompanhamento das escolas e a forma como são publicizados seus resultados. De qualquer modo, por meio dos testes avaliativos, são reorientadas as práticas desenvolvidas pelas instituições educacionais, que tentam se adaptar à nova lógica, pois a não adaptação coloca em cheque sua própria sobrevivência em tempos de focalização dos investimentos em políticas sociais e de regulação estatal do desempenho institucional (YANNOULAS; SOUZA; ASSIS, 2009).

Além disso, os peritos envolvidos no desenvolvimento dos testes avaliativos, em sua coleta e análise de dados, se transformaram em instrumentos fundamentais para garantir à regulação uma base de cientificidade às suas ações (CARVALHO, 2009). Apesar de não ser um fenômeno recente, na última década, os especialistas têm ganhado relevância cada vez maior nos processos de decisão política (BARROSO, 2011).

As seis tendências apontadas são sustentadas por dois modelos de governança, o Quase-Mercado e o Estado Avaliador, ou a governança por resultados, os dois qualificados como pós-burocráticos (MAROY, 2011). Isso porque não se fundamentam na racionalidade da lei e no controle dos atos, em função das regras e dos procedimentos, duas características do modelo burocrático anterior. Em linhas gerais, o que diferencia os dois modelos pós-burocráticos é o papel do mercado, enfatizado pelo primeiro, ou da competição, destacado pelo segundo, como promotores da qualidade. Apesar da diferença, os dois reservam ao Estado um papel importante na definição dos objetivos do sistema, da mesma forma como ocorria no modelo burocrático.

No Quase-Mercado, o Estado delega às escolas ou a outras entidades a autonomia para escolher os melhores meios para atingir os objetivos. Mas, para alcançar a qualidade, instaura a competição entre as escolas, por fundos de financiamento atrelados ao número de matrículas existentes. Nessa competição, são oferecidos serviços educacionais a clientes bem informados, a família ãavisadaö, mediante os resultados de testes avaliativos sobre sua

realidade educacional. Segundo Maroy (2011), a principal justificativa para a adoção desse modelo de governança baseia-se no argumento de que o caráter burocrático do sistema é o que o torna ineficaz e, para tirá-lo desse estado de letargia, é necessário incentivar a pressão da concorrência, comandada pelos usuários, para que o sistema melhore.

Elliot (2002), analisando a situação educacional na Inglaterra, afirma que o Estado vem abandonando progressivamente o papel de provedor de serviços públicos, tornando-se ente regulador de um mercado educacional cada vez mais privatizado. Com isso, o conceito de educação como bem público, necessário ao desenvolvimento e à perpetuação de uma ordem social justa e democrática, é substituído por outro, em que as instituições educacionais passam a ser vistas como prestadoras de serviços, e os cidadãos encarados como consumidores.

Maroy (2011) cita um interessante trecho do livro de Whitty, Power e Halpin, em que os autores trazem os argumentos para a defesa do quase-mercado afirmando que as reformas como essa acarretaram benefícios para famílias provenientes de comunidades mais vulneráveis, as quais foram mal atendidas pelos arranjos burocráticos tradicionais (p. 31). Apesar de abrir espaço para a privatização das escolas, permitindo a entrada de empresários ávidos por uma nova fonte de negócios, o argumento é o de garantir mais oportunidades aos desassistidos. Esse trecho revela bem a substituição do Estado-Providência, que gasta muito e, mesmo assim, não consegue atingir os mais pobres por uma ação pública não estatal eficiente e mais barata.

O modelo de governança denominando Estado Avaliador garante grande autonomia à gestão econômica e pedagógica das escolas, mas sujeitando-as ao cumprimento de metas, expressas, em alguns casos, em contratos de gestão. A expressão é criada por Neave (1988) para designar o modelo cujo foco do controle não está nos processos pedagógicos, mas nos resultados alcançados, utilizando-se das avaliações externas para obter informações sobre o desempenho das escolas e acompanhamento do cumprimento das metas. A melhoria da qualidade, nesse contexto, reduz-se à valorização da eficácia, que se dissocia de suas finalidades (MAROY, 2011). O autor destaca como uma condição essencial para o sucesso desse modelo a necessidade de haver uma maior centralização do Estado, principalmente naqueles mais descentralizados, para que a fixação de metas e a cobrança por resultados possa ser efetiva.

Para Elliot (2002), o Estado passa a relacionar-se com as instituições, tendo por base o princípio da responsabilidade contratual e, para normatizar essa relação, são criadas instâncias para medir o atingimento de objetivos, auditando-se o desempenho de cada uma delas.

Segundo Maroy (2011), esse modelo faz uso de um conjunto de incentivos para estimular a melhoria do desempenho e para o cumprimento do contrato assinado. Esses estímulos podem ser simbólicos, por meio do reconhecimento das melhores escolas ou pela identificação das piores, ou ainda pela difusão de referências do tipo melhores práticas. Os estímulos podem ser também materiais, por meio da distribuição de prêmios tanto para as escolas como para seus profissionais ou mesmo para os alunos, ou por meio de punições que podem chegar, em alguns contextos, ao fechamento das escolas. O autor afirma, ainda, que uma das características mais marcantes nesse modelo é a desconfiança com relação à autonomia que as professoras sempre gozaram, pois essa autonomia não estaria sendo suficiente para garantir a qualidade e, por isso, é necessário enquadrá-las nos moldes das avaliações realizadas.

Oliveira (2011) chama a atenção para outro aspecto, ligado à regulação autônoma. Cada escola, ao se comparar com outras de sua rede, começa um processo de autorregulação e acaba assumindo as responsabilidades que lhes são imputadas, buscando, normalmente de forma individualizada, formas para melhorar seu desempenho e superar suas metas. Estratégias como essa, portanto, favorecem o isolamento em detrimento do trabalho cooperativo.

Como em qualquer sistema complexo, esses modelos, discutidos aqui, acabam produzindo diferentes combinações, hibridizando-se em cada contexto e criando composições do tipo mosaico, em que sobrevivem, lado a lado, situações do modelo burocrático com aquelas típicas do pós-burocrático (BARROSO, 2005).

A avaliação, sendo um mecanismo-chave nos processos de regulação pós-burocráticos que se instauram, precisa ser mais bem compreendida a partir de sua história, principalmente o movimento que paulatinamente a vai afastando de sua importante função como elemento capaz de contribuir com a aprendizagem dos alunos. É o que será discutido no tópico seguinte.

### ***A avaliação educacional e a avaliação externa da educação***

A avaliação é uma atividade muito cara para o processo educacional e muito se tem escrito e pesquisado sobre ela, destacando-se os trabalhos de Vianna (2005), Freitas (2002), Afonso (2005) e Villas Boas (2008). Durante os cursos de formação de professoras, esse é um

tema bastante discutido, não apenas nas disciplinas específicas, como também ao longo de todo o curso e nos programas de formação continuada, enfatizando-se muito seu significado e sua importância para aprimorar o aprendizado, tanto das professoras em formação como de seus futuros alunos.

Ocasionalmente, confunde-se a avaliação com os instrumentos de medida utilizados para medir a aprendizagem, uma das partes de qualquer processo avaliativo. Nomeiam-se os instrumentos de medida como prova, teste ou exame. Cada um desses termos tem um significado próprio que é importante destacar.

A palavra *prova*, apesar de ser um termo corriqueiro nas salas de aula, carrega em si uma carga semântica pesada, pois indica que aquele que se submete a esse instrumento deve provar alguma coisa a alguém, e isso não faz o menor sentido se estamos tratando de aprendizagens. A palavra *teste* indica o que verdadeiramente se quer de um instrumento de medida educacional: verificar, sob diversos ângulos, se a aprendizagem dos alunos vem ocorrendo. Por ser um teste, ele não prova nada! Apenas mostra uma situação que deve ser investigada, pesquisada. É exatamente por ter o sentido de uma pesquisa que a avaliação externa utiliza esse termo para referir-se a seu instrumento de medida.

A palavra *exame*, em contextos educacionais, remete a uma verificação, com o objetivo de subsidiar uma decisão em relação ao prosseguimento para um novo ciclo ou nível. Assim, por exemplo, o exame de final de ano habilitaria, ou não, determinado aluno a prosseguir para a etapa seguinte de seus estudos; o exame vestibular forneceria uma habilitação para o ingresso em outro nível de ensino, dependendo da classificação do participante, da quantidade de vagas disponíveis na instituição de educação superior; os exames certificadores da educação de jovens e adultos habilitariam seus participantes, dependendo de desempenhos mínimos estipulados previamente, a obter o certificado de conclusão de uma etapa da educação básica. O conceito de exame também é utilizado em ambientes profissionais, em situações relacionadas a habilidades que são desenvolvidas em ambientes educacionais, como é o caso dos exames para a obtenção de certificados para o exercício profissional, como os exames da Ordem dos Advogados do Brasil (OAB). Nesse caso, como nos outros, o resultado habilita seu participante a algo.

Independentemente dos sentidos das palavras tratadas aqui, todas se referem a instrumentos de medida de aprendizagens. Depois de sua aplicação, apuração dos resultados e divulgação, encerra-se um ciclo, e o papel desse instrumento se encerra.

Por sua vez, avaliar é um processo mais complexo que medir, e está ligado a um posicionamento, a uma tomada de decisões (LUCKESI, 2009). Medir uma determinada

realidade não significa que a estamos avaliando. A medida é a etapa inicial que enceta uma avaliação. Assim, se o objetivo é conhecer determinada realidade, ela deve ser identificada a partir da realização de um conjunto de medidas que trazem informações sobre essa realidade que se procura conhecer. As medidas sozinhas, obtidas a partir de observação empírica, têm pouco significado. Para que tenham significado, é preciso que elas sejam comparadas a outras medidas, ou a uma situação ideal, um padrão definido por um dado conjunto de atores, com base em critérios e com o objetivo de atingir determinados fins. A partir dessa comparação, é possível fazer um julgamento, emitir um juízo de valor, sobre o que foi medido. Mas mesmo esse juízo de valor ainda se constitui uma das etapas do processo avaliativo, não o encerrando. Apenas julgar que determinada realidade não é a adequada não fará com que ela, por si só, se modifique. Assim, é preciso identificar as causas dessa não adequação e atuar para superá-las, encerrando-se aí o ciclo da avaliação.

Durante o processo, que deve ser ético, pois lida com valores e emoções, geram-se aprendizagens, e aquele que está avaliando também se avalia, ainda mais se o objeto da avaliação tem relação com o trabalho do avaliador (VILLAS BOAS, 2008).

Para melhor compreender as discussões que se farão a seguir, é importante destacar que duas palavras inglesas, *evaluation* e *assessment*, são tratadas pelos dicionários como sinônimas e traduzidas para o português também como sinônimos de avaliação. No entanto, *assessment* refere-se ao ato de testar, apurar os resultados e divulgá-los, e neste texto será empregada, na falta de uma palavra mais adequada, como *teste avaliativo*. Por outro lado, *evaluation* tem o sentido de formar juízos de valor e atuar sobre o processo educacional, como tratado por Vianna e Franco (1991) e neste trabalho será tratada como *avaliação*.

Quando se tratam de avaliações externas, Ravela (2005) diz que a primeira tarefa a ser executada é escolher quais modelos conceituais serão utilizados e quais perguntas se quer responder. O autor alerta que é preciso ter cuidado para que instrumentos desenhados para determinado objetivo não sejam usados para atender a outros, não previstos inicialmente, pois isso poderia criar problemas que mascarariam os dados obtidos, enviesando as análises posteriores. Outra face do problema é citada por Vianna (2001), que pode ser considerado o grande mestre das avaliações em larga escala no Brasil. Vianna destaca a tendência nefasta de se desejar que em um único processo seja testado o máximo de atributos com um número limitado de instrumentos, trazendo com isso prejuízos para a validade das análises que serão realizadas. Ou então utilizar os resultados da avaliação para propósitos diferentes daqueles para os quais os instrumentos foram desenvolvidos, por exemplo, para o fechamento de escolas, algo que está ocorrendo atualmente nos EUA.

Nessa mesma linha de preocupação com os usos que se fazem dos resultados, Ristoff (1995) afirma que avaliar é uma forma de restabelecer compromissos com a sociedade; de repensar objetivos, modos de atuação e resultados; de estudar, propor e implementar mudanças nas instituições e em seus programas; deve-se avaliar para poder planejar, para evoluir (p. 63).

Expostas essas questões, defendemos, baseados em Ravela (2005) e Vianna (2005), que a construção de uma avaliação educacional externa envolveria diferentes etapas:

- Identificar quais perguntas se quer responder: essa etapa é de fundamental importância e irá orientar todo o processo avaliativo, pois se, por exemplo, o objetivo é buscar respostas quanto ao desempenho dos alunos durante determinada fase da educação, todas as etapas seguintes do processo de desenvolvimento da avaliação devem estar sendo orientadas nessa direção e não para outra que não tenha sido prevista inicialmente. Essa etapa também definirá a pertinência de todos os passos do processo avaliativo e a validade das generalizações que venham a ser feitas a partir dos resultados da pesquisa.
- Escolha do marco teórico que irá iluminar o processo avaliativo: nesse ponto é necessário ter clareza sobre a teoria de aprendizagem que guiará o processo, pois ela é a base para a construção dos instrumentos de medida. Além disso, é importante não desconsiderar o sistema educacional do qual os alunos fazem parte, pois as medidas realizadas não acontecem em situações abstratas, mas sim em situações reais.
- Estabelecimento de um padrão de comparação: a comparação será *interna* entre os grupos participantes do processo (avaliação pela norma) ou *criteria*, envolvendo o estabelecimento de critérios? No segundo caso, quais seriam esses critérios e como seriam selecionados? No caso brasileiro, para uma avaliação nacional, a situação é crítica por não existir um currículo nacional que possa orientar as discussões. Como alternativas, os testes desenvolvidos até aqui se baseiam nos currículos estaduais e dos municípios das grandes capitais e nos Parâmetros Curriculares Nacionais. Além dessas duas alternativas, utilizam-se também os conteúdos dos livros didáticos, que são adquiridos centralizadamente pelo MEC e distribuídos para as escolas, de acordo com as opções feitas por elas.
- Desenvolvimento de instrumentos de medida: devem-se levar em consideração todas as preocupações anteriores, incluindo as questões quanto ao tamanho do teste, ao tempo para respondê-lo e sua adequação ao público que participará do processo.

Deve-se estabelecer, também, o formato dos itens que comporão o teste e se eles serão de múltipla escolha ou com questões abertas, ou ainda um misto dos dois, opções que determinarão os custos de sua realização. Além disso, como o uso de estatísticas sofisticadas é usual nas avaliações em larga escala, no momento de aferir os resultados das medidas, é necessário ter em mente que o cálculo estatístico deve ser mediado por discussões pedagógicas, para que esses cálculos reflitam os marcos teóricos escolhidos.

- **Julgamento de valor:** essa etapa também é bastante crítica, pois ela orientará as mensagens que divulgarão os resultados. Assim, julgamentos como bom ou ruim, adequado ou não, com qualidade ou não, e as gradações entre eles, devem se ater aos referenciais teóricos utilizados e ao padrão de comparação adotado. Utilizar critérios diferentes desses implicará julgamentos espúrios. Como cada palavra carrega em si diferentes cargas semânticas, a escolha dos termos mais apropriados e que reflitam as opções feitas não é tarefa fácil.
- **Identificação de causas:** um processo avaliativo que não busca formas de identificar as causas dos resultados obtidos não faz jus a esse nome. Normalmente, utilizam-se questionários, respondidos pelos participantes do processo avaliativo (alunos, professoras, equipe gestora da escola), para a obtenção de informações que possam fornecer pistas para compreender as possíveis razões dos desempenhos observados. Além disso, o processo avaliativo pode recomendar a realização de pesquisas qualitativas e quantitativas, locais ou nacionais, para investigar alguns temas considerados críticos.
- **Ações de aprimoramento:** essa etapa é a que traz consequências para todo o processo, pois atua sobre os problemas identificados a partir do desenho adequado de intervenções públicas para superá-los. Nesta etapa, deve-se estar atento para evitar intervenções em sentido contrário nos marcos teóricos escolhidos e para que as causas identificadas não sejam utilizadas como se fossem consequências daquilo que se está avaliando. Isso é particularmente importante porque tem crescido muito a utilização dos resultados dos testes, desenhados para aferir o desempenho dos sistemas, como o único instrumento para julgar o trabalho da professora ou a qualidade da escola.

Pela natureza das etapas listadas, a avaliação deve envolver profissionais das mais diferentes formações e especializações: pedagogos, psicólogos, psicometristas, estatísticos,



especialistas em gestão e nas áreas de conhecimento que serão usadas para a elaboração dos itens dos testes. Sem uma equipe multidisciplinar, com conhecimentos profundos e sólidos na área de atuação e sobre o processo avaliativo, os resultados do trabalho podem levar a equívocos gravíssimos. Montar uma equipe como essa requer longo tempo de preparação, que, aliado aos custos com a aplicação dos instrumentos de medida, transformam esse processo em um investimento elevado.

### ***A emergência da avaliação externa***

O amadurecimento da avaliação educacional passou por diversos percursos até ganhar o destaque que possui nas políticas educacionais atualmente em curso em diversos países. Períodos marcantes foram as estratégias de Tyler para aproximar as medidas do aprimoramento dos currículos escolares, seguido do período da expansão da avaliação externa como instrumento para induzir políticas públicas, até se chegar à época do *accountability*.

### **A trajetória da avaliação: da época da criação dos sistemas educacionais nacionais até a época da profissionalização<sup>7</sup>**

Uma boa estratégia para compreender melhor determinado objeto é analisar como ele vem se comportando ao longo do tempo. Isso é ainda mais importante para aquele objeto que, com o passar dos anos, tenha sofrido importantes modificações nas suas concepções, funções e usos. Esse é exatamente o caso da avaliação educacional.

Escorza (2003) afirma que é possível classificar a história da avaliação, a partir do século 19, em seis diferentes épocas, cada uma com suas características: 1) época da criação de sistemas educacionais nacionais (1800-1900); 2) época da eficiência e do *testing* (1900-1930); 3) época de Tyler (1930-1945); 4) época da inocência (1945-1956); 5) época da expansão (1957-1972); e 6) época da profissionalização (a partir de 1973). Com o objetivo de compreender melhor como a avaliação educacional vem se desenvolvendo, cada uma dessas etapas será discutida a seguir.

---

<sup>7</sup> Este texto se baseou parcialmente em Horta Neto (2012) e Horta Neto e Yannoulas (2011).

Na primeira época analisada por Escorza (2003), aquela marcada pela criação dos sistemas nacionais de educação (1800-1990), fez-se necessário desenvolver alguns instrumentos capazes de obter informações que permitissem tomar decisões sobre o progresso dos alunos, pois os instrumentos existentes se baseavam em critérios subjetivos, permitindo julgamentos muitas vezes díspares. Segundo o autor, duas iniciativas americanas inovaram na época. Uma em 1845, quando algumas escolas começaram a utilizar testes escritos, e outra em 1897, quando se realizou um estudo comparado sobre o papel do ensino da ortografia, utilizando-se como dados de análise os resultados obtidos em testes aplicados aos alunos dessas escolas.

A época seguinte, a da eficiência (1900-1930), é marcada pela influência do método positivista e pelo avanço dos métodos estatísticos. É uma época marcada também por grande interesse pelas pesquisas relacionadas a medidas das condutas humanas, algumas delas buscando compreender melhor como ocorria a aprendizagem. Como consequência, desenvolve-se uma intensa atividade conhecida como *testing*, ou seja, o desenvolvimento de instrumentos estandardizados para a medição do comportamento humano. Com isso, julgar comportamentos e medi-los tornam-se ações indissociáveis, tratadas inclusive como sinônimos (ESCORZA, 2003). Buscava-se detectar as diferenças individuais em relação à aprendizagem, com ênfase nos modelos de inteligência, levando à classificação e à ordenação dos indivíduos. Como toda a atenção estava voltada para as características inatas dos sujeitos, pouca atenção se dava à relação entre os atributos medidos e o processo educacional.

Durante os anos 1920, as forças armadas americanas, com o apoio de psicólogos, desenvolveram testes para medir a inteligência dos recrutas, como forma de aprimorar o processo de seleção (RAVITCH, 2010). Para evitar a subjetividade dos testes com itens abertos, os psicólogos desenvolveram os novos testes estandardizados com itens de múltipla escolha, na época a fronteira avançada da eficiência científica. Em busca dessa eficiência, gradativamente, as escolas também passaram a utilizá-los.

Ainda nos anos 1920, desenvolve-se na França a docimologia, que propunha a elaboração de taxonomias para formular objetivos, a diversificação de instrumentos de medida, a unificação de critérios de correção e a introdução da dupla correção ou média de diferentes corretores.

A época seguinte (1930-1945), a de Tyler, reconhecido como o pai da avaliação educativa, marca o uso da avaliação a serviço da melhoria da qualidade da educação. Entre 1932 e 1940, foi desenvolvida a importante pesquisa conhecida como *Eight Years Study*, desenvolvida no Estado americano de Ohio. Partindo de currículos e de objetivos

educacionais previamente estabelecidos, buscava-se saber o quanto deles seria atendido. Tyler foi convidado a participar dela em 1934, como seu diretor de avaliação, com o objetivo de introduzir na pesquisa os instrumentos avaliativos que já havia desenvolvido no Bureau of Educational Research, da Ohio State University, onde atuava.

Essa pesquisa foi realizada em um período marcado pela grande depressão mundial, que, entre outras consequências, reduziu drasticamente as oportunidades de emprego, ampliando-se significativamente a matrícula na *high school* americana. Segundo Nowakowski (1981), que entrevistou Ralph Tyler sobre sua produção intelectual, o ensino médio não fazia parte dos planos dos jovens daquela época e passa a ser a única alternativa de ocupação deles, pois o desemprego atingia 100% da juventude. Antes da depressão, a taxa líquida de matrícula na *high school* era de 10% e nesse período atingiu 50%. Como esses jovens poderiam ter acumulado déficits educacionais importantes, havia uma preocupação em conhecer como esse déficit era tratado pela escola.

O objetivo da pesquisa, de caráter longitudinal, que envolveu 30 escolas e que teve a duração de oito anos, era verificar como as escolas estavam lidando com o grande contingente de pessoas que passam a frequentá-la e, ao final do período escolar, acompanhar os egressos por mais cinco anos, para verificar o impacto da escolarização em sua vida futura. Embora o alvo fosse determinar a mudança ocorrida nos alunos, também havia a função de explicitar essa mudança aos próprios alunos, aos pais e às professoras. Ainda mais, seria um meio para informar sobre a eficácia de um programa educacional e para a professora seria um instrumento de formação continuada. Portanto, a pesquisa extrapolava a simples medida.

Segundo Escorza (2003), para Tyler, quatro questões delimitariam o currículo: 1) Que objetivos se deseja atingir? 2) Com que atividades se pode alcançá-los? 3) Como se pode organizar essas experiências? 4) Como se pode comprovar que os objetivos foram alcançados? Tyler também apresentou cinco condições necessárias para uma boa avaliação: 1) proposta de objetivos claros; 2) determinação das situações nas quais devem se manifestar as condutas esperadas; e 3) escolha de instrumentos de avaliação apropriados; 4) interpretação dos resultados das provas; e 5) determinação da confiabilidade e da objetividade das medidas. Como nesse processo se inclui a interpretação dos resultados, o processo deixa de ser uma simples medição, pois indica a necessidade de um juízo de valor sobre os dados obtidos e a tomada de decisões acerca dos acertos ou dos fracassos do programa de ensino, em função dos resultados dos alunos. À época, esse ciclo composto de desenho dos instrumentos, realização de medidas, análise de resultados, julgamento de valor e ação sobre o processo era algo novo. Tyler chamou-o de *evaluation* para diferenciá-lo de *assessment*.

Para Tyler, o processo de avaliação é essencialmente o processo de determinar em que medida os objetivos educacionais estão se concretizando a partir do currículo e do plano de ensino (TYLER, 1969, p. 105) e em que grau as mudanças nas condutas dos alunos ocorrem. Portanto, para o autor, o foco da análise são o currículo e a sala de aula.

Escorza (2003) destaca, ainda, que Tyler, no final da vida, nos anos 1990, aprimorou suas ideias iniciais, propondo que, antes de se iniciar um processo avaliativo, seus propósitos devam ser cuidadosamente analisados, não perdendo de vista que o mais importante é ajudar os alunos a aprenderem, levando em consideração os aspectos que podem interferir no processo. Além disso, a verdadeira avaliação deveria levar em consideração as diferenças e as peculiaridades dos alunos e das escolas, devendo as professoras prestar contas de seu trabalho às famílias dos alunos, interagindo mais frequentemente com elas. Essas mesmas concepções estão presentes nos trabalhos de Vianna (2005, 2003a, 2001). Como veremos mais adiante, esses princípios da avaliação vão gradativamente ficando cada vez mais distantes, conforme vão se apresentando as diferentes demandas das épocas seguintes.

A próxima época, a da inocência (1945-1956), segundo Escorza (2003), foi aquela que se seguiu à Segunda Guerra Mundial e foi marcada pela expansão de bens e serviços, depois de um longo período de escassez. Com relação à educação, em todos os graus de ensino, houve um expressivo e acelerado aumento das matrículas escolares. Segundo Prost (*apud* Nogueira, 1990, p. 52), jamais se tinha visto um movimento tão geral, de tal amplitude e de ritmo tão rápido. Foi uma mutação brusca e global, e a expressão explosão escolar se impõe para designá-la. Esse fenômeno, em grande medida, respondia à necessidade de formação de mão de obra e foi estimulado pelos aumentos expressivos dos gastos com educação.

A época da expansão (1957-1972), que se seguiu, coincide com o auge do período da guerra fria e teve como marco o lançamento do satélite Sputnik pela antiga União Soviética. Esse fato abriu grande discussão na sociedade americana sobre o que foi considerado como o fracasso de seu sistema educacional, trazendo em seu bojo o desencanto com a escola pública e uma pressão sobre as escolas pela prestação de contas. Além disso, passa a ser parte das preocupações identificar o potencial de talentos disponível com que contava o país, de bem gerenciá-lo e de não desperdiçá-lo, enfrentando assim a concorrência internacional (NOGUEIRA, 1990, p. 52).

Influenciado por esse cenário, o governo federal americano edita a lei *Elementary and Secondary Act*, prevendo a reestruturação da educação básica com a utilização de fundos federais. Para ter acesso a esses fundos no ano seguinte, os Estados deveriam realizar avaliações, com ênfase nos resultados obtidos, dos programas subvencionados no ano anterior

(ESCORZA, 2003). Percebe-se aqui a prática comum de impor mudanças a partir de fundos federais e de exigir a evolução de alguns indicadores para continuar a acessá-los, prática essa que se torna muito comum em diversos países.

Com a pressão sobre a escola americana por melhores resultados educacionais, já existia nos anos 1950 um mercado nos EUA para os testes avaliativos desenvolvidos por algumas empresas e aplicados com regularidade pelos distritos escolares. Koretz (2008) destaca que esses testes eram elaborados para ajudar professoras e administradores a identificar as forças e as fraquezas no interior das escolas e dos distritos escolares e que eram centrados no ensino, com o objetivo de aprimorá-lo. O autor destaca, ainda, que nessa época a maior parte dos testes era referenciada à norma, ou seja, indicava o desempenho dos alunos, quando comparados com outros grupos de alunos, sem que houvesse uma referência definida *a priori*. Esse é um dado importante, pois paulatinamente foram sendo introduzidas referências e, a partir delas, metas por meio das quais as escolas passaram a ser julgadas.

Para a história da avaliação educacional, um dos marcos foi o estudo *Equality of Educational Opportunity*, conhecido como Relatório Coleman, encomendado pelo Congresso norte-americano. O objetivo dessa pesquisa, coordenada por James Samuel Coleman, era identificar como a lei dos direitos civis, aprovada em 1964, estava repercutindo nas escolas americanas e se ãas crianças de diferentes raças, cores, religiões e origem nacional estavam tendo oportunidades educacionais iguais (COLEMAN, 1966, p. iii). O estudo foi direcionado a uma amostra nacional de escolas, e se compunha de testes e questionários aplicados aos estudantes dos primeiros, terceiros, sextos, nonos e décimo segundos anos de escolaridade e por questionários aplicados às professoras e aos diretores de escola. Os testes, aplicados a 900 mil estudantes, mediram o desempenho deles nas áreas de habilidades verbais, associações não verbais, compreensão leitora, Matemática e Conhecimentos Gerais. Os questionários investigaram, entre outros temas, o *background* socioeconômico e as atitudes dos participantes com relação às raças. O formato desse estudo influenciou os testes avaliativos atuais, como, por exemplo, o Saeb e o Pisa.

Esse relatório, parte dele traduzido por Brooke e Soares (2008), teve um forte impacto na época. O desempenho médio dos mais ricos foi superior ao dos mais pobres em todos os anos escolares e áreas do conhecimento testadas. Em busca das causas dessa diferença, concluiu-se que a escola pouco contribuía para superar as dificuldades de aprendizagem com as quais os mais pobres chegavam à escola, mesmo depois de vários anos de escolaridade. Como a origem social e econômica dos alunos era o fator mais importante para o desenvolvimento de habilidades, as escolas não eram capazes de superar qualquer

combinação de fatores não escolares que influenciavam esse desenvolvimento. Tal descoberta contrariava as ideias liberais americanas da instituição educacional como promotora da igualdade de oportunidades entre os indivíduos. Segundo Broke e Soares (2008), a partir dessa conclusão, generalizou-se no campo educacional a crença de que a escola não faz diferença (p. 16). Como o estudo era denso e longo, algumas conclusões não receberam o mesmo destaque da conclusão anterior. Uma delas argumentava que a escola de boa qualidade, com destaque para aquela com boas professoras, traz mais benefícios, em termos de aumento de desempenho, para os alunos mais pobres do que para os mais ricos. Dessa forma, é importante continuar investindo nas escolas, pois os mais pobres é que serão os mais privilegiados. Argumentava também que a escola que não segregava trazia mais benefícios para os mais pobres, reforçando a luta pelos direitos civis.

O National Assessment of Educational Progress (Naep), criado em 1964, aplica seu primeiro teste em 1969, sob o impacto dos resultados do Relatório Coleman, dando grande destaque, nos relatórios que publica, às diferenças de desempenho entre brancos, negros e outras minorias, como os orientais e os latinos.

Outro *survey*, desta vez sob encomenda do governo da Inglaterra e conhecido como *Relatório Plowden*, foi realizado em 1967. Seu objetivo era pesquisar as escolas dos anos iniciais da educação obrigatória e seu impacto no desenvolvimento cognitivo dos alunos. Foram entrevistados pais, professoras, diretores e correlacionaram-se os resultados das entrevistas com os do desempenho dos estudantes, medido a partir de testes de Leitura e Matemática, envolvendo 173 escolas. O relatório resultante, também traduzido em parte por Brooke e Soares (2008), apontou três fatores como os mais importantes para garantir o sucesso dos alunos: 1) o interesse da família pelo desempenho dos filhos, o mais importante dos três; 2) as condições dos domicílios, portanto relacionadas à renda familiar; e 3) as condições das escolas. Novamente, fatores externos à escola aparecem como importantes para o desempenho dos alunos.

Outro momento histórico importante foi a fundação na Alemanha, em 1958, da International Association for the Evaluation of Educational Achievement (IEA), organização independente financiada por 70 países-membros, entre eles o Brasil. Seu primeiro estudo foi realizado em 1960 e ficou conhecido como *Pilot Twelve-Country Study*. Seu objetivo era investigar a possibilidade de condução dos testes avaliativos internacionais em larga escala. O estudo envolveu alunos de 13 anos de idade em 12 países, e as áreas testadas foram: Matemática, compreensão leitora, Geografia, Ciências e competências não verbais. Quatro anos mais tarde, em 1964, realizou-se o *First International Mathematics Study*, que envolveu

12 países e dois públicos: alunos de 13 anos de idade e alunos que estavam concluindo o ensino secundário (equivalente aos anos finais do nosso ensino fundamental). Os dois estudos concluíram que, apesar das diferenças culturais, era possível aplicar testes e comparar seus resultados. Esses estudos serviram, tempos depois, de base para o desenvolvimento do Pisa.

Nos estudos conduzidos pela IEA, assim como nos estudos conduzidos por Coleman e Plowden, não se observa qualquer objetivo classificatório. Buscavam-se dados sobre aquilo que estaria impactando os resultados educacionais, tentando encontrar possíveis soluções para superar os problemas encontrados.

Ainda nessa época, muitos trabalhos e autores passam a discutir as questões conceituais e metodológicas das avaliações, que por sua importância passam a ser parte integrante delas. Entre os autores, Cronbach (2002) defende que a avaliação deveria estar ligada à tomada de decisões; Scriven (1967) diferencia o que seria a avaliação formativa da avaliação somativa e defende que o processo deve avaliar tanto os objetivos traçados como os alcançados; Suchman (1976) defende a ideia de que avaliadores externos sejam utilizados para que não haja qualquer tipo de influência durante as análises dos resultados.

Em outro campo de estudos, também nessa época, o economista Theodore Schultz, expõe sua teoria do capital humano:

Eu proponho tratar a educação como um investimento no homem e tratar suas consequências como uma forma de capital. Já que a educação torna-se uma parte da pessoa que a recebe, me referirei a ela como capital humano. Já que ela se torna uma parte integral da pessoa, ela não pode ser comprada, vendida ou tratada como propriedade sujeita às regras de nossas instituições. [...] A principal hipótese subjacente a esse tratamento da educação é que alguns aumentos importantes na renda nacional são consequências da adição ao estoque dessa forma de capital (SCHULTZ, 1960, p. 571, tradução nossa).<sup>8</sup>

A citação mostra pela primeira vez a relação entre a educação e o desenvolvimento econômico, via aumento da renda nacional. Como se o mercado fosse totalmente elástico, ela afirma que o aumento da escolaridade geraria rendas maiores.

Com o sucesso dessa teoria, o estudo dos efeitos da educação sobre o desenvolvimento econômico se torna um tema de interesse para os economistas, que passaram a pesquisá-lo cada vez mais. Como consequência natural da relação entre educação e desenvolvimento,

---

<sup>8</sup> I propose to treat education as an investment in man and to treat its consequences as a form of capital. Since education becomes a part of the person receiving it, I shall refer to it as a *human capital*. Since it becomes an integral part of the person, it cannot be bought or sold or treated as property under our institutions. [...] The principal hypothesis underlying this treatment of education is that some important increases in national income are a consequence of additions to the stock of this form of capital.

emerge a necessidade de se definir uma qualidade para a educação, ou seja, uma referência do que seria desejável em termos educacionais, para se obter melhores resultados econômicos.

A partir de 1973, que Escorza (2003) definiu como a época da profissionalização, a investigação avaliativa se consolida e se expande. Ao mesmo tempo que utiliza instrumentos de medida cada vez mais sofisticados, mais esses instrumentos se afastam dos indivíduos avaliados, ganhando destaque aqueles desenhados para avaliar os impactos dos recursos dispendidos nos programas educativos. Apesar disso, alguns autores trazem contribuições importantes, no sentido de destacar a importância do contexto daquilo que se vai avaliar, reforçando seu caráter ético, entre eles, Stufflebeam (1994).

A discussão realizada até aqui mostrou como surgiu a necessidade de instrumentos menos subjetivos para medir o desempenho dos estudantes, primeiro, por conta da criação dos sistemas nacionais de educação e, depois, com mais ênfase, durante o período de industrialização acelerada, pela necessidade de selecionar os melhores profissionais. É nessa fase que surgem os itens de múltipla escolha, considerados na época um eficiente instrumento de medida. A concepção desse instrumento possibilitou o desenvolvimento de grandes pesquisas que procuraram medir o desempenho dos alunos, no início, para avaliar as mudanças introduzidas nos currículos e, posteriormente, para avaliar a efetividade das políticas sociais. O próprio conceito de avaliação, como algo que extrapola a simples medida, surge dessas pesquisas. Até então, a discussão se referenciava nos processos educacionais. O avanço da influência da teoria do capital humano e o destaque dado à relação entre educação e desenvolvimento econômico impõem uma ênfase no resultado educacional, fazendo com que as discussões sobre os processos educacionais fossem deixados em segundo plano. A próxima fase, marcada pelo *accountability*, enfatiza mais ainda a importância dos resultados em detrimento dos processos.

### **A fase da prestação de contas e da responsabilização (*accountability*)**

Inicialmente, é preciso esclarecer o significado do termo *accountability*, que possui dois sentidos: um ligado à prestação de contas e outro à ideia de responsabilização. Portanto, de forma a preservar os dois sentidos, a palavra significa õuma cobrança por bons resultados e a demanda de que cada um dos atores envolvidos assuma a sua responsabilidade na produção desses resultadosõ (BROOKE, 2011, p. 21). Para poder garantir esses dois sentidos, esta pesquisa adotará, por simplificação, a palavra *accountability*.



Nos anos 1980, começam a ser implementadas em larga escala as políticas neoliberais, em que o Estado deve ser mínimo e manter somente serviços considerados essenciais. Além disso, os agentes públicos passam a ser fortemente responsabilizados pelos serviços prestados à população. Nesse contexto, os testes avaliativos ganham uma centralidade sem precedentes nas políticas educacionais. A preocupação das políticas definitivamente deixa de ser com o aprendizado e passa a estar fortemente focada no desempenho dos alunos nos testes. Parte-se do princípio de que Matemática e Leitura são áreas essenciais para a vida em sociedade, e que um bom resultado dos alunos nessas áreas é um excelente preditivo para a formação de cidadãos, para a garantia de seu desenvolvimento pessoal e para o crescimento econômico dos países.

Vai ser percebido ao longo deste texto que a *accountability* acaba sendo a justificativa para uma série de políticas que começaram a ser implementadas. Apesar de o discurso oficial enfatizar a busca pela melhoria da educação, cada vez mais, as ações desenvolvidas se afastam do processo educacional, levando Dorn (2007) a cunhar, apropriadamente, a expressão *Accountability Frankenstein*, como o título de seu livro, em que discute as origens dessa política.<sup>9</sup> Os instrumentos, inicialmente desenvolvidos pelos pesquisadores como alternativa para compreender melhor a aprendizagem dos alunos e as necessárias adequações do currículo para facilitá-la, passam a ser utilizados para julgar a qualidade do ensino oferecido e justificar uma aplicação melhor dos recursos financeiros em educação.

Para que essa mudança pudesse realmente ser concretizada, algumas necessidades técnicas envolvidas na análise dos resultados dos testes deveriam ser superadas. Uma delas era a questão da comparabilidade dos resultados dos testes entre si, mesmo quando aplicados em populações e períodos de tempo diferentes. Como essa comparabilidade não existia, havia dificuldade em confrontar desempenhos e, por meio deles, as escolas. O desenvolvimento de novas técnicas estatísticas, em especial a Teoria da Resposta ao Item,<sup>10</sup> conhecida como TRI, tornou isso possível, pois utilizava, como base de comparação, o comportamento dos itens. Essa teoria, surgida nos anos 1930, se desenvolveu principalmente com o avanço dos computadores nos anos 1980, possibilitando, assim, a criação de sofisticados *softwares* que permitiriam a realização dos complexos cálculos envolvidos na teoria.

Antes da expansão do uso da TRI, os resultados dos testes eram expressos pelo número de itens certos ou errados, independentemente de o item ser fácil ou difícil. A partir do uso corriqueiro da teoria, os resultados, mostrados em uma escala de proficiência,

---

<sup>9</sup> Outros títulos interessantes que tratam dos testes e da *accountability* são: *Standardized Minds*, de Sacks (2000); *The Tyranny of Testing*, de Hoffmann (2003); e *So Much Reform, So Little Change*, de Payne (2010).

<sup>10</sup> Duas referências sobre o tema: Baker (2001) e Pasquali (1998).

passaram a refletir o desempenho do aluno em função do grau de dificuldade dos itens presentes no teste: quanto maior sua proficiência, maior a probabilidade de o aluno ter acertado os itens mais difíceis, além de ter acertado também os mais fáceis. Inegavelmente, a medida ficou muito mais precisa, pois passou a retratar não somente acertos e erros, e aparentemente mais justa, ao diferenciar desempenhos a partir das dificuldades dos itens. O problema é que essa maior precisão criou um problema para a compreensão dos resultados: comentar sobre o acerto de 12 itens em um teste composto por 30 itens fornece uma imagem mais simples de ser compreendida do que afirmar que o desempenho do aluno foi de 275 pontos em uma escala que vai de 0 a 500, por exemplo. Essa nova linguagem está muito distante da prática diária das professoras, dos alunos e seus familiares, e apenas alguns poucos envolvidos em medidas realizadas por meio de testes conseguem compreender seu significado. Mesmo assim, as medidas baseadas na TRI passaram a ser fundamentais nessa nova fase, pois, mais importante que a transparência, era ter em mãos um instrumento que garantisse a comparação.

Apesar da centralidade que a TRI adquire nesse processo, algumas vezes começam a questionar as premissas dela quando aplicadas aos testes avaliativos. Um exemplo disso são os trabalhos de Pham (2009), nos EUA, e de Tavares (2012), no Brasil.

Se a escola de qualidade é aquela onde os alunos obtêm um desempenho elevado nos testes avaliativos, é fundamental identificar o desempenho de cada uma. Começa assim o ranqueamento de escolas: primeiro, para identificar as escolas de qualidade e, principalmente, para transformá-las em exemplos a serem seguidos; segundo, por se acreditar que as famílias, ao tomarem conhecimento dos desempenhos relativos das escolas, irão pressionar a equipe da escola de seus filhos por mudanças, ou até mesmo matricular seus filhos em outras escolas com desempenho melhor. Além disso, espera-se, também, que a competição entre as escolas sirva de incentivo para que cada equipe escolar se mobilize em busca da excelência dos resultados. Segundo essa lógica, típica do modelo do Estado Avaliador, esses três fatores, por si sós, deveriam garantir mudança significativa na escola. Assim, o ranqueamento passa a ser uma obsessão e leva os governos a utilizarem os testes avaliativos de forma ampla e indiscriminada, inclusive com premiações às escolas, às suas equipes e aos alunos.

Para entender melhor esse processo, é importante acompanhar o movimento que se desenvolveu nos EUA, porque muitas das ideias lá desenvolvidas vêm sendo alardeadas mundo afora, no Brasil inclusive, como aquelas que podem melhorar a qualidade da educação. Para essa tarefa, uma autora importante é Diane Ravitch, historiadora que acompanhou de perto algumas das principais reformas do governo George W Bush e depois

se transformou em uma das principais ativistas contra os prejuízos que os testes avaliativos causaram e vêm causando ao sistema público de ensino daquele país.

Em seu livro (RAVITCH, 2010), a autora relata que, em 1983, durante a administração de Ronald Reagan, o governo divulgou o documento *A Nation at Risk* (Uma Nação em Risco), preparado por uma comissão de especialistas por solicitação do governo federal. No documento, alertava-se que os EUA corriam sérios riscos de perder sua liderança mundial por conta dos graves problemas educacionais que estaria causando a perda de competitividade das empresas americanas. Uma das soluções apontadas era aumentar o padrão de desempenho acadêmico dos estudantes. Por ser um estudo, o documento propunha soluções, mas os Estados eram livres para segui-las ou não. Em 1989, já durante a administração de George Bush, os governadores foram convocados para uma conferência sobre educação, sendo pactuadas seis metas ambiciosas até o ano 2000, que seriam monitoradas pelo recém-criado Painel Nacional de Metas Educacionais. Era desejo do presidente solicitar ao Congresso autorização para testar nacionalmente os alunos, mas isso só ocorreu no governo seguinte. Quando Bill Clinton assume a presidência, em 1993, defende de forma entusiasta as metas estipuladas pela conferência e acrescenta mais duas.

Preocupado com a escalada na direção dos testes de desempenho como remédio para resolver os problemas educacionais americanos, Ralph Tyler, em outra entrevista concedida desta vez a Diana Hiatt, em agosto de 1993, comenta que os testes fornecem uma evidência marginal sobre a aprendizagem e que os aplicados em escala nacional não podem ser considerados uma panaceia para resolver as preocupações com o que acontece nas escolas, pois o aprendizado das crianças é muito mais complexo que simplesmente focar no que lá acontece (HIATT, 1994). Tyler comenta também com a entrevistadora o que acha da preocupação do governo americano sobre as crianças em risco:

Ó Poucas crianças frequentavam o ensino médio na virada do século. Agora estamos preocupados com os jovens que não se formam no ensino médio e os rotulamos ãem riscoõ. Estamos muito menos em risco do que estávamos antes com relação ao nosso atendimento escolar. Os jovens, que são a nossa presente preocupação, pertencem aos mesmos grupos com que estávamos preocupados no passado ó os grupos de imigrantes que têm pouca ou nenhuma escolaridade. Devemos concentrar nossos esforços em atender às necessidades educacionais desses jovens, em vez de reunir os

resultados sem significância dos testes deles e ficar reclamando (p. 786, tradução nossa).<sup>11</sup>

Mas essa preocupação não teve eco e, em 1997, Bill Clinton pede ao Congresso autorização para aplicar testes nacionais voluntários em Leitura e Matemática, para estudantes dos 4º e 8º graus, mas o Congresso recusou a proposta, pois feria a autonomia dos Estados. Com relação às metas pactuadas, nenhuma delas foi cumprida, e o Painel silenciosamente desapareceu (RAVITCH, 2010).

No início dos anos 2001, George W. Bush, depois de um encontro com mais de 500 educadores para apresentar seus planos para melhorar a educação americana, centrado em estratégias que incrementavam os testes e o *accountability*, como fizera na época em que foi governador no Estado americano do Texas, envia ao Congresso a lei chamada *No Child Left Behind* (NCLB) [Nenhuma Criança Deixada para Trás]. Apesar do grande alarde, existiam grandes resistências no Congresso à aprovação da Lei, que representava uma mudança de direção do controle educacional, do âmbito local para o nível federal (TEMKIN; ROELLKE, 2009). Como se afirmou antes, quando da discussão sobre as condições necessárias para a governança do Estado Avaliador, é necessário que os sistemas muito descentralizados percam parte de sua autonomia para que a regulação possa ocorrer.

As resistências do Congresso americano se dissiparam depois da grande comoção com os atentados em setembro de 2001, sendo a lei promulgada em janeiro do ano seguinte. Essa lei foi um marco na nova fase da avaliação educacional, nos EUA e em outros países. Em vez de várias metas, a lei determina uma única: que as escolas fossem responsáveis por garantir que todos os alunos, entre o terceiro e oitavo graus, alcançassem um determinado nível de desempenho, conhecido como *proficiency*,<sup>12</sup> em Leitura e Matemática no final de 2014.

O grande destaque dado durante a discussão da lei pelo Congresso americano foi realçar que a Lei NCLB tinha como foco especial os grupos da sociedade que tradicionalmente haviam sido deixados para trás, os mais pobres. A lei foi aprovada com folga

---

<sup>11</sup> "Few children attended high school at the turn of the century. Now we are concerned about the youths who do not graduate from high school and label them at risk. We are less at risk in our educational attainment than we have ever been. The youths who are our present concern belong to the same groups we were concerned about in the past - the groups of incoming immigrants who have little or no schooling. We should focus our efforts on meeting the educational needs of these youths rather than on gathering meaningless test scores from them and crying woe is me."

<sup>12</sup> O Naep prevê, para cada uma das áreas testadas, três níveis de proficiência: *basic*, *proficiency*, *advanced*, indicando, para cada um, intervalos da escala de proficiência que varia de 0 a 500. Esses três níveis foram definidos, em 1990, pelo National Assessment Governing Board (NAGB), órgão criado pelo Congresso americano para acompanhar o desenvolvimento da educação daquele país (BOURQUE, 2009). Na época em que isso ocorreu, houve resistência dos meios acadêmicos, que argumentaram que essa divisão era arbitrária e que pouco informava sobre o real progresso dos alunos (BROWN, 2000). Apesar das críticas, os três níveis tornaram-se referência na discussão sobre os problemas educacionais americanos.

pelo Congresso americano tanto pelos democratas como pelos republicanos, que concordaram com a importância do *accountability* para professoras e para as escolas, principalmente aquelas em que os alunos apresentavam baixa proficiência, e para os diretores dessas escolas. Para isso, determinava que os Estados americanos deveriam encaminhar planos de responsabilização ao Departamento de Educação, órgão federal responsável pela coordenação federal das políticas educacionais, detalhando as regras e as políticas a serem usadas para acompanhar o progresso anual em busca da meta estipulada (CRONIN; DAHLIN *et al.*, 2009). Dadas as resistências que ainda existiam no Congresso para aplicar testes nacionais, foi prevista a obrigatoriedade de os Estados desenvolverem testes próprios e aplicá-los em seu território. Para acompanhar o progresso anual, cada Estado deveria definir uma meta para cada uma das escolas, conhecida como Adequate Yearly Progress (AYP), que compreendia: o desempenho em Leitura, Artes e Matemática em testes avaliativos, taxa de matrícula anual e conclusão da *high school*. Caso as metas não fossem atingidas, o Estado deveria encaminhar um pedido de perdão (*waiver*), detalhando as providências a serem tomadas para ele se adequar às metas estabelecidas, uma clara intervenção branca do governo federal.

Como atingir o grau *proficiency* era uma meta muito ambiciosa, em 2011, o governo federal, durante a administração Barack Obama, sugeriu aos Estados solicitarem flexibilização com relação às exigências do NCLB, mas com o compromisso de assumirem algumas metas estabelecidas no programa *Race to the Top* (Corrida ao Topo), discutido adiante, que aprofunda mais ainda o controle sobre as escolas e a equipe escolar, em troca de metas menos ambiciosas. As metas definidas originalmente na lei eram tão irrealistas, e fixadas sem nenhum critério científico, que dos 50 estados americanos e mais o distrito federal, até maio de 2013, 46 solicitaram a flexibilização. Começa a se repetir o que aconteceu com o Painel Nacional de Metas Educacionais durante o governo Bill Clinton, quando as metas fixadas foram esquecidas.

Para Ravitch (2010), o *accountability* tornou-se o *slogan* dos administradores públicos e dos líderes empresariais (p. 149).

A relação entre a qualidade educacional e o desenvolvimento econômico, que na época anterior era uma teoria, ganha, nos anos 2000, evidências empíricas, determinando a relação entre as habilidades cognitivas, uma das dimensões do que seria a qualidade educacional, e o crescimento econômico. Em um desses trabalhos, elaborado por um dos mais importantes economistas da educação, Eric Hanushek, da Universidade de Harvard, o fator crucial seria o desempenho acadêmico, pois melhorando os resultados da aprendizagem, juntamente com a expansão da escolarização, melhoraria a produtividade do trabalho, o que

iria se refletir no rendimento dos trabalhadores, contribuindo para taxas mais elevadas e sustentáveis de crescimento da renda nacional (HANUSHEK; PETERSON; WOESSMANN, 2007). Para que isso ocorresse, bastaria determinar quais políticas e programas poderiam melhorar os resultados da aprendizagem, tomando como base os estudos do IEA e do Pisa, pois esses estudos colocariam os desempenhos dos alunos americanos em uma perspectiva internacional.

Em 2009, durante o governo Barack Obama, na esteira da crise econômica de 2008, o Departamento de Educação do governo americano recebeu do Congresso recursos na ordem de US\$ 100 bilhões, como parte de um pacote econômico maior destinado a superar os efeitos da crise do ano anterior. Desse total, US\$ 4,3 bilhões foram destinados ao programa de reformas conhecido como *Race to the Top* (RAVITCH, 2010). Esse programa foi definido por Petrilli (2009) como um programa que coloca a cenoura na ponta de uma vara,<sup>13</sup> ou seja, com o objetivo de alcançar a cenoura, no caso os recursos federais, os Estados seguem a direção determinada pelo governo federal, implantando determinadas reformas. Essas reformas, segundo o autor, acabam impondo aos Estados restrições às suas liberdades de definir políticas e cobrando deles o cumprimento de metas, isso em um país que preza muito seu federalismo e a independência de seus entes federados, o que pode ser um indicativo de uma excessiva centralização das políticas educacionais nas mãos do governo federal e uma ameaça ao federalismo, algo que já foi discutido como condição essencial para a implantação do Estado Avaliador. O programa prevê a competição entre Estados, para que os melhores planos recebam mais recursos, com o objetivo de implementar reformas significativas que permitam: intensificar o desenvolvimento de padrões educacionais (*standards*) e testes; melhorar a coleta e a utilização de dados educacionais; aumentar a eficácia das professoras, por meio de medidas de valor agregado, que serão discutidas adiante; distribuir as melhores professoras, igualmente, entre as escolas; e transformar as escolas com extremas dificuldades (USA, 2010).

Estudos liderados pelo economista Eric Hanushek têm fornecido as evidências necessárias para a aplicação dessas políticas. Em um deles, com o sugestivo título *Are U.S. students ready to compete?* (Os Estudantes Americanos Estão Preparados Para Competir?), Hanushek e outros (2011) afirmam que encontraram apenas 6% dos estudantes norte-americanos classificados no nível avançado em Matemática, o menor percentual dentre 30 outros países. Afirmam, ainda, que apenas 32% dos alunos do 8º grau eram proficientes em Matemática, colocando os EUA na 32ª posição, entre vários países. Em um estudo de 2012,

---

<sup>13</sup> A expressão usada pelo autor é *the carrot that fells like a sitckö*.

novamente Hanushek e outros comparam o desempenho dos EUA com 48 países desenvolvidos e em desenvolvimento, durante 1995 e 2009, destacando como resultado que 24 deles tiveram ganhos de desempenho maiores que os dos EUA. Os autores também afirmam que, quando se comparam os ganhos de desempenho dos Estados americanos entre si, no período de 1992 a 2011, é possível perceber grande variação no desempenho de cada um deles, maior inclusive que a observada entre os países participantes do estudo. Portanto, os dados levantados reforçam mais ainda as ideias apresentadas em 1983, quando do lançamento do relatório *A Nation at Risk* (Uma Nação em Risco), criando, assim, o clima adequado para a propagação do *accountability* como sendo a solução para resolver os problemas da economia americana, como se os problemas educacionais fossem os únicos responsáveis por esses problemas.

Dentro da linguagem do *accountability*, existem dois tipos de implicações com relação aos resultados daqueles que estão sendo avaliados: as *low stakes*, quando não existem sanções ou então quando elas são muito pequenas, e as *high stakes*, aquelas que preveem sanções graves. As primeiras estariam relacionadas apenas à aplicação dos testes e à divulgação de seus resultados, enquanto as outras seriam as que podem trazer consequências para as escolas, suas professoras e, mais recentemente, para os gestores educacionais.

Existe um debate acalorado nos EUA que defende e critica as políticas *high stakes*. Vários estudos analisaram essas políticas. Carnoy e Loeb (2004), por exemplo, concluíram que a responsabilização por resultados resultou em melhoria de desempenhos e diminuição das diferenças deste entre os alunos. Hanushek e Raymond (2003) encontraram melhorias nos desempenhos, mas aumento nas diferenças entre os alunos. Amrein e Berliner (2002, 2003) concluíram que em determinados contextos houve quedas no desempenho, mas em nenhum foram encontrados ganhos. Como se pode ver, não existem comprovações claras de que o estabelecimento de sanções tenha impactado positivamente o desempenho dos alunos nos testes. Apesar disso, muitas políticas *high stakes* começaram e continuam sendo implantadas.

O aumento dos testes para medir o desempenho dos alunos, um mercado que já existia nos anos 1950, como mencionado, possibilitou o aumento da chamada indústria de venda de instrumentos de medidas educacionais. Segundo estimativas de Julie Broom, diretor de um importante centro de pesquisas educacionais americano, o Institute for Research and Reform in Education (IRRE), esse mercado gira em torno de US\$ 20 a US\$ 50 bilhões por ano (GRILLO, 2011). Grande parte desse mercado se desenvolveu depois da Lei NCLB, ou seja, o próprio governo federal abriu espaço para o crescimento dessa indústria, que envolve,

principalmente, as gigantes americanas do mercado editorial como McGraw-Hill, Pearson e Houghton Mifflin Harcourt.

Criou-se na sociedade americana certo consenso de que os testes e as medidas de *accountability* são a solução para a educação naquele país, apesar de estar aumentando o dissenso. Segundo Ravitch (2010), criou-se o mito de que a solução para a baixa pontuação nos testes é a aplicação de mais testes. E esse contexto se reflete no fato de que vários governos, mesmo com visões políticas diferentes em diferentes aspectos, insistem nessas ações, aprofundando-as cada vez mais, como será visto a seguir.

### **As políticas com consequências (*high-stakes*)**

Neste tópico serão abordadas as políticas *high-stakes*, que, ao se basearem em ações de *accountability*, tentam introduzir no sistema educacional as lógicas do mercado empresarial. É possível perceber ações estabelecidas sempre sob a justificativa da melhoria da qualidade da educação, em benefício das crianças e dos jovens, que tradicionalmente são deixados para trás, e ao mesmo tempo criam-se mercados educacionais bastante lucrativos. Além disso, com o aumento da pressão, surgem as estratégias para enganar o sistema, que envolvem o ensino para o teste e as fraudes. Esses tópicos serão discutidos a seguir.

#### *Medidas de valor agregado*

Como a Lei NCLB determinava que o desempenho nos testes de Matemática e Leitura subisse anualmente, e como as professoras, individualmente, são o principal fator de influência no desempenho dos alunos, surge o interesse pelo desenvolvimento de técnicas de medidas de valor agregado, para acompanhar e prever o desempenho nos testes, conectando-o com a efetividade da professora (RAVITCH, 2010). Assim, procura-se verificar no final de um período letivo o quanto cada professora agrega ao desempenho dos alunos.

O desenvolvimento de uma medida com essa característica baseou-se em estudos que buscavam saber se as professoras faziam diferença no desempenho dos alunos. Carnoy (2010) destaca que as medidas de valor agregado devem ser relativizadas, pois existem algumas questões que não podem ser deixadas de lado, por exemplo: as professoras não são designadas de forma aleatória e, habitualmente, as melhores professoras acabam ensinando os alunos que



têm mais chances de aprender; também, habitualmente, os resultados individuais das professoras não se mantêm estáveis ano a ano, o que causa problemas quando se pretende recompensá-las com base nessa medida; existem resultados divergentes com relação ao efeito das qualificações das professoras, ou seja, se elas têm influência ou não no desempenho dos alunos; por fim, a dúvida maior é se são as habilidades da professora ou os fatores escolares os mais decisivos para determinar o desempenho dos alunos.

Além dessas questões, Carnoy (2010) destaca alguns problemas nas ações que se baseiam nas medidas de valor agregado. A baixa estabilidade, ano a ano, dessa medida, principalmente entre os quintis superior e inferior de desempenho, poderá contribuir para diminuir sua utilização; no entanto, aqueles que permanecem seguidamente nessas faixas merecem alguma recompensa, no primeiro caso, ou algum tipo de ação, no segundo. Se a decisão for afastar os de pior desempenho, o autor tem dúvidas se existiriam bons substitutos para eles. Por fim, o autor levanta uma dúvida sobre qual seria o comportamento das professoras que não são recompensadas. Outros trabalhos também levantaram dúvidas sobre a estabilidade das medidas obtidas (SASS, 2008; KANE; STAIGER, 2009).

Apesar das questões apontadas, o jornal americano *Los Angeles Times* publicou, por duas vezes a primeira em 2010, baseando-se em medidas para o período 2003-2010, e a segunda em 2011, com os dados desse ano, um *ranking* das professoras do Estado da Califórnia, baseado em medidas de valor agregado e organizado por um grupo de especialistas contratados pelo jornal. Durso (2012) revisou os dados daqueles estudos e acrescentou algumas dúvidas àquelas discutidas aqui, e a mais importante referia-se à possibilidade de comparar, na mesma métrica, professoras que ensinam populações estudantis muito diferentes. Sobre a precisão da medida, o autor, tomando, por exemplo, o caso relatado pelo jornal, de uma professora que teve como medida de seu valor agregado 45 pontos, em uma escala que chega até 100 e na qual a faixa de erro é de +/-20, significando que a medida dela estaria em algum lugar entre os 25 pontos, posição das professoras menos efetivas, e 65 pontos, reservado às mais efetivas. Por fim, como principal conclusão, a autora afirma que existe evidência significativa de que a inclusão de contribuições à aprendizagem do aluno não pode ser atribuída apenas às professoras, indicando assim a fragilidade dessa medida. Ravitch (2012d) reforça a conclusão afirmando que, por conta do índice de pobreza americano de 22%, existem alunos que já entram defasados em relação a outros, por conta de má nutrição; ausência dos pais, ou porque estes possuem pelo menos dois empregos; falta de óculos ou de aparelhos auditivos; questões de higiene, abuso ou negligência; acesso precário a cuidados

médicos; mudanças frequentes e falta de experiências fora das imediações da sua comunidade. São condições muito sérias que não podem ser desconsideradas.

Ravitch (2012m), comentando a insistência do prefeito de Nova York, Michael Bloomberg, em utilizar as medidas de valor agregado para classificar as professoras das escolas do município, assim se manifestou:

Parece que o prefeito pensa que os familiares são consumidores que estão aptos a ir às compras de professoras. Se eles não gostarem da classificação de Mr. Smith, eles estão aptos a transferir suas crianças para a sala de Ms. Jones, porque ela tem uma classificação maior. [...] Diferentemente dos negócios, onde os consumidores podem decidir mudar de fornecedor, uma professora pode acomodar um número limitado de crianças. [...] O que acontece se os familiares de 200 estudantes ou todos os 500 estudantes quiserem estar em sua classe? Isso não funciona, e não faz sentido. Ainda mais, dado o que sabemos sobre as enormes margens de erro das classificações, Ms. Jones pode absolutamente não ser a melhor professora. Os consumidores podem ser mal informados.<sup>14</sup>

Portanto, segundo a autora, ações baseadas em medidas de valor agregado, que a princípio valorizariam o mérito das professoras, servem apenas como forma de aumentar a pressão sobre o corpo docente, trazendo poucos benefícios reais.

Outro aspecto maléfico dessa ação é que ela acaba destruindo o trabalho em equipe (SAHLBERG, 2011), fundamental para o trabalho docente, porque impossibilita a discussão sobre como melhorar as aprendizagens baseadas em estratégias utilizadas individualmente em sala de aula.

Mesmo com todas essas evidências, é crescente a utilização, tanto nos EUA como em outros países, de medidas de valor agregado para mensurar a efetividade das professoras.

#### *Padrões Estaduais Básicos Comuns (Common Core State Standards)*

O direito legal e a responsabilidade com relação à educação, nos EUA, recaem sobre cada um dos 50 Estados americanos (BAKER; SCIARRA; FARRIE, 2012). Portanto, definir

---

<sup>14</sup> It appears that the mayor thinks that parents are consumers who should be able to go teacher-shopping. If they don't like Mr. Smith's rating, they should be able to transfer their child into Ms. Jones's class because she has a higher rating. [...] Unlike a business, where consumers may decide to shift their patronage, a teacher can accommodate a limited number of children. [...] What happens if the parents of 200 students or all 500 students want to be in her class? It doesn't work, and it makes no sense. Furthermore, given what we already know are the huge margins of error built into the ratings, Ms. Jones may not be the best teacher at all. The consumers may be misinformed.

padrões nacionais sempre foi uma tarefa muito complexa, pois envolve discutir a autonomia dos Estados.

Em 1992, começou uma discussão nos EUA sobre a implantação de um sistema educacional baseado em padrões (*standards*) que estabelecesse as habilidades que os alunos deveriam desenvolver, facilitando, assim, a elaboração de testes nacionais baseados nas mesmas referências (ROTHMAN *et al.*, 2002). Essa iniciativa não progrediu.

Uma organização independente, criada em 1996, chamada Achieve & The American Diploma Project Network (Achieve),<sup>15</sup> que congrega líderes empresariais e governadores de Estados, começou a liderar um movimento para levar adiante a ideia dos padrões e aprofundar as reformas educacionais em curso no país. Com o apoio político que tinha, algumas de suas ideias começaram a se transformar em ações efetivas. Uma delas foi a implementação de padrões para o ensino médio (*high school standards*), que envolviam o estabelecimento de requisitos para a conclusão do curso, medidos através dos testes, condição para o recebimento do diploma. Esta sistemática está sendo utilizada por mais da metade dos Estados americanos.

A partir dessa vitória, a Achieve liderou o movimento para estabelecer padrões para cada ano escolar de toda a educação básica. Esse trabalho recebeu o nome de *Common State Core Standards* (Padrões Estaduais Básicos Comuns) e ficou pronto em 2010, para as áreas de Matemática e Língua Inglesa (envolvendo leitura de diferentes gêneros, redação, expressão e compreensão oral e gramática). Segundo a Achieve (2012), esses padrões são referenciados internacionalmente, garantindo que os estudantes estejam preparados para ter êxito nas economias e sociedades globais. Além disso, segundo a organização, são padrões adequados à carreira e ao trabalho, rigorosos, claros, focados, cuja formulação foi baseada em pesquisas.

Juntamente com o desenvolvimento dos padrões, a Achieve obteve apoio da entidade que congrega todos os governadores, a National Governors Association (NGA), e da que congrega os gestores educacionais, o Council of Chief State School Officers (CCSSO). Com esse apoio, os padrões estão em processo de implantação em todos os Estados.

Chama a atenção o fato de todo esse processo ter sido desenvolvido por meio da união de interesses entre os governadores, pressionados pela legislação da Lei NCLB, que exige dos Estados ações efetivas para que todos alcancem metas de proficiência, e os empresários, preocupados com a crise americana e o avanço de outras economias no mercado mundial, que podem minar o poder de influência dos EUA, e ter sido liderado por uma organização privada como a Achieve. Aliás, organizações como essa acabam, por meio de diversas articulações,

---

<sup>15</sup> Mais sobre essa organização pode ser obtido em seu *site*: <<http://www.achieve.org>>.

conseguindo fazer com que suas ideias sejam assumidas pelo poder público, não só nos EUA como no Brasil também.

O governo federal americano, apesar de não participar diretamente do processo de discussão e implantação dos *Common State Core Standards*, tem apoiado financeiramente os Estados na implantação deles, implantação essa que está movimentando todo o mercado editorial dos EUA.

#### *Rankings e ampliação dos testes*

Com base no desempenho das escolas, em alguns Estados americanos, elas vêm recebendo notas na forma de letras, de A a F. Segundo Ravitch (2012g), essa ideia começa no Estado da Flórida, em 1999, depois foi adotada pela cidade de Nova York e passa a ser adotada por outras localidades simplesmente pelo fato de já estar sendo usada pelas primeiras. A autora destaca que uma letra não significa nada, pois o objetivo da avaliação é dar suporte e desenvolver, não estigmatizar.

Um dos efeitos perversos dessa medida é piorar ainda mais a situação das escolas indicadas como de baixo desempenho, aquelas classificadas pelas letras entre D e F, pois as professoras estariam sujeitas a sanções, ameaçadas de serem despedidas e expostas ao ridículo e as melhores professoras, em massa, acabam se afastando delas (DARLING-HAMMOND, 2010). Para exemplificar essa situação, a autora destaca o depoimento de um diretor de escola, que pergunta se alguém vai querer dedicar sua vida a uma escola que já foi rotulada como fracassada? (2010, p. 79). Se a escola já apresenta dificuldades e seus alunos convivem com baixos desempenhos, sua situação ficará ainda pior se não puder contar com boas professoras. Estudos baseados na realidade da Inglaterra revelaram que políticas como essa acabam levando à marginalização dos estudantes de baixo desempenho e a um clima, percebido pelas professoras, menos tolerante com estudantes que têm dificuldades acadêmicas ou de comportamento, e esses efeitos foram maiores quanto pior o desempenho da escola nos testes avaliativos (RUSTIQUE-FORRESTER, 2005).

Se alguns consideram importante comparar escolas em um único país, por que não compará-las internacionalmente? De acordo com Andreas Schleider, chefe da Divisão de Análises e Indicadores da área de educação da OCDE, que tem como uma de suas atribuições o Pisa, em entrevista ao jornal *The New York Times*, brevemente isso será possível, permitindo que qualquer familiar, com o resultado da escola de suas crianças e jovens em

mãos, possa perguntar ao gestor escolar quais as razões para a escola de seu filho não ter desempenho semelhante a outra da China ou da Finlândia (FRIEDMAN, 2012). Assim como, nessa época de globalização, buscam-se melhores condições para a produção de bens, buscar-se-iam as melhores escolas para garantir os melhores desempenhos das crianças e dos jovens.

Com a fixação sobre os testes, algumas propostas de estendê-los a populações que nunca haviam sido examinadas antes começam a aparecer. Uma delas foi desenvolvida pela americana American College Testing (ACT), importante organização que desenvolve diferentes testes educacionais. Seu mais novo produto é um teste para acompanhar o desenvolvimento de carreiras a partir da educação infantil.<sup>16</sup> A ferramenta rastreia os interesses de carreira dos estudantes, seu desempenho acadêmico e o progresso em direção a objetivos, acompanhando os estudantes do jardim de infância até o ensino médio (*high school*). A esse respeito, Ravitch (2012e) informa que vários Estados e distritos estão desenvolvendo testes para os alunos da educação infantil, a fim de avaliar suas habilidades cognitivas, e que contam com o apoio do governo federal a essas iniciativas, fazendo com que cada vez mais dados sejam coletados para rastrear as pessoas desde a mais tenra idade. Ravitch (2012f) afirma que políticas como essas poderão ser vistas daqui a cem anos com os mesmos olhos que vemos hoje uma notícia publicada no jornal *The New York Times*, de 1912, que informava sobre uma nova política em curso, que iria eletrificar internamente as paredes das salas de aula para melhorar o desempenho das crianças deficientes, experiência que já havia sido testada com sucesso na Suécia.<sup>17</sup>

#### *Soluções para as escolas de baixo desempenho*

Quando os testes revelam escolas com baixo desempenho, algumas soluções estão sendo comumente usadas nos EUA. Serão analisadas três delas, dado o impacto que estão gerando: fechamento de escolas, escolas *Charter* e sistema de *Voucher*.

As escolas de baixo desempenho são denominadas de *failing schools*. Se essas escolas estão falhando em seu processo de ensinar, começa-se uma discussão sobre o que é mais efetivo: atuar para melhorá-las ou simplesmente fechá-las (SPRENG, 2005). Na cidade de Nova York, as escolas com baixo desempenho são fechadas com a justificativa de que fechar as escolas ruins significa salvar os alunos de uma situação de não aprendizagem. Como

---

<sup>16</sup> <<http://www.act.org>>

<sup>17</sup> <sup>o</sup>Electricity for defectives. Disponível em: <<http://query.nytimes.com/mem/archive-free/pdf?res=9400E5DB123CE633A25752C3A9619C946396D6CF>>. Acessado em: 30 maio 2012.

consequência, os diretores são despedidos e as professoras ficam à disposição da administração em um banco de reservas, podendo ser chamadas para atuar em escolas onde existem faltas temporárias de professoras. Assim, comenta Spreng (2005), as professoras vão se deslocando pelas diferentes escolas, como se fossem almas penadas, sem passar mais de uma semana em cada uma delas. Fechada uma escola, abrem-se outras menores, com uma nova equipe escolar, funcionando no mesmo prédio. Rubinstein (2012) comenta que essa política permitiu à administração alardear falsamente o aumento da taxa de conclusão em 7%. Esse aparente sucesso, segundo o autor, não se deve a uma mudança pedagógica ou a melhores professoras, mas sim pelo fato de a nova escola não possuir repetentes, pois são todos alunos novos: compara-se uma situação não mais existente com uma nova.

Existem outras estratégias menos radicais que o fechamento de escolas. São elas o sistema de *Vouchers* e o arrendamento de escolas públicas (*Charters Schools*).

No caso do sistema de *Vouchers*, é dada às famílias a opção de escolher a melhor escola para matricular seus filhos e, para isso, recebem recursos públicos para arcar com os custos educacionais em escolas privadas. Essas políticas estão fundamentadas no conceito da escolha da família (*Parent Choice*), algo que já foi discutido neste texto, um debate que ocorre nos EUA desde os anos 1950 e que se consolidou como uma decisão da suprema corte americana, permitindo às famílias que, na época, não queriam ter seus filhos frequentando as escolas públicas, que implantavam a integração entre brancos e negros, matriculá-los em uma escola privada, obrigando o Estado a arcar com os custos.

Poucos anos depois, Milton Friedman defende que o governo deveria sustentar a educação, mas não ser o responsável pelas escolas. Esse conceito permaneceu adormecido e só começou a ser disseminado durante os anos 1980, durante o governo Reagan (RAVITCH, 2010), principalmente para validar a política de *Vouchers*. Os *Vouchers* foram utilizados nessa mesma época em larga escala na Inglaterra e no Chile, onde ainda hoje geram discussões. Nos EUA, são defendidos arduamente pelas escolas confessionais, mas não conseguiram legislação que lhe desse suporte e massificasse sua utilização (RAVITCH, 2010).

No caso das escolas *Charters*, as escolas são arrendadas para organizações privadas que, em contrapartida, recebem financiamento do Estado com base no número de alunos matriculados, em troca da garantia de elevação do desempenho dos alunos, tudo isso definido em contrato com o poder público. Seu nascimento, nos anos 1990, é marcado por um movimento defendendo que a educação pública era incapaz de ser reformada porque a instituição havia sido capturada por interesses que envolviam os sindicatos dos professores, os

diretores de escolas, as próprias escolas, as administrações educacionais e outros beneficiários do *status quo* institucional (RAVITCH, 2010, p. 118). A solução para isso seria eliminar a interferência de toda a administração de alto nível da escola, deixando-a a cargo de instituições que prestassem contas de seu trabalho à sociedade e ao Estado.

Hoje, as escolas *Charters* competem com as escolas públicas para receber os melhores alunos e por mais recursos, algumas contando com o apoio de grandes grupos empresariais e conseguindo garantir junto à população uma imagem de eficiência e sucesso. Além dessas escolas, foram criadas também escolas *Charters* virtuais, onde a educação básica ocorre a distância.

Para se ter uma ideia de como essas escolas se alastraram, dados de 2009 informam que existiam mais de 4.700 escolas *Charters* nas quais estudavam mais de 1,4 milhão de alunos e, nas listas de espera, um sistema em que, por sorteio ou por seleção, são escolhidos os alunos que a frequentarão, existiam mais de 350 mil pessoas (CREDO, 2009). Em 2010, um ano depois, já eram 5.300 escolas, 218 delas *on-line*, com mais de 1,8 milhão de alunos, sem contabilizar as listas de espera, um aumento de quase 30%. Ravitch (2010) destaca que o sistema de sorteios por essas vagas, antes de ser um sistema justo, tende a beneficiar os melhores alunos das regiões mais pobres, deixando os alunos com maiores dificuldades de aprendizagem dessas regiões frequentando as escolas regulares, piorando ainda mais a situação deles.

Esse tema é tão atual na discussão americana que, em 2012, foi lançado um filme com o título de *Won't Back Down*,<sup>18</sup> que retrata a luta vitoriosa de duas mães, uma delas professora, contra a burocracia local e o sindicato dos professores, para transformar uma escola em *Charter*, de uma pequena cidade do interior, onde seus filhos estudam, que apresenta baixo desempenho, por ser classificada como *failing school*.

Muitas das organizações criadas para arrendar as escolas públicas são responsáveis por várias delas em diversos Estados e atuam fortemente na mídia, no sentido de alardear os benefícios que tal arranjo institucional tem para a melhoria da educação. São verdadeiras redes de ensino. Para apoiar ainda mais essa política e para dar guarida ao clamor popular que a defende, no Estado da Califórnia, em janeiro de 2010, foi aprovada, e está se discutindo em outros vinte Estados americanos (NCSL, 2012), uma legislação conhecida como *Parent Trigger*,<sup>19</sup> Por meio dela, permite-se que familiares de alunos de escolas com baixo desempenho decidam pelo voto, bastando a maioria simples, que a escola seja transformada

---

<sup>18</sup> O trailer do filme pode ser assistido em: <<http://www.youtube.com/watch?v=lvYPxZpstkE>>.

<sup>19</sup> Segundo Diane Ravitch, essa ideia surgiu de um grupo chamado *Parent Revolution*, que é financiado por fundações ligadas aos bilionários americanos Bill Gates, Eli Broad e Sam Walton.

em *Charter* e entregue a uma companhia privada, que sejam despedidas todas suas professoras e seu diretor, ou até mesmo que a escola seja fechada. O argumento utilizado para defender essa lei é que ela fornece às famílias de baixa renda mecanismos para superar os entraves da burocracia, que não se mobiliza para modificar situações de baixo desempenho educacional. Baker (2012a) lembra que políticas como essa podem ser utilizadas para a criação de escolas voltadas para determinados valores e, com isso, oprimir qualquer população minoritária, devido a razões de preferência sexual, raça, cultura ou religião.

Ravitch (2010) relata uma série de abusos cometidos pelas redes de escolas *Charters*, sendo que um deles chama a atenção. Como qualquer negócio privado, essas redes também podem enfrentar dificuldades financeiras, como a que ocorreu com uma rede no Estado da Califórnia, com 60 escolas e 6 mil alunos, que faliu, apesar de ter recebido fundos de US\$ 100 milhões.

Com relação ao desempenho dos alunos que frequentam as escolas *Charters*, alguns estudos (CREDO, 2009, 2012; BAKER, 2012b) revelam que apenas 17% delas superaram o desempenho das escolas regulares na mesma região e que, quando seus resultados são melhores, observa-se certa segregação dos alunos com baixo desempenho ou dos alunos de minoria étnica. Isso demonstra que o mito da administração privada conseguir melhores resultados que a pública é falso. Com relação às escolas virtuais, o desempenho dos alunos é ainda pior. Além da diferença no desempenho, as escolas *Charters* possuem proporcionalmente menos crianças com deficiências matriculadas do que as escolas regulares, 8,2% contra 11,3% (RAVITCH, 2012b). Backer (2012c) fez um estudo para a cidade de Nova York, no período entre 2008 e 2010, e concluiu que, além do problema relatado pela autora, o gasto por aluno é superior às escolas do sistema, por receberem muitas doações de empresas.

As escolas *Charters* são um excelente negócio para os grupos que as administram e têm despertado cada vez mais o interesse dos empresários. Uma entrevista com o responsável por uma empresa que tem em carteira cerca de US\$ 3 bilhões de investimentos, sobre qual a opinião dele quanto a elas, pode dar uma ideia real da importância desse novo negócio:

ó Eu acho que é um negócio muito estável, muito resistente à recessão. É um produto de alta demanda. Há 400.000 crianças em listas de espera para as escolas *charters*, a indústria está crescendo cerca de 12%, 14% ao ano. [...] se você faz negócios com Estados com tesouros sólidos, então é um negócio muito sólido. [...] É



uma oportunidade anual de dois e meio bilhões de dólares (REGUNBERG, 2012, tradução nossa).<sup>20</sup>

Portanto, as escolas *Charters*, na visão das grandes corporações, nada mais são que uma ótima oportunidade de negócios. Não é por outro motivo que elas vêm se expandindo também em outros países, como a Inglaterra e a Nova Zelândia.

Na Inglaterra, as escolas *Charters* são conhecidas como *Academy* e na Nova Zelândia como *Partnership Schools*. Se nos EUA a justificativa para seu aparecimento foi a necessidade de melhorar o desempenho dos estudantes americanos, em comparação com seus colegas de outros países, qual seria a justificativa para que esses dois países, que possuem desempenho maior no Pisa que os americanos, a adotassem? Para discutir essa questão, construiu-se o Quadro 1.1 indica os desempenhos dos estudantes na Finlândia, EUA, Nova Zelândia e Inglaterra:

**Quadro 1.1 – Proficiência média no PISA e classificação entre os 63 países participantes: média OCDE – Finlândia, EUA, Nova Zelândia e Inglaterra – 2009**

	Média OCDE	Finlândia	EUA	Nova Zelândia	Inglaterra
Matemática	500	545 (6°)	480 (31°)	520 (13°)	495 (28°)
Ciências	500	555 (2°)	500 (23°)	535 (7°)	515 (16°)
Leitura	500	535 (3°)	500 (17°)	520 (7°)	495 (25°)

Fonte: OCDE, 2011, com elaboração do autor.

Pelo Quadro 1.1, observa-se que o desempenho da Nova Zelândia é superior à média da OCDE e do resultado americano e próximo do resultado finlandês; portanto, a justificativa de implantar, a partir de 2012, como anunciado pelo governo local, reformas como as americanas, para melhorar o desempenho no teste, não parece lógico. Quanto à Inglaterra, apesar de o desempenho ser pior que o da Nova Zelândia, ainda assim é superior ao dos EUA. Dessa forma, parece que a opção pelas escolas *Charters* nesses dois países está mais ligada a questões relacionadas à ampliação do espaço para a educação privada, dentro da lógica do Quase-Mercado.

#### *Formação de professoras e sua retenção nas escolas*

<sup>20</sup> I think it's a very stable business, very recession-resistant. It's a high-demand product. There's 400,000 kids on waiting lists for charters schools, the industry's growing about 12-14% a year. [...] if you do business with states with solid treasuries then it's a very solid business. [...] It's a two and a half billion dollar opportunity set annually.

Estão se buscando alternativas para uma nova forma de avaliar os cursos de formação de professoras. Ravitch (2012h) comenta sobre uma nova proposta do governo federal, de avaliar os cursos de formação inicial a partir do desempenho nos testes avaliativos, dos alunos ensinados pelos graduados. Segundo a autora, quer-se ampliar o espectro das punições relacionadas ao baixo desempenho dos alunos, punindo-se sua professora, seu diretor, sua escola e a universidade que preparou sua professora. Opondo-se a essa possível política, a autora destaca que diversas organizações se manifestaram, em junho de 2012, contrárias a essa proposta, reivindicando a necessidade de discussão dos princípios para avaliar programas de preparação de professoras.

Outra preocupação é com a permanência das professoras nas escolas. O resultado de um estudo produzido por uma associação chamada National Commission on Teaching and America's Future (NCTAF) (2003) revela que está se observando o uso de critérios menos rigorosos para a contratação de professoras e se difundindo a ideia de que bons cursos de formação de curta duração, tipicamente de três meses, qualquer profissional pode se tornar uma boa professora. O estudo também revela que o problema não é com a falta de professoras, mas sim com a permanência dessas profissionais na profissão: cerca de 33% das professoras abandonam a profissão depois do 3º ano de exercício, e esse número chega a quase 50% depois do quinto ano. Segundo Darling-Hammond (2010), a alta rotatividade está relacionada a uma sensação de falta de efetividade pela professora, pois as iniciantes se sentem despreparadas, procuram desesperadamente por treinamento adicional, desenvolvem um estilo de ensinar focado em mecanismos de controle que, frequentemente, empobrecem o currículo para aquilo que pode ser controlado e, finalmente, entram em desespero quando se deparam com o resultado. Seriam essas as razões que as levariam a desistir da profissão.

Esse é um problema sério para as finanças públicas. Para se ter uma ideia desse custo, as conclusões de um estudo para o Estado do Texas mostrou que o Estado perdia entre US\$ 321 milhões e US\$ 2,1 bilhões por ano com a rotatividade das professoras (TCER, 2000).

Uma das estratégias que alguns sistemas estão implantando para não perder as melhores professoras é premiá-las financeiramente. Fulbeck (2012), em artigo publicado no *blog* do Instituto Albert Shanker,<sup>21</sup> identifica alguns problemas dessa proposta. Segundo a autora, essa política, baseada na premissa de que a professora toma decisões relativas à sua carreira em resposta ao dinheiro que recebe, não garante que as professoras se mobilizem todos os anos para receber o prêmio, a menos que o prêmio suba constantemente. Mesmo

---

<sup>21</sup> <<http://shankerblog.org>>

assim, continua a autora, outros fatores devem ser levados em conta, como, por exemplo, o fato de algumas professoras não se sentirem confortáveis com essa situação de depender do prêmio, ou se sentirem desprestigiadas, como se estivessem evitando melhorias no desempenho e na aprendizagem de seus alunos, à espera de um pagamento mais alto, ou podem se sentir ineficientes por não saberem o que mais pode ser feito para aumentar o desempenho de seus alunos para garantir a premiação ou, ainda, podem gerar sentimentos de falta de confiança entre as professoras e os gestores. Por poderem causar tantos impactos negativos, essa estratégia deve ser mais bem estudada.

*Burlando o teste: ensinando para o teste e as fraudes*

As políticas de pressão sobre as escolas e suas professoras, por meio das políticas *high stakes*, têm causado nos EUA efeitos perversos. O primeiro deles é orientar os cursos com base no que vai ser solicitado pelo teste a que os alunos serão submetidos, estreitando-se, assim, os currículos. Quanto a esse caso, existe um estudo que mostra que no Estado da Florida são gastos em média entre 38 e 40 dias por ano, cerca de 20% do ano escolar, para a preparação para os testes e suas aplicações (RAVITCH, 2012c). Em outro estudo, 85% das professoras de escolas públicas informaram que suas escolas dão menos atenção aos conhecimentos que não fazem parte do teste estadual (CEP, 2007). Outro mostra que 40% das professoras americanas informam que as professoras de suas escolas conseguem aumentar a pontuação nos testes sem melhorar a aprendizagem, e três quartos acreditam que a pontuação e o *ranking* de escolas não retratam com fidedignidade a qualidade da educação (PEDULLA; ABRAMS *et al.*, 2003).

Sobre essa realidade, Darling-Hammond (2010) traz o depoimento de um professor sobre o resultado do teste avaliativo aplicado no Texas, o *Texas Assessment of Academic Skills*, mais conhecido como TAAS:

Tenho visto mais alunos que conseguem passar no TAAS, mas não conseguem aplicar essas habilidades em nada que não esteja no formato TAAS. Tenho alunos que se saem bem no teste, mas não conseguem procurar palavras em um dicionário e

entender seus diferentes significados... Como um ensino de maior qualidade, eu não tenho certeza se eu o chamaria disso (p. 72, tradução nossa).<sup>22</sup>

Portanto, percebe-se um claro descompasso entre os resultados dos testes e o que os alunos conseguem fazer de fato. Mesmo que os testes venham a apontar um bom desempenho, não necessariamente isso de fato estará acontecendo, principalmente, ao que parece, pelas atividades de preparação para o teste, que se concentra no que será testado, deixando como secundários outros temas importantes.

Outro efeito são as fraudes para enganar o sistema (*gaming the system*), como os ocorridos nas cidades de Atlanta e Nova York. Em 2009, na cidade de Atlanta, uma investigação descobriu o envolvimento de 178 professoras e diretores que atuavam em 40 das 56 escolas investigadas, por fraudarem os resultados dos testes. Dos acusados, 80 deles confessaram a prática, que durou mais de uma década, envolvendo a alteração das respostas dos alunos no teste, depois do fim das aulas ou durante sua aplicação (STRAUSS, 2011). Em 2010, depois de uma investigação independente, comprovou-se que os resultados de Nova York foram artificialmente melhorados pela aplicação de testes com itens que, a cada ano, ficavam mais fáceis (RAVITCH, 2010). Apesar disso, o exemplo da cidade de Nova York tornou-se, durante um breve período, referência não somente nos EUA, mas em todo o mundo, inclusive no Brasil,<sup>23</sup>

Outra forma de fraudar o sistema tem sido selecionar quais alunos responderão ao teste, deixando que façam o teste aqueles que vêm demonstrando, ao longo do ano escolar, os melhores desempenhos. Os com piores desempenhos são submetidos a testes alternativos, desenvolvidos para crianças deficientes. Aliás, em tempos de pressão por resultados, passa a existir uma medicalização das crianças com problemas de aprendizagem, conforme reportagem de 2012 do jornal *The New York Times*, com o título "Remédios são usados para melhorar o desempenho de alunos pobres nos EUA".<sup>24</sup>

Darling-Hammond (2010) cita um estudo longitudinal coordenado por ela, que investigou um distrito escolar do Estado do Texas que submetia seus alunos a dois testes: o TAAS, de caráter *high-stakes*, e o Stanford-9, que o distrito também aplicava aos alunos e às escolas, sem nenhuma ligação com qualquer decisão de *accountability*. Os repetidos aumentos obtidos no TASS não foram observados no Stanford-9, nem em Matemática nem

---

<sup>22</sup> "I have seen more students who can pass the TAAS but cannot apply those skills to anything if it's not in the TAAS format. I have students who can do the test but cannot look up words in a dictionary and understand the different meanings ... As for higher quality teaching, I'm not sure I would call it that."

<sup>23</sup> Exemplo disso foram as publicações de Gall e Guedes (2009), de Margolis (2010) e da Fundação Victor Civita (2011).

<sup>24</sup> A reportagem está disponível em: <<http://ultimosegundo.ig.com.br/educacao/2012-10-12/remedios-sao-usados-para-melhorar-rendimento-de-alunos-pobres-nos-eua.html>>. Acesso em: 14 maio 2013.

em Leitura, pois, segundo a autora, tudo indicava que o TAAS, a cada ano, deixava de ser aplicado a uma grande quantidade de alunos, principalmente negros e latinos, considerados os mais vulneráveis. A autora afirma que, utilizando técnicas de regressão em busca de mudanças na pontuação do 10º grau de todas as escolas entre 1972 e 2002, observou-se que aumentos na taxa de reprovação e abandono no 9º grau produziam aumentos na pontuação de Matemática e Leitura. Darling-Hammond (2010) destaca que quando a exclusão dos alunos foi denunciada, todos passaram a ser testados e a pontuação do TAAS, evidentemente caiu, aproximando-se mais do Stanford-9; no entanto, observou-se um aumento significativo no número de itens sem marcação no cartão de respostas, indicando nova tentativa de enganar o sistema.

#### *Quem ganha com as políticas high-stakes?*

Pelo exposto, percebe-se que o *accountability*, como proposto pelas políticas americanas, tem trazido sérias consequências, sem que tenha havido mudança no desempenho dos estudantes, nos *rankings* internacionais, objetivo primeiro das reformas introduzidas no país. A esse respeito, Darling-Hammond (2010) afirma, que:

É claro que uma consequência negativa não intencional dos testes *high-stakes*, relacionada a sanções e sem os apoios adequados, tem sido a redução do acesso à escola para os alunos mais vulneráveis (p. 97, tradução nossa).<sup>25</sup>

Assim, com o objetivo inicial de não deixar nenhuma criança para trás, o que de fato vem ocorrendo é que o sistema está se esforçando para impedir que os mais vulneráveis frequentem a escola. Não se cuidam das diferenças, e a homogeneização e a performatividade passam a ser os parâmetros a ser seguidos.

A partir da análise das mudanças efetuadas pelo estado do Texas, que está provocando uma alta rotatividade de professoras, substituindo as mais experientes e mais bem pagas por outras mais novas e com salários mais baixos, Darling-Hammond (2010) destaca:

[ ] um sistema como esse conhecido como *accountability* não fornece incentivos ou capacidade para as escolas trabalharem com alunos que são difíceis de ensinar. O

---

<sup>25</sup> It is clear that an unintended negative consequence of high-stakes testing, linked to sanctions and without adequate supports, has been to reduce access to school for the most vulnerable students.

*accountability* que os alunos e familiares desejam e devem ter direito ó ou seja, um ensino de alta qualidade e o apoio pedagógico necessário para aprender o que os padrões demandam ó está faltando quando os testes substituem os investimentos como uma alavanca para a melhoria (p. 93, tradução nossa).<sup>26</sup>

A autora comenta que a responsabilização e a prestação de contas são desejáveis, e que as famílias são favoráveis a elas, mas o que de fato o *accountability* acarreta é a substituição de investimentos mais custosos e com retorno de longo prazo por outros mais baratos e com possibilidade de resultados mais rápidos. Sem um ensino de qualidade e o apoio necessário para que a aprendizagem aconteça, dificilmente o que está sendo medido refletirá melhoria na aprendizagem.

Assim, se a questão-chave colocada pelos políticos é o *accountability*, o que se precisa saber é como responsabilizar os governantes e fazê-los prestar contas sobre seu baixo desempenho em garantir que os alunos tenham as condições e os recursos necessários para aprender (DARLING-HAMMOND, 2010).

Ravitch (2012j, tradução nossa) vai além e chama a atenção para as verdadeiras intenções daqueles que clamam pelo aprofundamento das reformas, afirmando que:

Eles insistem que nossas escolas públicas estão fracassando, apesar da clara evidência das avaliações nacionais de que as pontuações dos testes nunca estiveram tão altas para qualquer grupo testado.

Eles insistem que o pagamento por mérito (*merit pay*) é necessário, apesar de nunca ter òfuncionadoö em nenhum sentido da palavra, seja elevando a pontuação dos testes, seja tornando a profissão de professora mais atraente.

Eles insistem que as escolas *Charters* são melhores que as escolas públicas, apesar de estudo após estudo mostrar que isso não é verdade.

Eles insistem que o sistema de *vouchers* ósalvaráö as crianças pobres, apesar disso não ter acontecido em nenhum distrito que o tenha adotado (Milwaukee, Cleveland, DC).

Seu objetivo é a privatização.

Seu objetivo é empurrar as escolas para um sistema de mercado, apesar de não haver qualquer evidência que tal sistema faça sentido para ninguém, exceto para aqueles que vendem material para as escolas ou que querem assumir o controle das escolas e realizar lucros cortando custos (professoras).<sup>27</sup>

---

<sup>26</sup> ò[...] such a system of so-called "accountability" provides no incentives or capacity for schools to work with students who are challenging to teach. The accountability that students and parents desire and should be entitled to ó that is, the high quality teaching and instructional support needed to learn what the standards demand ó is lacking when testing replaces investing as a lever for improvement.ö

<sup>27</sup> òThey insist that our public schools are failing, despite the clear evidence in the national assessments that test scores have never been higher for every group tested. They insist that merit pay is necessary, even though it has never "worked," in any sense of the word, not in

A autora comenta o fracasso das últimas ações levadas a cabo nos EUA e destaca o avanço da privatização do sistema público de ensino, privatização essa que só interessaria aos donos desse negócio.

É importante esclarecer, no entanto, um ponto levantado pela autora, quando afirma que está havendo aumento na pontuação dos testes. Sua afirmação baseia-se em estudo que ela produziu, baseado em dois instrumentos do Naep, conhecidos como *Main Naep*, que testa os alunos anualmente, e *Long-term Naep*, que os testa com o mesmo instrumento a cada quatro anos. Os dados, para os testes de Matemática, que comprovariam o aumento crescente destacado pela autora, estão presentes no Quadro 1.2.

**Quadro 1.2 – Variação da pontuação obtida pelos estudantes nas versões *Long-term* e *Main*, do Naep, para diversos grupos de estudantes – 1973 a 2011**

<i>Long-term</i> Naep entre 1973 e 2008			
	Idade de 9 anos	Idade de 13 anos	Idade de 17 anos
Estudantes brancos	+ 25 pontos	+ 16 pontos	+ 4 pontos
Estudantes negros	+ 34 pontos	+ 34 pontos	+ 17 pontos
Estudantes hispânicos	+ 32 pontos	+ 29 pontos	+ 16 pontos
<i>Main</i> Naep entre 1990 e 2011			
	4º grau	9º grau	
Estudantes brancos	+ 45 pontos	+ 23 pontos	
Estudantes negros	+ 36 pontos	+ 25 pontos	
Estudantes hispânicos	+ 29 pontos	+ 24 pontos	
Estudantes asiáticos	+ 29 pontos	+ 28 pontos	

Fonte: RAVITCH, 2012j.

Dessa comparação, percebe-se que estaria havendo um aumento da proficiência dos alunos, como comentado por Ravitch.

Kyung Hee Kim, psicóloga educacional e professora associada do College of William & Mary, no Estado de Virginia, em uma interessante entrevista, comenta os possíveis efeitos da Lei NCLB, como a perda de criatividade nas crianças americanas, observada na pesquisa que ela conduziu. Segundo a autora,

---

raising test scores or in making teaching more attractive as a profession. They insist that charters are better than public schools, even though study after study shows this is not true. They insist that vouchers will save poor children, even though this has not happened in any of the districts that have vouchers. (Milwaukee, Cleveland, DC). Their goal is privatization. Their goal is to push schools into a market-system despite any evidence that such a system makes any sense for anyone except those selling stuff to schools or wanting to take over schools and make a profit by cutting costs (teachers).ö

Os testes padronizados obrigam à aprendizagem rotineira, em vez de pensar crítico, criativo, e diminuem a curiosidade natural dos alunos e a alegria de aprender sozinho.

Além disso, a NCLB pode sufocar a criatividade das professoras, porque a alta pressão para cobrir o conteúdo necessário para produzir o desempenho esperado substitui o desejo (e o tempo) para estimular a imaginação e a curiosidade das crianças. A NCLB não valoriza as habilidades das professoras, que poderiam incentivar a aplicação criativa do que se aprende em sala de aula a situações da vida real. Profissionais do ensino são reduzidos a técnicos de ensino com menos capacidade de desenvolver abordagens criativas para envolver os alunos, porque são obrigados a cobrir o que é cobrado nos testes.

[í ] A NCLB tem sufocado qualquer interesse em desenvolver as diferenças individuais, o pensamento inovador e criativo ou o potencial individual<sup>28</sup> (KIM, 2010, tradução nossa).

O alerta de Kim sobre o possível efeito dos testes remete ao aumento das preocupações que o uso desenfreado deles pode estar causando ao sistema educacional americano. Uma conclusão que merece destaque é quando a autora comenta que os profissionais da educação estão se transformando em técnicos de ensino, descaracterizando a profissão, e a perda que isso traz para a aprendizagem dos alunos.

A partir das análises de Diane Ravitch e Linda Darling-Hammond, é possível perceber que existe uma aliança, ainda que informal, entre os governantes que buscam reduzir custos e melhorias a curto prazo e aqueles que querem desenvolver o mercado educacional com o objetivo de abrir uma nova frente de negócios em que possam atuar e obter mais lucros. Dessa forma, a educação nos EUA está cada vez mais se transformando em produto de consumo, em vez de um bem público, mas não sem resistência. Além de diversos ativistas que mantêm *blogs* bastante ativos e lidos,<sup>29</sup> desde junho de 2012, tem circulado um abaixo-assinado intitulado *National Resolution to Oppose High-Stake Testing*,<sup>30</sup> cobrando das autoridades

---

<sup>28</sup> Standardized testing forces emphasis on rote learning instead of critical, creative thinking, and diminishes students' natural curiosity and joy for learning in its own right. Further, NCLB may stifle teachers' creativity because the high pressure to cover the content required to produce passing test scores overrides the desire (and time) to stimulate children's imagination and curiosity. NCLB does not value teachers' skills that could encourage the creative application of classroom learning to real life situations. Teaching professionals are reduced to teaching technicians with less ability to develop creative approaches to engage students because they are required to cover what is on the tests. [í ] NCLB has stifled any interest in developing individual differences, creative and innovative thinking, or individual potential.

<sup>29</sup> Podemos citar aqui pesquisadores como Diane Ravitch (<<http://dianeravitch.net>>) e Bruce Baker (<<http://schoolfinance101.wordpress.com>>), jornalistas como Valerie Strauss (<<http://www.washingtonpost.com/blogs/answer-sheet>>), instituições como New York City Public Schools Parents (<<http://nycpublicschoolparents.blogspot.com>>) e o Albert Shanker Institute (<<http://shankerblog.org>>).

<sup>30</sup> Disponível em: <<http://timeoutfromtesting.org/nationalresolution>>.



locais, estaduais e federais o fim dos testes *high-stakes*, que contava, em maio de 2013, com assinaturas de 543 organizações e mais de 17 mil pessoas.

Todo esse movimento em direção ao Quase-Mercado da educação pode levar à morte a educação pública, pois quanto mais as pessoas enxergarem a educação como uma escolha do consumidor, mais elas estarão dispostas a não pagar pela educação dos filhos dos outros, e o senso de responsabilidade comum, fundamental em uma sociedade democrática, acaba se dissolvendo (RAVITCH, 2012o). O público deve sempre ser o responsável pela educação, pois toda a sociedade acaba se beneficiando, tanto quando nossos filhos são educados como quando os filhos dos outros o são.

### **O *accountability* é o único caminho?**

Pasi Sahlberg, um finlandês que vem participando do desenvolvimento de políticas educacionais em seu país e as discutindo mundo afora, afirma que existe um fenômeno que está se espalhando pelos sistemas escolares ao redor do mundo. A esse fenômeno, Sahlberg dá a sugestiva alcunha de *Germ* (germe), abreviatura de *Global Educational Reform Movement* (Movimento Global de Reforma Educacional).

É como uma epidemia que se espalha e infecta os sistemas de ensino através de um vírus. Ele viaja com especialistas, mídia e políticos. Os sistemas de educação emprestam políticas dos outros e são infectados. Como consequência, as escolas ficam doentes, as professoras não se sentem bem, e as crianças aprendem menos (SAHLBERG, 2012, tradução nossa).<sup>31</sup>

Essa analogia de relacionar as reformas educacionais a um germe que se dissemina e que cria doenças com sintomas muito claros é muito feliz, pois é isso exatamente o que estamos observando em diferentes países: a aplicação de fórmulas muito semelhantes em diferentes contextos, com objetivos semelhantes. Três seriam os sintomas dessa doença.

O primeiro é a maior competição na educação. Essa competição se baseia na crença de que a educação se aprimora quando as escolas competem entre si, criando o que Afonso (1999, 2005) chamou de Quase-Mercado. Para isso, o aumento dos testes avaliativos e a medida da efetividade das professoras são essenciais.

---

<sup>31</sup> It is like an epidemic that spreads and infects education systems through a virus. It travels with pundits, media and politicians. Education systems borrow policies from others and get infected. As a consequence, schools get ill, teachers don't feel well, and kids learn less.

O segundo sintoma é o *Parent Choice*, já discutido anteriormente, que se refere ao ato de oferecer às famílias a oportunidade de elas matricularem seus filhos na escola que acharem a mais adequada, seja ela pública ou privada, reforçando o desenvolvimento do Quase-Mercado, com as escolas atraindo essas famílias. De acordo com Sahlberg (2011), políticas como essa vêm se tornando comuns, na forma das escolas *Charters* nos EUA, das *Secondary School Academies* na Inglaterra ou das *Free Schools* na Suécia. Para o autor, países que seguem esse caminho têm observado o declínio de seus resultados escolares e a ampliação da segregação escolar.

O terceiro sintoma é a forte responsabilização das escolas e a busca da efetividade da professora, ambos tendo por base os testes avaliativos aplicados aos estudantes, hoje em dia, a forma mais comum de decidir se as escolas e as professoras estão ou não realizando um bom trabalho. Sahlberg (2011) diz que o aumento dos testes avaliativos têm causado efeitos danosos e indesejados, como o aumento do ensino focado nos testes, o estreitamento do currículo para priorizar a Leitura e a Matemática e aproximação da Pedagogia a uma instrução mecânica.

O autor destaca, também, os efeitos danosos sobre a vida dos alunos e cita o crescente número de alunos, na Coreia do Sul e no Japão, que estão cometendo suicídio por não conseguirem suportar as pressões dos adultos por desempenhos cada vez maiores. Afirma, ainda, que é preciso parar esse germe que coloca tamanha pressão nas crianças mediante competição, *Parent Choice* e *accountability*.

Mas esse modelo, que se baseia no *accountability* para buscar melhorias nos sistemas educacionais, apesar de hegemônico, não é o único existente. Um exemplo disso é a Finlândia.

A Finlândia, por séculos, lutou muito por sua própria identidade, por sua língua materna e por seus próprios valores; primeiro, durante seis séculos, sob o reinado da Suécia; depois, por mais de um século, sob o império russo; e depois, por mais um século, como uma nação livre, espremida entre a antiga União Soviética e as potências da globalização (SAHLBERG, 2011).

Esse pequeno país nórdico com 5,5 milhões de habitantes está no topo de vários *rankings*, como os das economias mais competitivas, de maior transparência corporativa, de maior qualidade de vida e bem-estar geral. Além disso, no campo educacional, possui indicadores que se destacam em relação aos dos outros países.

No blog *Bridging Differences*, da revista americana *Education Week*, Diane Ravitch e Deborah Meier debateram tempos atrás seus diferentes pontos de vista sobre educação e o

fizeram na forma de cartas dirigidas uma a outra. Em uma dessas cartas, Ravitch (2012n), bem ao estilo dos antigos comparativistas, que narravam o que vivenciavam em suas viagens ao estrangeiro, procurando traçar paralelos com fatos de seu próprio país (BERADAY, 1972), relata sua visita à Finlândia para conhecer o sistema educacional do país e destaca alguns pontos que, segundo ela, explicariam o sucesso desse sistema.

A autora apresenta uma longa lista de observações sobre o sistema educacional finlandês. Segundo ela, os estudantes nunca são submetidos a testes estandardizados até o fim da educação básica, quando prestam exame de seleção para a educação superior. Como são as próprias professoras que elaboram seus testes, elas sabem o momento da aprendizagem em que estão seus alunos e o tipo de reforço que eles precisam. Com relação ao currículo, o que existe são amplas diretrizes gerais, com o objetivo de assegurar que todos os estudantes tenham educação completa, mas as professoras têm bastante autonomia e responsabilidade para elaborar currículos para sua escola e decidir quais estratégias pedagógicas utilizar, e são elas as únicas autoridades que monitoram o progresso dos alunos.

Quanto aos prédios escolares, Ravitch destaca que são arquitetonicamente bonitos e funcionais, cheios de luz. As turmas são pequenas e, apesar de, oficialmente, o número permitido de crianças por sala nos anos iniciais ser de 24 alunos, nas diversas escolas que visitou, ela não observou menos de 19 alunos por turma. Em apenas uma, observou uma turma com o máximo de 24 alunos, mas havia duas professoras assistentes para ajudar as crianças com dificuldades de aprendizagem.

Com relação à profissão docente, a autora destaca que ela é altamente valorizada e disputada, concorrendo com carreiras como as de médico e advogado, sendo procurada pelos alunos com maior desempenho na etapa equivalente ao nosso ensino médio, no qual menos de 10% dos candidatos são aceitos. O programa de formação, em que os alunos são preparados para diversas situações como, por exemplo, lidar com crianças com deficiências ou com dificuldades de aprendizagem, ou ainda com aquelas que não dominam a língua, dura cinco anos em período integral, sendo três anos de bacharelado e dois de mestrado. Segundo relato das professoras, a chave do sucesso da escola finlandesa é a confiança, pois as famílias confiam nas professoras, porque elas são profissionais; as professoras confiam umas nas outras e colaboram entre si na solução de problemas; e há também a confiança mútua entre as professoras e os diretores, porque estes já foram professores e possuem ampla experiência. Ravitch (2012n) destaca, também, que dificilmente as professoras deixam a profissão, porque sentem possuir um bom trabalho e se consideram respeitadas.

Em relação ao desempenho dos alunos, Darling-Hammond (2010) destaca que, de acordo com os resultados do Pisa de 2006, a variância entre escolas na escala do teste de ciências era de apenas 5%, enquanto que a variância média entre os outros países da OECD era de quase 33%. Esse fato é mais impressionante porque em algumas escolas urbanas o total de crianças imigrantes, vindas de mais de 60 diferentes países, ou aquelas em que a língua materna não é o finlandês, aproxima-se de 50%. A autora afirma que não existem testes avaliativos externos, mas que, a exemplo do Naep americano, o governo finlandês desenvolveu testes avaliativos aplicados a amostras de estudantes no final do 2º e do 9º ano de escolaridade, não com o objetivo de aplicar sanções às escolas, mas para obter informações que permitam desenhar soluções que auxiliem na aprendizagem dos alunos.

Sahlberg (2011), por sua vez, afirma que o país chegou a tal situação depois de uma grande mobilização ocorrida com a Queda do Muro de Berlin e do colapso da protegida economia soviética nos anos 1990. O caminho para as mudanças foi pavimentado por sucessivos aprimoramentos que começaram nos anos seguintes à Segunda Guerra e que ganharam impulso nas três décadas, a partir de 1980, cada uma delas marcada por objetivos específicos. Nos anos 1980, buscou-se repensar os fundamentos teóricos e metodológicos, envolvendo as concepções de conhecimento e aprendizagem, os métodos de ensino e o currículo. Os anos 1990 foram dedicados ao aprimoramento da implantação do currículo, tendo por base a colaboração entre as escolas, a discussão intensa nas comunidades de aprendizagem recém-criadas e as mudanças relacionadas a sua implantação, fundamentadas em um grande movimento de descentralização, que levou ao aumento da autonomia escolar. Nos anos 2000, buscou-se o aumento da eficiência das estruturas e da administração, que começou logo depois dos primeiros resultados do Pisa, indicando o alto desempenho dos alunos finlandeses, que, comparados com os alunos de outros países que estavam no topo da lista, não necessitavam de aulas particulares, atividades escolares extraclases ou grande quantidade de trabalho de casa. Sahlberg, contudo, demonstra preocupação com o futuro das conquistas finlandesas, pois a atual preocupação com o aprimoramento da eficiência e da produtividade tem levado a uma diminuição do orçamento escolar, que procura fazer o mesmo com menos recursos, e isso vem inquietando as lideranças dos diretores escolares e das professoras.

Os aprimoramentos iniciados nos anos 1980 e a mobilização dos anos 1990 deram origem a mudanças que transformaram o ambiente educacional existente, baseadas em algumas premissas. Entre estas, Sahlberg (2011) destaca a concepção de que as mudanças educacionais deveriam estar conectadas com estratégias que privilegiassem a inclusão e a

multiplicidade de soluções, em vez de estratégias estandardizadas. Outra premissa é que deveria haver confiança nas professoras, e que estas deveriam ser de alta qualidade, bem treinadas, e que fossem atraídas para a profissão pela sua missão social, por sua autonomia e pelo suporte que poderiam ter para desenvolver um bom trabalho. Previa-se, também, que as professoras ensinassem menos, e que os estudantes gastassem menos tempo estudando dentro e fora da escola. Não se podia negligenciar uma estratégia de educação especial inclusiva, que permitisse que os estudantes com dificuldades de aprendizagem tivessem o suporte para se desenvolver. Todo o trabalho deveria ser baseado na crença de que o isolamento seria o inimigo de qualquer melhoria, e que as professoras e as escolas se desenvolvem aprendendo juntas com outras professoras e outras escolas. Por fim, e talvez o mais importante, era considerar a educação não como um bem individual, mas sim como um esforço público que serve a um propósito público e, sendo assim, com qualidade para todos.

Sahlberg (2011), como que antecipando as críticas que atribuem o sucesso do modelo finlandês a fatores que só se encontram naquele país, destaca que, para tornar essas premissas realidade, o fato de o sistema educacional finlandês estar entrelaçado com o Estado de Bem-Estar Social contribui bastante, mas apenas isso não seria suficiente para manter um sistema com elevados e consistentes resultados educacionais. Para isso, foi necessário basear-se em pesquisas e contar com a colaboração de acadêmicos, agentes públicos, diretores de escola e professoras, para construir estruturas que dessem suporte ao sistema escolar. Uma dessas estruturas refere-se ao financiamento educacional, que garante igualdade de recursos para todas as escolas, independentemente de sua localização ou da prosperidade de sua comunidade. Outra é a atenção dada à criança, que, por lei, deve ter acesso à assistência à infância, cuidados de saúde e pré-escola em sua própria comunidade, exigindo, inclusive, que toda escola tenha uma equipe que garanta sua felicidade. Além disso, a Constituição finlandesa considera a educação um direito humano e, assim toda a educação, da pré-escola até a universidade, é gratuita para todos os moradores da Finlândia, permitindo, com isso, que a educação superior esteja disponível e acessível a todos. Portanto, o que o autor destaca é que as estruturas que por lá se desenvolveram estavam todas integradas e ligadas a objetivos de longo prazo, não eram ações isoladas das quais se esperavam resultados imediatos.

Outro ponto destacado por Sahlberg (2011) é que muitas delegações oficiais, principalmente originárias de países desenvolvidos em que a educação pública tem falhado em proporcionar oportunidades de aprendizagem adequadas a todas as crianças, têm visitado a Finlândia, buscando compreender melhor como seu sistema educacional pode produzir resultados tão consistentes. De início, concentram sua curiosidade na busca por inovações

pedagógicas e tecnologias de ponta e, ao contrário, têm encontrado professoras ensinando e alunos aprendendo, como se faz em qualquer outro país, com a diferença de que estes encontram um sistema educacional bem menos formal e mais sereno. O autor enfatiza, surpreendentemente, que muitas das estratégias utilizadas hoje nas escolas finlandesas foram fortemente inspiradas nas ideias educacionais americanas e inglesas e estão intimamente alicerçadas em resultados de pesquisas desenvolvidas em vários países, destacando, como exemplos, a aprendizagem cooperativa e a avaliação por portfólio, dos EUA; o ensino de ciência e matemática dos EUA, da Austrália e da Inglaterra; e o desenvolvimento de lideranças locais, do Canadá. Por outro lado, destaca o autor, o sonho finlandês de educação é um produto local, de propriedade do povo e não importado de nenhuma parte.

Baseado em um quadro elaborado por Sahlberg (2012), comparando os elementos-chave do *Germ* e as políticas educacionais do início dos anos 1990 na Finlândia, montou-se o Quadro 1.3.

**Quadro 1.3 – Elementos chaves do Movimento Global de Reforma Educacional (Germ), em comparação com as políticas educacionais da Finlândia, desde os anos 1990**

<b>Germ</b>	<b>A maneira finlandesa</b>
Ensino e aprendizado estandarizados.	Ensino e aprendizado customizados.
Foco em leitura, matemática e ciências.	Foco no aprendizado criativo, envolvendo, de forma equilibrada, além das diferentes áreas do conhecimento humano, as habilidades necessárias para o desenvolvimento da personalidade do indivíduo.
Currículo escolar prescritivo.	Encorajamento do currículo baseado na escola e desenvolvido pelas professoras, em que é aceitável correr os riscos e as incertezas quanto ao ensino e à aprendizagem.
Reformas orientadas para o mercado.	Aprender a partir do passado e inovar.
<i>Accountability</i> baseado em testes e controle.	Responsabilidade compartilhada e confiança.

Fonte: SAHLBERG, 2012, com elaboração do autor.

Pelo quadro, é possível verificar que existe realmente uma concepção diferente sobre como melhorar as aprendizagens dos alunos. Uma concepção, que apesar de ter sido desenvolvida em um pequeno país, pode ser implantada, em partes e com as devidas adaptações, em outros países, pois está baseada nos mais fundamentais conceitos do que seja

o ato de educar. Tanto é assim, que o papel mais importante nessa nova concepção tem sido aquele reservado às professoras.

Reijo Laukkanen, chefe do órgão educacional máximo finlandês, em uma conferência internacional sobre a qualidade das professoras, realizada na Finlândia, em 2009, patrocinada pela Fundação Person e pelo CCSSO americano, da qual participaram líderes americanos e de outros de nove países, assim se manifestou sobre a importância das professoras:

O empoderamento da profissão docente produz bons resultados. Os profissionais docentes devem ter espaço para a inovação, porque eles devem tentar encontrar novas maneiras de melhorar a aprendizagem. As professoras não devem ser vistas como técnicas cujo trabalho é implementar programas de estudo estritamente ditados, mas sim como profissionais que sabem como melhorar a aprendizagem para todos. Tudo isso cria um grande desafio que certamente exige mudanças nos programas de formação de professoras. As professoras são as mais importantes, porque os sistemas de ensino trabalham através delas (PEARSON FOUNDATION; COUNCIL OF CHIEF STATE SCHOOL OFFICERS, 2009, p. 6, tradução nossa).<sup>32</sup>

Portanto, a profissão docente recebe um tratamento de destaque na Finlândia, quando comparada com outros países. Quando Laukkanen se refere à profissão docente, reafirma aquilo que foi comentado por Kim (2010): que as profissionais docentes são mais do que técnicas educacionais, são professoras, e que elas foram um dos pontos centrais das mudanças introduzidas na Finlândia, a partir dos anos 1980. Foi através delas que as mudanças puderam ser implementadas com o sucesso que os testes avaliativos internacionais mostram, apesar das ressalvas com relação ao significado dos resultados de qualquer teste avaliativo.

Com relação à qualidade da professora, Sahlberg (2011) afirma que:

A questão da eficácia da professora ou as consequências de ser uma professora ineficaz não é relevante na Finlândia. [...] as professoras têm tempo para trabalhar juntas durante o dia escolar e compreender como os seus colegas ensinam. Esta é uma importante condição para a professora refletir sobre sua forma de ensinar e também para construir o compartilhamento de responsabilidades entre as professoras. O sistema de inspeção escolar, que anteriormente fornecia *feedback* externo e avaliava como as professoras ensinavam e como a escola funcionava, foi abolido no início dos anos 1990. [...] O pressuposto básico nas escolas finlandesas é

---

<sup>32</sup> "Empowerment of the teaching profession produces good results. Professional teachers should have space for innovation, because they should try to find new ways to improve learning. Teachers should not be seen as technicians whose work is to implement strictly dictated syllabi, but rather as professionals who know how to improve learning for all. All this creates a big challenge [ ] that certainly calls for changes in teacher education programs. Teachers are ranked highest in importance, because educational systems work through them."

que as professoras, por definição, são profissionais bem formadas e que estão fazendo o seu melhor nas escolas (p. 91, tradução nossa).<sup>33</sup>

Portanto, cria-se na sociedade finlandesa uma forte relação de compromisso entre os gestores do sistema educacional e os responsáveis pela escola. É importante destacar que essa filosofia é anterior aos primeiros resultados do Pisa e, portanto, que as medidas adotadas tomaram como base princípios pedagógicos sobejamente conhecidos e reconhecidos como ideais em todo o mundo e não baseadas em preocupações fundamentadas nos desempenhos dos alunos em testes avaliativos.

Darling-Hammond (2010) afirma que os finlandeses compreenderam que se as professoras aprenderem como ensinar os alunos com dificuldades, elas estarão preparadas para ensinar todos os estudantes, para assim não deixar nenhuma criança para trás (p. 171). Essa afirmação é uma clara provocação à política americana da lei do NCLB, baseada quase que exclusivamente no *accountability*, e prova que para conseguir a aprendizagem são necessários bons profissionais docentes e uma rede de apoio para garanti-la.

Com relação aos salários das professoras, Sahlberg (2011) comenta que elas ganham, em média, o mesmo que suas colegas da OCDE, usando como medida o dólar paridade de poder de compra (ppp), um pouco acima da média dos salários finlandeses, e recebem aumentos em função dos anos de experiência adquiridos, independentemente de qualquer tipo de avaliação. Com relação à sindicalização, 95% das professoras de todos os níveis são filiadas ao Trade Union of Education in Finland, o sindicato nacional. Desde 1973, a principal atribuição do sindicato tem sido negociar os termos dos contratos de emprego e atuar como um dos interlocutores sobre educação.

É interessante destacar, para finalizar esta abordagem sobre a Finlândia, um trecho do livro de Sahlberg (2011):

As políticas de educação destinadas a aumentar o desempenho do aluno na Finlândia têm destacado o ensino e a aprendizagem, incentivando as escolas a criar ambientes de aprendizagem ideais e a desenvolver conteúdos instrucionais que melhor ajudem os alunos a alcançar os objetivos gerais de escolaridade. Assumiu-se muito cedo, no processo de reforma da Finlândia, que a instrução é o elemento-chave que faz a diferença no que os alunos aprendem na escola, não os padrões, os testes avaliativos,

---

<sup>33</sup> "The question of teacher effectiveness or consequences of being an ineffective teacher is not relevant in Finland. [...] teachers have time to work together during a school day and understand how their colleagues teach. This is an important condition for reflecting on teachers' own teaching and also building shared accountability between teachers. The school inspection system that previously provided external feedback and evaluation of how teachers taught and school operated was abolished in the early 1990s. [...] The basic assumption in Finnish schools is that teachers, by default, are well-educated professional and are doing their best in schools."



ou os programas de instrução alternativos. Com o aumento gradual do nível de profissionalismo da professora nas escolas durante os anos 1990, a prevalência de métodos de ensino eficazes, os projetos pedagógicos escolares e de salas de aula aumentaram. Uma nova flexibilidade dentro do sistema de ensino finlandês possibilitou que as escolas aprendessem umas com as outras, permitindo, assim, universalizar as melhores práticas, adotando abordagens inovadoras para organizar a escolaridade. Isso também incentivou as professoras e as escolas a continuarem a expandir seus repertórios de métodos de ensino e a individualizar o ensino, a fim de atender as necessidades de todos os alunos (p. 39, tradução nossa).<sup>34</sup>

Essa reflexão mostra uma nítida preocupação com a aprendizagem, com o processo pedagógico, e não com o desempenho baseado em pontos obtidos em testes avaliativos; mostra um caminho diferente em busca da qualidade educacional. Para o autor, a experiência finlandesa demonstra que um foco consistente na equidade e na cooperação, não na competição e na *accountability*, pode construir um sistema educacional em que todas as crianças consigam aprender. Para se chegar a esse ponto, é necessário um grande acordo nacional, a definição clara de uma política educacional ó não apenas ações pontuais com objetivos também pontuais ó , a confiança e a solidariedade. Além disso, no caso de países como o Brasil, também são necessários muitos recursos.

### ***Para se compreender os instrumentos utilizados***

Para que seja possível avançar nas análises que serão feitas nos próximos tópicos desse texto, é necessário esclarecer o significado de uma série de expressões e discutir as teorias subjacentes aos instrumentos utilizados nos testes avaliativos.

Como ocorre em qualquer processo de medição, é necessário estabelecer uma referência tanto para comparar os resultados obtidos como para a elaboração dos instrumentos que serão utilizados para realizar a medição. No caso dos testes avaliativos, essa referência são as Matrizes de Referência, que mostram o que será medido e como isso será realizado; portanto, elas são a base para a elaboração dos itens que comporão o teste, o instrumento de

---

<sup>34</sup> Education policies designed to raise student achievement in Finland have put a strong accent on teaching and learning by encouraging schools to craft optimal learning environments and establish instructional content that will best help students to reach the general goals of schooling. It was assumed very early in Finland's reform process that instruction is the key element that makes a difference in what students learn in school, not standards, assessment or alternative instructional programs. As the level of teacher professionalism gradually increased in schools during the 1990s, the prevalence of effective teaching methods and pedagogical classroom and school designs increased. A new flexibility within the Finnish education system enabled schools to learn from one another and thus make the best practices universal by adopting innovative approaches to organize schooling. It also encouraged teachers and schools to continue to expand their repertoires of teaching methods, and to individualizing teaching in order to meet the needs of all students.

medida utilizado. Nesse processo, é fundamental a definição de quais habilidades, entre outras possíveis, serão medidas. Mas essa definição não pode ser arbitrária; deve estar referenciada a uma realidade sobre a qual se quer obter informações. Por isso, as Matrizes devem estar intimamente relacionadas com o currículo escolar, para que a medida tenha relação com o processo que se desenvolve na escola. Caso contrário, medem-se outras coisas que não aquelas relacionadas com a educação escolar, as quais podem até ser muito importantes, mas se distanciam do principal objetivo, que é compreender melhor o processo educacional. Nesse processo inicial, é importante, também, que a escolha das habilidades a serem medidas possam ser testadas, em especial por um teste composto por itens de múltipla escolha.<sup>35</sup> Para ilustrar esse problema da escolha das habilidades a serem medidas, tome-se a habilidade *cálculo mental*, fundamental para, entre outras coisas, permitir que o aluno possa rapidamente avaliar se os cálculos que fez usando os algoritmos aprendidos estão corretos ou não. Apesar de ser uma habilidade importantíssima, que todos nós usamos na vida cotidiana, ela não é passível de ser medida em um teste de múltipla escolha, já que qualquer item que tente medir essa habilidade permite que o aluno utilize diversas estratégias que não somente o cálculo mental para resolver o problema proposto. Essa observação é muito importante para compreender que, por melhor que seja o instrumento que se utilize, a realidade é muito mais complexa e, portanto, nem tudo que deve ser trabalhado em sala é passível de medição. Sendo assim, classificar o resultado obtido relacionando-o com uma pretensa qualidade que poderia estar relacionada a uma preparação para a vida em sociedade é inadequado e absolutamente impreciso.

Para que o resultado dos testes possa fornecer uma grande quantidade de informações, o ideal é que seja utilizada uma grande quantidade de itens, pois, dessa forma, pode-se testar um número significativo de habilidades e cada uma delas em diferentes contextos e abordagens. Assim, seria possível, para uma mesma habilidade, observar o comportamento dos alunos quando submetidos a diferentes situações. Mas, para isso, existem limites de tempo de aplicação do teste e de esforço que o aluno pode dispendir para que a fadiga não venha a prejudicar seu desempenho, mascarando o resultado obtido.

Se um dos objetivos dos testes avaliativos é acompanhar os possíveis efeitos de alguns programas educacionais implementados, é fundamental ter instrumentos que permitam que os resultados obtidos em diferentes testes possam ser comparáveis ao longo do tempo. Uma das possibilidades para isso ocorrer seria submeter os alunos aos mesmos testes ano após ano, o

---

<sup>35</sup> Como na maior parte dos testes avaliativos utilizam-se itens de múltipla escolha, este texto fará referência a eles, apesar de que técnicas semelhantes podem ser utilizadas para itens de resposta parcialmente ou totalmente construída. A elaboração de itens de múltipla escolha é um processo complexo. Sobre o tema, consultar Haladyna. (1997, 2004).

que seria impraticável, pois, com o tempo, é possível que os alunos se preparem previamente para a aplicação do teste, trazendo perda de confiabilidade nos resultados obtidos.

A solução para esse problema veio com a utilização da Teoria da Resposta ao Item (TRI).

### **A Teoria da Resposta ao Item (TRI)**

Para compreender sua utilização, é necessário esclarecer que é uma teoria estatística desenvolvida para auxiliar, na medida de traços latentes, fenômenos existentes que não podem ser medidos diretamente, como, por exemplo, a proficiência em Matemática. Mas também pode ser aplicada em outras áreas, como, por exemplo, para medir o grau de estado de depressão de um paciente ou o grau de satisfação com um serviço.

Como toda teoria, ela apresenta alguns pressupostos. Um deles é sua propriedade de invariância:

- Nos parâmetros dos itens. Depois de o item ter sido aplicado a uma população e com base nas respostas dos respondentes, é possível obter três atributos para classificar o item, os quais, uma vez determinados, não sofrem qualquer alteração se esse item for aplicado a diferentes populações, em diferentes tempos, e se estiver presente em testes também diferentes. Os três atributos obtidos são a *discriminação*, a capacidade de distinguir de forma eficiente populações com diferentes proficiências, ou seja, para acertar determinado item, exige-se do respondente determinada proficiência mínima que, se o respondente não a possuir, a probabilidade de ele acertar o item será baixa; a *dificuldade*, relacionada com a quantidade de pessoas que acertam o item, ou seja, quanto maior a quantidade de pessoas que acertam o item, mais fácil este é, e mais difícil quanto menor o número de pessoas a acertá-lo; a *probabilidade* de indivíduos com baixa proficiência acertarem o item, ou seja, a probabilidade de acertar a resposta õchutandoõ. Para se obter esses atributos, os itens são previamente testados em populações semelhantes àquelas que se submeterão ao teste. Depois disso, é possível escolher itens com determinados atributos para compor o teste, garantindo, assim, maior precisão e confiabilidade na medida.

- Na proficiência dos indivíduos. Uma vez determinada a proficiência de um indivíduo a partir do seu desempenho frente a um grupo de itens, ela será a mesma se o mesmo indivíduo responder a outros itens com parâmetros semelhantes. Mas é claro que conforme o indivíduo for desenvolvendo suas habilidades ou aprendendo novas habilidades pertencentes a um mesmo traço latente, sua proficiência para esse traço latente tende a ser maior.

Como a TRI é uma teoria, se todas as suas premissas não forem respeitadas, corre-se o risco de que o resultado obtido não seja confiável. Autores como Tavares (2012) se valem exatamente desse aspecto para colocar em dúvida os resultados divulgados depois da aplicação dos testes e propalados como precisos.

Portanto, com a TRI resolve-se a dificuldade quanto à comparabilidade dos testes. Não é mais necessário aplicar testes iguais para diferentes populações; basta ter itens calibrados, ou seja, itens para os quais se conheçam seus três atributos, possibilitando com isso comparar as diferentes proficiências medidas. Dessa forma, se entre dois períodos quaisquer observa-se um aumento de proficiência de determinado traço latente, isso significa que, entre as gerações de populações testadas, as habilidades mais complexas foram desenvolvidas.

Outra característica da TRI é que a proficiência é um dos termos de uma função probabilística. Assim, para determinado item, existe uma probabilidade de acerto que depende da dificuldade do item e da proficiência do indivíduo. Portanto, uma proficiência não se refere a um valor exato, mas a uma probabilidade, baseada nas respostas dadas a um conjunto de itens.

### **A pré-testagem dos itens**

Depois de os itens terem sido elaborados, observando-se algumas técnicas que garantam sua qualidade pedagógica (INEP, 2010), e antes de fazerem parte de um teste, eles devem ser pré-testados, tanto para obter seus parâmetros como para verificar se existe alguma inconsistência com relação ao gabarito.

A pré-testagem implica submeter os itens a uma população com características semelhantes àquela que será testada. Não é aconselhável pré-testar os itens na mesma população que participará do teste, pois, dessa forma, estar-se-á dando a conhecer a eles o instrumento que seria aplicado no futuro. Assim, por exemplo, se deseja aplicar o teste ao

final do 5º ano do ensino fundamental, uma boa alternativa seria pré-testar os itens com a população frequentando o início do 6º ano.

Depois de os itens serem pré-testados e com base nos parâmetros obtidos, decide-se quais deles estão aptos para compor os testes, quais podem ser aprimorados para serem novamente pré-testados e quais devem ser descartados. Os itens aprovados são armazenados, juntamente com seus parâmetros, em um banco de dados conhecido como Banco de Itens. Como pode acontecer de apenas uma quantidade pequena de itens ser aprovada para fazer parte do Banco de Itens, recomenda-se testar quatro ou cinco vezes o número de itens que se quer utilizar no teste avaliativo.

Uma alternativa para melhorar a qualidade dos itens, antes de eles serem pré-testados, é submetê-los a um processo conhecido como Laboratório Cognitivo. De forma bastante simplificada, esse processo envolve aplicação de testes com pequenas quantidades de itens a grupos de alunos voluntários. Durante a resolução dos itens, os pesquisadores discutem com os alunos as estratégias que eles estão utilizando para resolver os itens e acompanham o desenvolvimento do grupo, desde a leitura do item até sua resolução. Com base nas informações coletadas, os itens podem ser mais aprimorados ou definitivamente descartados. Por ser um processo de menor custo que um pré-teste, é interessante utilizá-lo, mas para isso é necessário desenvolver toda uma metodologia, algo que o Inep vem realizando.

### **A montagem dos testes**

Para a montagem dos testes, selecionam-se, entre os itens que compõem o Banco de Itens, aqueles que satisfaçam algumas das condições previamente estabelecidas pelos testes, principalmente com relação às habilidades que se quer medir. Como comentado antes, quer-se utilizar a maior quantidade de itens, mas é impraticável que todos os alunos respondam a todos eles. Para resolver isso, os itens são agrupados por características comuns, e cada grupo constitui um caderno. Normalmente, em cada caderno, utiliza-se uma proporção de 15% de itens fáceis, 70% de itens médios e 15% de itens difíceis, e esses itens medem uma parte das habilidades que se quer testar. Cada aluno recebe um teste composto, geralmente, por dois ou quatro dos cadernos montados. A distribuição dos itens pelos diferentes cadernos deve ser tal que, no total, dentro do menor extrato que se quer medir, por exemplo, a escola, todas as habilidades sejam testadas. Nesse modelo, como cada aluno responde a cadernos contendo itens com diferentes habilidades, não se calcula sua proficiência individual, uma vez que o

traço latente que se está medindo leva em consideração todas as habilidades previstas na Matriz de Referência. Mas como todos os itens do teste foram utilizados no interior da escola, calcula-se a proficiência da escola. Essa proficiência, na realidade é a proficiência média dos alunos que participaram do teste e que são alunos dessa escola.

### **A escala de proficiência e sua interpretação pedagógica.**

Aplicado o teste, calcula-se, com base na TRI, as proficiências. Essas proficiências são expressas por valores numéricos dentro de um intervalo arbitrário, por exemplo, entre 0 e 500, como é o caso do Saeb. Esse intervalo constitui a escala de proficiência. Essa escala é dividida em diversos intervalos, chamados *níveis de proficiência*. Para se compreender melhor a divisão da escala, pode-se, por exemplo, dividi-la em cinco intervalos: nível 1 ó entre 0 e menor que 125; nível 2 ó maior ou igual a 125 e menor que 200; nível 3 ó maior ou igual a 200 e menor que 275; nível 4 ó maior ou igual a 275 e menor que 350; nível 5 ó maior ou igual a 350.

Como a medida obtida para uma escola refere-se a uma proficiência média, ou seja, uma proficiência calculada a partir dos diferentes alunos que participam do teste, faz todo sentido associar essa proficiência média a um nível. Assim, seguindo o exemplo anterior, se em uma escola a proficiência média dos alunos que participaram do teste é 150, essa proficiência estaria localizada no nível 2. Nesse caso, existem duas possibilidades para se trabalhar com os níveis. Uma, utilizá-los para ranquear os desempenhos; outra, compreender melhor as diferenças de complexidade existentes entre as diferentes habilidades testadas. A primeira não ajuda em nada o trabalho da escola e da professora, mas a segunda pode trazer informações importantes sobre as limitações existentes, que impedem o avanço na aprendizagem.

Determinar que a proficiência está associada a um número ou que ela se encontra em determinado nível da escala de proficiência de nada adianta. É preciso dar significado pedagógico ao número, pois, afinal, qual o sentido de se afirmar que a proficiência média dos alunos de uma escola encontra-se no nível 2? Ou então, qual a diferença pedagógica entre escolas que estão no nível 2 e o nível 3?

Para essa tarefa, primeiramente, é preciso identificar quais itens fornecem as proficiências que correspondem a determinado nível. Esses itens são analisados para encontrar os chamados itens âncora, aqueles que, depois da análise pedagógica, servirão para

dar significado pedagógico ao nível escolhido. Para ser âncora, os itens devem, simultaneamente, ter sido respondidos corretamente por pelo menos 65% daqueles que possuem proficiência superior e 50% daqueles que possuem proficiência inferior ao do nível em análise (VALLE, 2001).<sup>36</sup> Tomando como exemplo o nível 2, referente ao intervalo de proficiência entre 125 e 200, os itens que serão âncora para esse nível serão aqueles em que pelo menos 65% dos alunos com proficiência entre 200 e 275 respondem corretamente ao item e, ao mesmo tempo, no máximo 50% dos alunos com proficiência inferior a 125 que respondem ao item corretamente. Assim, alguns itens servirão para interpretar o nível e outros não poderão ser utilizados para essa tarefa, pois não teriam as propriedades necessárias. Como a proficiência faz parte de uma função probabilística, a interpretação da escala também deve se basear em probabilidades para ter alguma precisão. Portanto, nem todos os itens que participaram do teste serão utilizados na etapa seguinte.

Identificados os itens âncora para cada nível, eles são apresentados a um grupo de especialistas da área do conhecimento testado, para que esclareçam pedagogicamente o que cada item está exigindo dos alunos. Esse processo é conhecido como *descrição dos itens*. Depois de descrever todos os itens de um mesmo nível, os especialistas apresentam uma síntese capaz de descrever cada nível, baseada no que os itens têm em comum.

Como se parte do princípio de que o conhecimento é cumulativo, é muito baixa a probabilidade de um aluno, cuja proficiência esteja localizada no nível 2, responder a um item que exija uma proficiência localizada nos níveis superiores 3, 4 ou 5, pois o aluno ainda não desenvolveu as habilidades necessárias para tal feito. Da mesma forma, para um aluno cuja proficiência esteja localizada no nível 3, é muito alta a probabilidade de ele ter respondido corretamente os itens localizados nos níveis 1 e 2.

Assim, cada escola, dependendo de qual seja sua proficiência média e em que nível está posicionada, terá informações sobre as habilidades já desenvolvidas pelos alunos. Ao mesmo tempo, a escola obtém informações que permitem a possibilidade de estruturação de seu projeto pedagógico, permitindo, também, que as professoras revejam seus planos de curso, para que os alunos possam alcançar maiores aprendizagens.

Diante do exposto, é importante destacar que essa metodologia de interpretação da escala de proficiência é uma aproximação com o objetivo de dar significado pedagógico aos resultados obtidos no teste e, portanto, ela não oferece um resultado exato, mas sim um resultado válido para as condições que foram estipuladas. Além disso, como o item de múltipla escolha utilizado nos testes não consegue medir todas as habilidades desenvolvidas

---

<sup>36</sup> Klein (2003) sugere percentuais ligeiramente diferentes.

na escola, fazer inferências sobre a escola com base em um conjunto tão limitado de dados é no mínimo temerário. Também não se deve esquecer que nem todas as áreas de conhecimento são testadas, normalmente Língua Portuguesa e Matemática. Assim, o trabalho de interpretação dos resultados obtidos pela escola vale apenas para essas áreas. Por isso, a equipe pedagógica da escola tem de agregar outras informações, vindas não somente de outras áreas, que permitam fazer ajustes consistentes no projeto pedagógico da escola e nos planos de ensino das professoras das áreas que não foram testadas. Contudo, para que isso aconteça, deve haver uma equipe escolar comprometida com o sucesso de seus alunos, que receba apoio dos órgãos gestores para se desenvolver; que tenha certa estabilidade na escola, não mudando com muita frequência; e que existam canais democráticos para a livre troca de ideias. Além disso, a comunidade escolar também deve estar imbuída do mesmo desejo e deve ter espaço para colaborar com a escola. Como se percebe, se o que se busca são mudanças, não basta apenas a escola ser informada sobre sua proficiência média.

Outra importante limitação, nem sempre divulgada, é que a interpretação realizada utiliza apenas uma parte dos itens utilizados no teste, como já comentado antes.

Dependendo da ênfase que os sistemas educacionais dão à divulgação dos resultados dos testes avaliativos e da relação que se faz com a interpretação da escala de habilidades, pode-se obter um efeito contrário àquele que se esperaria de uma escola comprometida com uma aprendizagem significativa. Poderia acontecer, por exemplo, de a escola decidir que o melhor caminho para criar uma boa educação é centrar o ensino nas Matrizes de Referências e exercitar os alunos para responder a itens de múltipla escolha, baseados nas habilidades descritas nas Matrizes. Com isso, além da redução do currículo, deixa-se de estimular a capacidade criativa das pessoas, algo muito necessário para garantir tanto o desenvolvimento intelectual delas, como aumentar suas possibilidades de enfrentar com mais preparo as mudanças de uma sociedade em constante movimento.



## CAPÍTULO 2

### O GOVERNO FEDERAL E A ORGANIZAÇÃO DOS TESTES AVALIATIVOS

Neste capítulo são discutidos os testes avaliativos direcionados ao ensino fundamental, organizados pelo governo federal, o Sistema Nacional de Avaliação da Educação Básica (Saeb), a Prova Brasil e a Provinha Brasil, e também o Índice de Desenvolvimento da Educação Básica (Ideb), que vem sendo utilizado por esse nível de governo, como indicador de qualidade da educação básica.

Para facilitar a discussão, dividiu-se este capítulo em quatro tópicos. O primeiro tópico descreve a evolução dos testes avaliativos no Brasil, enfatizando o desenvolvimento da *expertise* nacional na área de avaliação educacional, fundamental para gerar os sofisticados instrumentos de coleta de dados, e como a legislação educacional vem dando suporte a esse processo de avaliação. Discute-se, também, a evolução do Saeb, dividindo esse processo em três fases bem demarcadas, acompanhando cada uma por meio dos relatórios técnicos publicados pelo Inep e identificando as mudanças introduzidas em cada ciclo e os possíveis impactos que elas tiveram. Também apresenta-se nesse tópico a Provinha Brasil.

Nos dois próximos tópicos, discutem-se, primeiro, os limites do Ideb como indicador de qualidade educacional e, a seguir, os programas federais, quais sejam, o Plano de Desenvolvimento da Educação (PDE) e um de seus componentes, o Pacto Nacional para a Alfabetização na Idade Certa (Pnaic), buscando compreender como os resultados dos testes avaliativos os influenciaram.

No último tópico apresentam-se dados de uma pesquisa realizada sobre como a qualidade educacional é discutida na mídia eletrônica escrita, tomando por base mais de dois mil textos produzidos entre 2009 e 2012.

#### ***A evolução dos testes avaliativos produzidos pela União***

A literatura nacional sobre as avaliações educacionais, como os trabalhos de Gatti (1991, 2002, 2009), Vianna (2005), Franco (2001), Bonamino (2002), Bauer (2012), Freitas (2007; 2013), Dirce Freitas (2007), Rodrigues (2007), Zanardini (2008), Coelho (2008), Souza e Oliveira (2007), Oliveira (2011), Horta Neto (2006), Alavarse e Bravo (2012) e Pestana (1992), vem tratando os diferentes aspectos do processo de desenvolvimento do Saeb,

que contribuíram para que a avaliação educacional atingisse o estágio que tem hoje. Alguns desses aspectos serão aprofundados, a seguir. Também serão discutidas as mudanças ocorridas no Saeb, à luz dos relatórios técnicos produzidos pelo Inep, e, por fim, a Provinha Brasil.

### **O desenvolvimento da *expertise* em avaliação educacional**

Entre os anos 1960 e 1980, uma série de acontecimentos no Brasil contribuiu para o desenvolvimento de *expertise* na área de avaliação educacional. O primeiro deles, sem dúvida, foi o surgimento dos vestibulares unificados nos anos 1960. Segundo Sigueta (1985), os dirigentes de alguns cursos da área médica do Estado de São Paulo, liderados pela Escola Paulista de Medicina, hoje Universidade Federal de São Paulo (Unifesp), decidiram unificar seus vestibulares em meados de 1963 e, para isso, criaram o Centro de Seleção de Candidatos às escolas Médicas e Biológicas (Cescem). Com isso, em vez de cada instituição isoladamente desenvolver seu instrumento próprio de seleção, uma equipe única desenvolveria um só instrumento, aplicado no mesmo dia e válido para várias instituições ao mesmo tempo. O primeiro vestibular unificado ocorreu em fevereiro de 1964, envolvendo sete instituições e os cursos de Medicina, Medicina Veterinária, Farmácia e Bioquímica e Biologia, ampliando-se nos anos seguintes o número de instituições participantes. O passo seguinte foi criar uma instituição que administrasse a organização e a aplicação desse vestibular, instituindo-se, então, em 25 de novembro de 1964, a Fundação Carlos Chagas (FCC). A seleção constava de duas etapas, sendo que, na etapa inicial, pela primeira vez, utilizaram-se testes de múltipla escolha com o objetivo de facilitar a correção dos testes.

Em seguida à criação do Cescem, unificam-se, também no Estado de São Paulo, os vestibulares das outras áreas. Em 1967, é criado o Centro de Seleção de Candidatos às Escolas de Economia e Administração (Cescea), que, inicialmente, reuniu as faculdades que ofereciam esses cursos na Pontifícia Universidade Católica de São Paulo (PUC-SP), na Universidade de São Paulo (USP) e na Fundação Armando Alvares Penteado (Faap), agrupando depois diversos cursos de Ciências Humanas. Assim como o Cescem, o Cescea também utilizava testes com itens de múltipla escolha. Em 1969, cria-se o Mapofei, sigla derivada das três instituições que ofereciam cursos de Engenharia e que resolveram unificar seus vestibulares: Escola Politécnica da USP, Escola de Engenharia Mauá e Escola de Engenharia da FEI. Também como ocorreu com os outros dois, com o passar do tempo, o

Mapofei acabou englobando diversas outras instituições que ofereciam cursos na área de Ciências Exatas. Diferentemente dos outros dois, o Mapofei utilizava um modelo de prova discursiva com questões abertas em todos os testes. Em 1971, cria-se no Rio de Janeiro o Centro de Seleção de Candidatos do Grande Rio (Cesgranrio), que também acabou unificando os diversos vestibulares naquele Estado. Nesse ano, inclusive, todos os jornais do País trouxeram uma imagem emblemática da unificação dos vestibulares no Rio: os candidatos fazendo os testes no estádio do Maracanã.

Esses vestibulares unificados alteraram a lógica anterior dos vestibulares, que exigiam uma nota mínima para o ingresso na faculdade, criando o que ficou conhecido como candidatos excedentes, aqueles que, apesar de terem atingido a nota mínima, não conseguiam matrícula, pois os aprovados excediam o número de vagas. A nova lógica era a da seleção de candidatos até o limite do número de vagas. O sucesso desse modelo e a utilização dos testes objetivos fizeram com que os itens de múltipla escolha passassem a ganhar destaque nos livros didáticos e começassem a ser amplamente utilizados pelos educadores (GATTI, 2002).

Para elaborar instrumentos cada vez mais sofisticados e confiáveis, capazes de dar conta do desafio de selecionar, entre milhares de candidatos, aqueles mais bem preparados para enfrentar a educação superior, foi necessário o desenvolvimento de equipes, com o objetivo de criar instrumentos de medida mais sofisticados e aprimorar a elaboração dos testes. Criavam-se, assim, as bases de conhecimento para as etapas seguintes.

Além dos grandes vestibulares, Gatti (2009) destaca que desde os anos 1960 houve alguns pequenos estudos financiados pelo Inep, que contribuiriam para o desenvolvimento da *expertise* na área de avaliação educacional, muitos deles ligados aos processos de seleção das universidades. A autora destaca, também, duas outras iniciativas como importantes para o desenvolvimento dessa *expertise*: a criação, em 1966, do Centro de Estudos de Testes e Pesquisas Psicológicas (CETPP), ligado à Fundação Getúlio Vargas (FGV) do Rio de Janeiro, que chegou inclusive a montar testes para as últimas séries do ensino médio nas áreas de Língua Portuguesa, Matemática, Ciências da Natureza e Ciências Humanas, e os estudos desenvolvidos, a partir de 1974, por iniciativa do Programa de Estudos Conjuntos de Integração Econômica Latino-Americana (Eciel), para testar o desempenho dos alunos das primeiras séries do ensino fundamental, abrangendo Língua Portuguesa (leitura e escrita) e Matemática.

Outra experiência importante ocorreu em 1980, na cidade de São Paulo, onde se desenvolveu uma pesquisa com o objetivo de verificar o desempenho dos alunos nas 1ª, 3ª, 5ª

e 7ª séries<sup>37</sup> do ensino fundamental e na 1ª série do ensino médio, envolvendo as áreas de Língua Portuguesa, Matemática e Ciências. O objetivo era utilizar os resultados no planejamento das atividades escolares e extraescolares, mas, como houve mudança de governo na capital paulista, a pesquisa não se realizou (GATTI, 2002).

Todas essas iniciativas contribuíram para que se criassem as bases do desenvolvimento das etapas seguintes da avaliação educacional no Brasil.

### **O surgimento da avaliação educacional na legislação**

A preocupação com a avaliação, seja em que área pública for, é recente no País. Tanto é assim, que, na Carta promulgada em outubro de 1988, apenas no inciso II do artigo 209, a avaliação é tratada, quando é determinado que a oferta de ensino pela iniciativa privada deve estar sujeita à avaliação de sua qualidade pelo poder público. Somente em 1998, dez anos depois e durante o primeiro governo de Fernando Henrique Cardoso, é que se incluem sete novas referências à avaliação na Emenda Constitucional nº 19, todas elas no âmbito da administração pública e relacionadas à reforma do Estado que se implantava, coordenada pelo então ministro Bresser Pereira. As referências eram relativas à participação do usuário da administração pública na avaliação da qualidade do serviço oferecido; à possibilidade de estabelecimento de contratos de gestão em que o cumprimento das cláusulas deveria ser objeto de avaliação permanente; a avaliação de desempenho como condição para se atingir a estabilidade no serviço público; a avaliação periódica das despesas com os serviços de saúde, nas três esferas de governo; e a avaliação de desempenho dos administradores das empresas públicas. Mesmo separados por dez anos, o sentido dado à avaliação nos dois períodos era o de verificar como determinado serviço público estava sendo oferecido.

O artigo 22, inciso XXIV, da Constituição determina que compete à União legislar sobre as diretrizes e as bases da educação nacional. O primeiro Projeto de Lei (PL 1.258), com esse objetivo, é encaminhado à Câmara pelo deputado Otávio Elísio, do PSDB-MG, em 29 de novembro de 1988, pouco mais de dois meses depois da promulgação da nova Carta. Na justificativa do Projeto (BRASIL, 1988), o deputado afirma que a proposta estava fundamentada no estudo elaborado pela Associação Nacional de Pós-Graduação e Pesquisa

---

<sup>37</sup> A Lei nº 11.274, de 6/2/2006, instituiu o ensino fundamental de nove anos, que deveria ser implantado pelos sistemas de ensino até 2010. Até então, o ensino fundamental era dividido em 8 séries e depois da lei passou a ser dividido em nove anos, deixando com cada sistema a liberdade de quando o fazer, dentro do limite determinado. Na rede estadual mineira, o ensino de nove anos foi implantado em 2004, portanto antes da lei, e na rede estadual paulista foi implantado em 2010. Este trabalho obedecerá à divisão adotada em cada um dos textos consultados.

(Anped), apresentado durante a XI Reunião Anual da instituição, realizada em Porto Alegre, de 25 a 29 desse ano (BRASIL, 1988). Segundo Oliveira (1997), o estudo foi elaborado pelo educador Demerval Saviani, com o título de "Contribuição à elaboração da nova LDB" e um início de conversa. No projeto, dividido em dez títulos e 83 artigos não é feita nenhuma menção ao tema avaliação.

O PL 1.258 foi alvo de intensa disputa e negociação, pois, além de modificações na forma como a educação escolar estava estruturada, mexia, entre outros pontos, na composição de poder na área educacional, criando um Sistema Nacional de Educação, responsável pela articulação das redes federal, estadual, municipal e particular de educação, que garantiria o direito à educação gratuita em todos os níveis e transformava o Conselho Federal de Educação em órgão normativo, de nível nacional, em que um terço de seus membros seria indicado pelas entidades representativas do magistério. O fato é que, depois de muitas emendas, um terceiro substitutivo, com 20 capítulos e 152 artigos, relatado pela deputada Ângela Amin, do PPR-SC, é aprovado em 13 de maio de 1993, quase cinco anos depois do começo de sua tramitação. No projeto aprovado na Câmara, o tema avaliação foi destacado em oito artigos, sendo que quatro deles relacionados à educação superior. Quanto aos outros quatro, um deles repetia o disposto no inciso II do artigo 209 da CF/1988, determinando que a oferta de ensino pela iniciativa privada deve estar sujeita à avaliação de sua qualidade pelo poder público, e outro, artigo 9º, inciso X, determinando que um dos princípios orientadores da organização da educação nacional é a valorização do processo de avaliação institucional (BRASIL, 1993, p. 7-8). O mesmo processo se repete no artigo 21, inciso III, que determina que as instituições de educação básica devem observar como diretriz a avaliação de desempenho institucional. Portanto, a ênfase dada pelo projeto não é nos testes avaliativos externos, mas sim em processos internos que contem com a participação da comunidade escolar, mas delegando ao Conselho Nacional de Educação, órgão que no projeto original recebeu o nome de Conselho Federal de Educação, a definição das diretrizes para a avaliação das instituições de ensino, conforme consta no artigo 23, inciso XIII do PL.

Não se sabe por quais motivos, talvez por conta do impasse que ocorria na Câmara, o senador Darcy Ribeiro, PDT-RJ, apresenta, em concorrida sessão do Senado Federal, do dia 21 de maio de 1992, outro projeto de lei para a Lei de Diretrizes e Bases da Educação nacional (LDB). Tratava-se do PL 67,<sup>38</sup> com dez títulos e 109 artigos, que, em sua

---

<sup>38</sup> O projeto, de autoria do senador Darcy Ribeiro, leva também a assinatura dos senadores Marco Maciel, do PFL/PE, e Maurício Corrêa, do PDT/DF, então líder do partido no Senado. Entre as pessoas a quem o senador agradece por tê-lo ajudado na elaboração do projeto, encontram-se o professor Candido Gomes, então consultor do Senado, e a professora da Universidade de São Paulo (USP), Eunice Ribeiro

justificativa, destaca como maiores preocupações o ensino fundamental e a meta de escolarização de tempo integral (BRASIL, 1992). O projeto dedica cinco artigos à avaliação, sendo que um deles reitera o disposto no inciso II do artigo 209 da Constituição, determinando que a oferta de ensino pela iniciativa privada deve estar sujeita à avaliação de sua qualidade pelo poder público, e outro, o art. 64, dedicado à educação superior. O artigo 23 trata da avaliação dos alunos na educação infantil, determinando que esta seja feita mediante acompanhamento e registro de seu desenvolvimento e proibindo a utilização de provas, exames, graus e menções. No título que destaca a qualidade do ensino, em seu artigo 63, § 1º, alínea b, determina que os sistemas de ensino estabeleçam a função de supervisão dessa qualidade, mediante a avaliação educacional. Nesse ponto, os dois projetos, o da Câmara, que estava prestes a ser aprovado, e o do Senado apresentam proposições diferentes para a avaliação das instituições escolares, no primeiro caso, centrado em um processo interno, com suas diretrizes fixadas pelo Conselho Nacional de Educação, e no segundo, com características de supervisão em relação à qualidade do ensino ofertado.

Mas o grande divisor entre as duas propostas, em matéria de avaliação, aparece no artigo 64, que determina que a União, em colaboração com os sistemas de ensino, deve manter processo nacional de avaliação do rendimento escolar em âmbito nacional. O § 1º desse artigo ressalta que essa avaliação deve orientar a política educacional, e o § 2º determina que a assistência financeira da União aos sistemas de ensino fica condicionada à sua ativa colaboração na implantação e no desenvolvimento da avaliação. Dessa forma, sem que haja obrigatoriedade formal, o dispositivo praticamente obriga os outros Entes da Federação a participarem do processo, já que a assistência financeira da União sempre foi significativa para grande parte dos sistemas de ensino.

Durante os debates que antecederam à leitura do projeto no plenário do Senado, o senador Darcy Ribeiro fez longo discurso, do qual alguns trechos são destacados a seguir. Sobre a importância de seu projeto, Darcy Ribeiro afirma que

[...] toda legislação vigente e também a programada é inclusive pelo Projeto de Lei de Diretrizes e Bases, que se discute na Câmara dos Deputados é só tendem a consolidar esse sistema educacional responsável pela produção em massa de analfabetos, por sua incapacidade de alfabetizar as crianças brasileiras (BRASIL, 1992, p. 3702).

---

Durham, na época Secretária Nacional de Educação Superior e que seria depois Secretária Nacional de Política Educacional, entre 1995 e 1997, durante a gestão do ministro Paulo Renato Souza.

Portanto, ele se contrapõe ao projeto da Câmara, que, segundo ele, não resolverá os problemas educacionais do País.

Com relação ao fracasso da escola, o senador destaca em diversos pontos de seu discurso que

[...] Isso se dá em razão da escandalosa inadequação entre a escola brasileira e seu alunado majoritário. Ela funciona como se sua tarefa fosse alfabetizar as crianças das camadas favorecidas, porque só estas progredem no estudo. As crianças que vêm de famílias que não tiveram escolaridade prévia fracassam. Não por culpa ou incapacidade delas próprias, mas em razão da hostilidade real, ainda que inconfessada, da nossa escola pública a seu alunado pobre ou negro, pior ainda se negro e pobre, como ocorre com tantíssimos brasileiros.

[...] A rede educacional brasileira, forçada a ampliar enormemente suas matrículas para atender ao crescimento vertiginoso das populações urbanas ó que, nas últimas décadas, saltaram de menos de trinta para mais de setenta por cento da população ó , em lugar de multiplicar o número de escolas, a desdobrou em turnos ó dois, quatro e até cinco, diariamente ó , mesmo no Estado de São Paulo. Isto implicou em dar uma atenção cada vez mais reduzida a seus alunos, até o ponto em que só pode progredir nos estudos quem tenha em casa uma outra escola.

[...] Somadas as carências de espaço, de escolas e de tempo docente com a ineficácia didática do magistério, o desprestígio da profissão e o peso dos preconceitos, o sistema nacional de educação se constitui como uma entidade aberrante (BRASIL, 1992, p. 3702-3703).

Difícilmente qualquer estudioso dos problemas educacionais brasileiros poderia discordar da análise dos problemas da escola pública feita por Darcy Ribeiro. Em seu discurso de mais de vinte anos atrás, o autor destacava que o principal problema da escola estava ligado à pobreza e à incapacidade da escola e de suas professoras em lidar com ela, acrescentando a isso políticas educacionais irresponsáveis e problemas na formação inicial das professoras e apresentando um quadro alarmante para aquela época, mas que ainda hoje faz parte da realidade atual.

O problema é que a forma encontrada pelo senador para colocar suas propostas em debate não levou em consideração a discussão que vinha sendo construída desde os primeiros embates na Constituinte, consubstanciada no PL aprovado na Câmara. Mesmo que a proposta apresentada pelo senador contivesse pontos interessantes e inovadores, a forma encontrada de colocá-los em pauta primou pelo fortalecimento do personalismo, em detrimento da construção coletiva.

O ponto que merece destaque no discurso de Darcy Ribeiro, dado o tema do presente trabalho, refere-se às razões que o levaram a incluir a avaliação externa em seu projeto. Segundo o autor, a avaliação externa seria

[...] indispensável a qualquer sistema de ensino.

Nesse momento, o Presidente dos Estados Unidos determinou a aplicação, em todo o sistema escolar de primeiro e segundo graus dos Estados Unidos, de uma prova de estado. Isso parece uma violência, sobretudo num País federativo em que os Estados são tão ciosos de sua independência, mas o Presidente George Bush considerou indispensável aplicar essa medida violenta como única forma de poder competir com o adiantamento muito maior do ensino no Japão e na Alemanha. Se isso é problema para os Estados Unidos, para poderem exercer-se plenamente na futura civilização, para nós o é em escala muitíssimo maior e muitíssimo mais grave (BRASIL, 1992, p. 3705).

Nesse trecho, provavelmente o senador se referia à iniciativa do presidente George Bush de pactuar metas nacionais com os governadores, acompanhadas pelo Painel Nacional de Metas Educacionais, já comentado antes neste texto, e seu desejo de enviar ao Congresso uma lei solicitando autorização para testar os alunos nacionalmente, algo que foi efetivamente feito por Bill Clinton em 1997, cinco anos depois do pronunciamento de Darcy Ribeiro. É surpreendente que um intelectual com larga tradição no pensamento de esquerda tenha ficado seduzido por argumentos como a de que a avaliação seria o remédio para reverter a queda de desempenho dos alunos americanos frente ao de outros países e que, em nosso caso, conseguiria reverter os baixos indicadores educacionais. Note-se que, mesmo considerando uma violência ao preceito da Federação, Darcy acredita que o resultado que se poderia obter com a avaliação a justificaria, mesmo que para isso fosse necessário cortar a assistência financeira da União aos Entes federados que não quisessem dela participar. É a velha crença de que na educação existe a õbala de prataõ capaz de liquidar todos os problemas, ou pelo menos eliminar os mais graves.

Essa contradição entre o pensamento de um homem de esquerda que consegue claramente identificar os problemas da escola, relacionando as dificuldades de aprendizagem com a pobreza da população, é o mesmo que usa uma das principais bandeiras do neoliberalismo para controlar uma pretensa qualidade da escola que se almeja, ainda que com o argumento de orientar a política educacional. Os testes avaliativos, gradativamente, com o passar dos anos, e com a sucessão de partidos no governo federal, transformam-se em um



fetichismo que acaba seduzindo as diversas administrações federais desde 1988, com diferentes perfis ideológicos.

Voltando à tramitação do PL nº 67, de 1992, a proposição foi encaminhada à Comissão de Educação do Senado, tendo sido aprovada, com alterações, pela Comissão, em fevereiro de 1993, menos de um ano depois de o projeto ter sido apresentado no Senado. Segundo Oliveira (1997), o projeto deixa de ser votado em Plenário por interferência do Fórum Nacional em Defesa da Escola Pública na LDB<sup>39</sup> e de alguns senadores.

Quando o PL nº 1.258, aprovado na Câmara, chega no Senado, em maio de 1993, recebe o título de PL nº 101. Na Comissão de Educação, o senador Cid Sabóia, do PMDB-CE, apresenta um substitutivo a ele, retomando alguns pontos do projeto original apresentado em 1988, e esse substitutivo é aprovado no final de 1994. Chega a ir a Plenário para votação, em 30 de janeiro de 1995, no último dia da legislatura, mas a votação não ocorre por falta de quórum.

Na nova legislatura, que se inicia em 1995, por meio de um recurso regimental, o substitutivo do senador Cid Sabóia é encaminhado novamente para a Comissão de Constituição, Justiça e Cidadania para apreciação. Nessa Comissão, o relator da matéria, senador Darcy Ribeiro, rejeita o substitutivo do senador Cid Sabóia e apresenta um novo substitutivo, retomando parte das ideias de seu Projeto. Depois de retornar à Comissão de Educação e receber novas emendas, o Projeto da LDB é aprovado em fevereiro de 1996. Segue novamente para a Câmara, onde é discutido por mais dez meses, e é votado em dezembro de 1996.

A LDB, contendo 9 títulos e 92 artigos, faz 22 menções ao termo avaliação. Reafirma o inciso II do artigo 209 da Constituição, determinando que a oferta de ensino pela iniciativa privada deve estar sujeita à avaliação de sua qualidade pelo poder público. Com relação à proposta inicial de Darcy Ribeiro sobre a avaliação, ela é ampliada e ao mesmo tempo se torna menos impositiva:

Art. 9º A União incumbir-se-á de:

[...]

VI ó assegurar processo nacional de avaliação do rendimento escolar no ensino fundamental, médio e superior, em colaboração com os sistemas de ensino, objetivando a definição de prioridades e a melhoria da qualidade do ensino.

[...]

---

<sup>39</sup> O Fórum congregava 26 entidades científicas, sindicais, estudantis, de dirigentes educacionais e de classe e dela participavam: SBPC, Anpae, Anped, Cedes, Anfope, CUT, CGT, Crub, Contag, Andes Ande, Fasubra, CBCE, Conam, CNRCFE, Fenaj, Fenase, Fenoe, Ubes, UNE, ANPG, Consed, Undime e OAB.

VIII ó assegurar processo nacional de avaliação das instituições de educação superior, com a cooperação dos sistemas que tiverem responsabilidade sobre este nível de ensino (BRASIL, 1996).

Assim, fica claro, pela concepção apresentada na lei, que a avaliação educacional surge com os objetivos de definir prioridades para as políticas educacionais e auxiliar na busca da qualidade do ensino, longe portanto de estabelecer metas e ranqueamento de escolas, determinando, também, que o processo ocorra em colaboração com os sistemas de ensino. No entanto, a expressão dessa concepção de forma tão genérica permitiu justificar outras práticas que dessem sustentação a políticas baseadas no *accountability*.

O artigo 9º também dava sustentação ao Saeb, que, desde 1995, além de testar os alunos do ensino fundamental, começou a testar também os do ensino médio. O artigo ainda dava apoio ao Provão, teste de desempenho dos estudantes do ensino superior, elaborado pelo Inep, que começou a ser aplicado em 1996. Por fim, propunha a avaliação da educação superior, que também começa a ser estruturada pelo MEC em 1996. Portanto, ao mesmo tempo que o governo desenvolvia uma série de programas ou realizava modificação nos já existentes, discutiam-se no Congresso formas de institucionalizá-los por meio de alterações na da LDB.

Com relação à avaliação externa, um avanço com relação ao projeto original de Darcy Ribeiro foi a eliminação da determinação de que Estados, DF e municípios que não colaborassem com a avaliação externa não teriam acesso à assistência financeira da União, uma obrigação disfarçada. Mas essa obrigatoriedade da participação nos testes nacionais retorna por meio da Lei nº 11.330, de 2006, durante o primeiro governo de Luiz Inácio Lula da Silva, que, entre outras coisas, altera o artigo 87 da LDB, em seu inciso IV do § 3º, determinando que os Entes da Federação integrem todos os estabelecimentos de ensino fundamental de seu território ao sistema nacional de avaliação do rendimento escolar. Com isso, obteve-se a sustentação legal para que o teste avaliativo pudesse ser aplicado em todas as escolas, algo que passou a ser realizado a partir de 2005 com a Prova Brasil.

As outras menções ao tema avaliação na LDB de 1996 envolvem as avaliações dos alunos pela escola e os processos de acreditação de cursos de educação superior presencial e a distância.

## Da criação do Saeb até a Prova Brasil

Pode-se classificar o Saeb pelas três fases que o modificam substancialmente<sup>40</sup>. A primeira fase, a de sua implantação, entre 1988 e 1993, começa, com a aplicação piloto de 1988, e envolve as aplicações que ocorrem nos próximos dois ciclos, de 1991 e 1993. A segunda fase, entre 1995 e 2003, envolvendo o começo do uso da TRI, no ciclo de 1995; a construção das Matrizes de Referências, no ciclo de 1997 e nos três próximos ciclos entre 1999 e 2003. A terceira fase se inicia no ciclo de 2005 com a transformação do teste amostral em censitário.

O que chama a atenção nesse processo é que não houve descontinuidade no desenvolvimento e no amadurecimento do Saeb, processo esse iniciado na época do regime militar, durante o governo de João Figueiredo (1979-1985), e que atravessou os governos José Sarney (1985-1990), Fernando Collor (1990-1992), Itamar Franco (1992-1995), os dois governos Fernando Henrique Cardoso (1995-2003), os dois governos Luiz Inácio Lula da Silva (2003-2011) e o atual governo Dilma Rousseff. Foram governos com diferentes orientações políticas, mas em todos privilegiou-se a realização de medidas de desempenho educacional como uma das estratégias para aprimorar a qualidade da educação. O que foi mudando ao longo do tempo foram a abrangência da avaliação e os usos de seus resultados.

### *Primeira fase: de 1988 a 1993*

As primeiras ações no sentido da criação do Saeb partem do Programa de Educação Básica para o Nordeste Brasileiro (Edurural), implantando em 1980, com financiamento do MEC e do Banco Mundial. Esse programa tinha como objetivos expandir o acesso às quatro primeiras séries do ensino, diminuir as taxas de repetência e evasão e melhorar o rendimento escolar dos alunos. Para acompanhar a evolução do último objetivo, foram realizados três estudos em 1981, 1983 e 1985, em 60 municípios de três Estados, envolvendo 600 escolas e 6 mil alunos, para medir o desempenho escolar das então 2ª e 4ª série da escola primária em Português (leitura e interpretação, escrita e gramática) e Matemática (HORTA NETO, 2006), a cargo da equipe da FCC, liderada pelo professor Heraldo Vianna e da qual fazia parte, entre outros pesquisadores, a professora Bernadete Gatti. Testava-se para acompanhar os resultados

---

<sup>40</sup> Bonamino e Souza (2012), propõem uma outra divisão através dos objetivos dos testes avaliativos, quais sejam, diagnóstica, nos primeiros anos do Saeb, responsabilização atrelada a mecanismos de consequências fracas, como a Prova Brasil e, por último, responsabilização atrelada a mecanismos de consequências fortes, como o Saresp a partir de 2008.

do programa que estava sendo implementado, utilizando-se os grupos que dele faziam parte e outros de controle, não havendo interesse em comparar e classificar as escolas. Segundo Gatti (1994), durante os cinco anos do programa não se observou melhoria no desempenho dos alunos, apesar dos vultosos recursos dispendidos em construções e reformas de escolas, treinamentos e material didático. A autora, com base nos resultados dos estudos, apresenta as razões para isso:

A rotatividade dos professores nas escolas, cuja permanência apenas dependia da boa vontade e interesse dos prefeitos, sendo inconstantes as condições desse atendimento. A complementação de salário prevista para os professores nem sempre era providenciada a tempo pela prefeitura e mesmo o salário regular dessa professora (muito baixo, em geral 1/3 do salário mínimo) nem sempre era pago. As condições de ensino multisseriado,<sup>41</sup> característico destas escolas, permaneceram precárias: muitos alunos em classe, infraestrutura deficiente, falta de condições do professor para o manejo de várias séries ao mesmo tempo. O tempo dedicado ao ensino propriamente dito era muito pequeno ó duas a duas horas e meia por dia. Frequência irregular dos alunos em função do trabalho, doenças das crianças ou na sua família, condições difíceis para chegar à escola. Prevalência, na escola, da concepção de que aprender significava memorizar e, portanto, preponderava como método de ensino a repetição sem significado. Não incentivo à manifestação autônoma da criança; passividade induzida ao aluno. Os professores, presos às cartilhas ou livro didático (aquele disponível, em geral um só e não de sua escolha), não se utilizavam quer das experiências das crianças, quer dos espaços disponíveis no entorno da escola para desenvolver seu trabalho didático, não tendo sido nem incentivada, nem orientada para isto pelo sistema de supervisão pedagógica que se montou (GATTI, 1994, p. 68-69).

Apesar dos avanços sociais obtidos quase trinta anos depois das condições descritas pela autora, muitas das condições que impactaram a aprendizagem naquela época continuam presentes. Assim, nos dias de hoje, por maior que sejam as pressões por melhores resultados nos testes, existe um limite para o que é possível ser feito.

Com a experiência do Edurural e com a experiência adquirida na realização dos grandes vestibulares, a professora Bernadete Gatti<sup>42</sup> lembra que ela e o professor Heraldo

---

<sup>41</sup> O ensino multisseriado, típico de comunidades rurais localizadas em regiões de baixa densidade populacional, distantes dos centros urbanos e que podem ter, ou não, carência de professoras, caracteriza-se pelo ensino simultâneo, em uma mesma sala de aula, para alunos de diferentes anos escolares.

<sup>42</sup> Palestra realizada durante o Ciclo de Debates: Vinte e Cinco Anos de Avaliação de Sistemas Educacionais no Brasil, organizado pela Fundação Carlos Chagas (FCC), em São Paulo, SP, entre os dias 29 e 30 de março de 2012.

Vianna foram convidados, no começo de 1987, para discutir com Aloísio Sotero, então ministro da Educação durante o governo do general João Figueiredo, a possibilidade de realizar um diagnóstico sobre como estava a educação, de forma a se ter elementos para influir nos processos desenvolvidos por Estados e municípios. Mesmo depois da indicação do novo titular da pasta, esses estudos preliminares tiveram prosseguimento e deram origem a três pesquisas realizadas em 1988: uma, envolvendo 4.518 crianças em 10 cidades (VIANNA; GATTI, 1988); outra, nesse mesmo ano, que testou 8.069 alunos em 20 cidades (VIANNA, 1989); e uma terceira, com 14.868 crianças em 39 cidades (VIANNA, 1990). Para viabilizá-las, devido a problemas orçamentários do MEC, utilizaram-se recursos de um convênio existente entre o Instituto Interamericano de Cooperação para a Agricultura (IICA) e o Ministério.

Vianna (1990) apresenta quais seriam os objetivos da terceira pesquisa:

- Identificar, na diversidade do quadro educacional brasileiro, os pontos curriculares mais críticos que deveriam merecer consideração de professores e administradores, a fim de solucionar problemas de aprendizagem que criam dificuldades ao longo do processo educacional;
- Verificar o desempenho dos alunos de 1ª, 3ª, 5ª e 7ª séries em determinados aspectos cognitivos que seriam, na opinião de professores militantes, fundamentais à aprendizagem no currículo da escola de 1º grau<sup>43</sup>;
- Fornecer subsídios para os professores atuarem no seu dia a dia, recuperando alunos ao longo do processo em suas possíveis deficiências (p. 3).

Além disso, o autor destaca a importância da avaliação externa, ao afirmar que:

As avaliações [que ocorrem na escola] são de natureza somativa e bimestrais e visam ao atendimento de uma contabilidade imposta pela burocracia educacional. Dessa forma, faz sentido um programa de avaliação externa que forneça subsídios para as avaliações do professor, complementando informações nem sempre precisas sobre o desempenho de seus alunos. Além do mais, uma avaliação externa pode proporcionar elementos aos especialistas para a reformulação da estrutura e dos conteúdos curriculares. A experiência diária dos professores e as experiências externas a ela associadas complementam-se e contribuem para um conhecimento mais fundamentado sobre o desempenho dos alunos (VIANNA, 1990, p. 17).

---

<sup>43</sup> Antes da aprovação da LDB de 1996, os anos escolares eram divididos em 1ª a 9ª séries do primeiro grau e 1ª a 3ª séries do segundo grau, conforme determinava a Lei 5.692/71. Com a aprovação da nova LDB, passou-se a se dividir em 1ª a 8ª série do ensino fundamental e 1ª a 3ª série do ensino médio.

Percebe-se nesse trecho a expectativa de que as avaliações externas fornecessem subsídios às professoras para aprimorar a avaliação de seus alunos, agregando novas informações às avaliações internas realizadas normalmente pelas escolas. Além disso as avaliações indicavam a possibilidade de seus resultados influenciarem os currículos. Os aspectos citados pelo autor reforçam a preocupação pedagógica que havia nesses momentos iniciais da avaliação educacional.

Para essa terceira pesquisa, partiu-se de uma amostra de escolas públicas, estaduais e municipais, elaborada por Philip Fletcher, do Instituto de Pesquisa Econômica Aplicada (Ipea), em que foram testados alunos das 1ª, 3ª, 5ª e 7ª séries da educação básica. A escolha desses anos escolares se deu porque õesses grupos escolares ocorrem com maior intensidade os problemas que afligem a escola de 1º grau, especialmente a repetênciao (VIANNA, 1990, p. 9). Para as 1ª e 3ª séries, foram usados testes de Português (interpretação de texto e gramática) e Matemática, cada um com 30 itens semiobjetivos, em que o aluno escrevia uma palavra, copiava uma frase, reescrevia frases, ordenava palavras, entre outras atividades. Nesses anos, o aplicador dos testes, contratado pela FCC, lia os itens antes de o aluno dar sua resposta. Para os anos escolares seguintes, foram utilizados testes de Português, Matemática e Ciências, cada um com 30 itens de múltipla escolha, com quatro alternativas e uma redação de, no mínimo, 15 linhas. Segundo o autor, objetivando não cansar os alunos e não afetar a rotina escolar, cada aluno respondeu a apenas um teste (ou Português, ou Matemática, ou Ciências), com exceção dos alunos das 5ª e 7ª séries que participaram também do teste de redação. Além dos testes, os diretores e professoras responderam a questionários de 50 e 53 itens, respectivamente.

Para referenciar a construção dos itens, foi realizada uma pesquisa com as professoras, sobre os conteúdos trabalhados em sala de aula, e outra sobre livros didáticos, buscando determinar quais seriam os conhecimentos mínimos esperados para cada uma das séries a serem testadas. A partir daí, construíram-se os itens para as 1ª e 3ª séries. Para as 5ª e 7ª séries, usaram-se itens calibrados pertencentes ao Banco de Itens (Banitem) da FCC. Infelizmente, dos documentos a que a pesquisa teve acesso não constavam as Matrizes de Referência utilizadas.

Os resultados foram analisados a partir da Teoria Clássica dos Testes, pois o objetivo era fazer com que os resultados fossem compreensíveis para as professoras, apesar de se saber da importância que a TRI adquiria ao tornar os testes independentes das características dos alunos (VIANNA, 1990).

Para garantir a fidedignidade dos resultados, reuniu-se uma equipe de professoras para, durante dois dias, analisar os testes utilizados e verificar se as premissas levantadas nas pesquisas estavam sendo cumpridas.

Os resultados obtidos nos testes, expressos em termos de quantidade de itens corretos, entre os 30 apresentados aos alunos, podem ser observados no Quadro 2.1.

**Quadro 2.1 – Média, desvio-padrão, fidedignidade e discriminação média das provas de Português, Matemática e Ciências, aplicadas a uma amostra de alunos de 1ª, 3ª, 5ª e 7ª séries do 1º grau da rede oficial de ensino, em 15 unidades da Federação – 1988**

Série	Matéria	Média	Desvio-padrão	Discriminação média
1ª	Português	16,29	7,81	0,66
	Matemática	11,28	6,25	0,51
3ª	Português	17,43	5,95	0,50
	Matemática	11,56	5,51	0,45
5ª	Português	16,23	4,15	0,34
	Matemática	9,39	2,81	0,22
	Ciências	12,64	3,39	0,28
7ª	Português	15,43	3,91	0,33
	Matemática	9,10	3,80	0,30
	Ciências	13,10	3,96	0,33

Fonte: VIANNA, 1990, p. 72.

O quadro mostra que, em termos das três áreas testadas, houve melhor desempenho em Português, depois em Ciências e, muito abaixo dos dois, em Matemática.

Como a pesquisa havia sido planejada para indicar pontos críticos da aprendizagem e não identificar inter-relações de variáveis que poderiam explicar este ou aquele problema (VIANNA, 1990, p. 23), não foi realizado nenhum estudo relacionando o desempenho dos alunos e os fatores que poderiam estar afetando esse desempenho. Segundo o autor, buscava-se determinar quais os problemas que os alunos de 1ª, 3ª, 5ª e 7ª séries estariam tendo em relação aos programas de Português, Matemática e Ciências, em escolas da rede oficial, após uma sequência instrucional, com base no desempenho em provas de escolaridade (VIANNA, 1990, p. 23).

Vianna (1990) destaca que as provas de 1ª e 3ª séries e as redações foram avaliadas individualmente por professoras do antigo 1º grau, que, em seguida, apresentaram relatórios avaliativos. São apontados pelo autor alguns aspectos destacados por essas professoras e, a título de exemplo, reproduzem-se aqui alguns deles. Na página 78, o autor, referindo-se à 1ª

série, observou que após um ano de escolaridade, um número expressivo de alunos não escrevia o próprio nome, sendo necessário o auxílio da professora, enfatizando que os vários problemas surgidos não se limitavam a uma única área geográfica [mas ocorriam] também em centros culturalmente mais desenvolvidos, onde supostamente haveria um nível de ensino mais elevado e que os sinais nas operações nem sempre eram atendidos, havendo uma tendência à realização da adição, independentemente do que estivesse indicado. Para a 3ª série, o autor diz que as falhas do processo de alfabetização impediram que um número considerável de crianças entendesse as questões relativas aos textos (VIANNA, 1990, p. 82) e que houve grandes dificuldades para operar com números de dois algarismos (VIANNA, 1990, p. 87). Nas provas de redação para a 5ª série, observou-se que alguns alunos apresentavam graves problemas de alfabetização (VIANNA, 1990, p. 89) e, de modo geral, falta de coordenação das ideias, dificuldade na estruturação da composição e na manutenção da sequência expositiva (VIANNA, 1990, p. 89). Ainda segundo Vianna, algumas poucas redações da 7ª série foram bem escritas e possuíam certa originalidade, ao lado de outras pobres e infantis. Nessa atividade, a fuga ao tema foi comum e houve casos de alunos que não conseguiram escrever com um mínimo de clareza (VIANNA, 1990, p. 98). O relatório faz também uma descrição detalhada dos erros mais comuns cometidos pelos alunos.

No final do relatório, Vianna (1990, p. 202-217) relaciona diversos aspectos do que ele chamou de lições aprendidas. Apesar de o estudo ter sido realizado em 1988, é de uma atualidade marcante, demonstrando que muito pouco parece ter sido realizado para melhorar a escola pública em mais de vinte e cinco anos.

Segundo o autor, os resultados devem ser tomados como referência para se compreender melhor os processos que ocorrem na escola e as dificuldades que os alunos apresentam, para assim serem desenvolvidas estratégias pedagógicas para superá-los. Entre os diversos pontos abordados no texto, destacam-se: as escolas de 1º grau eram frequentadas em sua maioria pelos pobres; não possuíam condições materiais adequadas, com carência de material didático e de recursos humanos (professoras e pessoal de apoio) qualificados; tinham necessidade de capacitar o corpo docente em termos de conteúdo, metodologias e estratégias para uma ação pedagógica mais efetiva; estavam situadas em locais de difícil acesso, sem instalações e mobiliários adequados e com deficiência de saneamento básico; clima de violência rondando a escola e a invadindo; falta de participação dos pais na vida escolar; ausência de assistência técnica dos vários órgãos gestores às escolas e assistência pedagógica às professoras e alunos; frequente ausência de professoras sem a sua substituição e constantes mudanças no quadro de professoras ao longo do ano; evasão de professoras e de pessoal



qualificado por falta de uma política salarial adequada, provocando o esvaziamento da escola; superlotação das classes; ensino baseado no apelo à memorização; marginalização dos alunos mais fracos pelos professores; deficiências no processo de alfabetização, com repercussões ao longo dos anos escolares seguintes; deficiência do sistema de avaliação utilizado pelas professoras, que age como fator de desestímulo junto ao aluno.

O texto relaciona, ainda, vários problemas detectados com relação ao desempenho dos alunos nas três áreas testadas, entre eles: problemas com o domínio da língua, com deficiências tanto na compreensão do texto como na escrita; em matemática, falta de domínio de conhecimentos básicos, como o de unidade, dezena e centena, além de falta de conhecimento das técnicas operatórias nas quatro operações fundamentais; com relação à redação, as diferenças entre os trabalhos de 5ª e 7ª séries foram mínimas.

Mesmo que nos anos seguintes tenham sido superados muitos dos problemas dessa época, como o grande número de professoras leigas, a taxa de matrícula das crianças em idade para frequentá-la inferior a 90%, a inexistência do Fundef e do Fundeb,<sup>44</sup> para reorientar o financiamento entre as redes dentro de um mesmo Estado e para injetar recursos federais na educação básica, os problemas com a aprendizagem pareciam continuarem.

## **Ciclo de 1990**

Em 1990, cria-se o Sistema de Avaliação do Ensino Público de 1º Grau (Saep), concebido como um projeto integrador e cooperativo entre a União e os Estados, que com seus resultados permitia a orientação das políticas educacionais dos níveis de governo federal e estaduais (PILATTI, 1995).

Como uma pesquisa, buscava-se responder a três questões (INEP, 1992):

- Em que medida a gestão educacional tornou-se mais eficiente e democrática?

---

<sup>44</sup> O Fundo de Manutenção e Desenvolvimento do Ensino Fundamental e de Valorização do Magistério (Fundef) foi criado em 1998, como um fundo setorial destinado a financiar em parte o ensino fundamental público, uma das etapas da educação básica. Foi substituído em 2006 pelo Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb), financiando todas as etapas e modalidades da educação básica. Constitucionalmente, Estados e municípios devem destinar 25% dos tributos arrecadados e das transferências para financiar a educação. O Fundo, regulamentado pela Lei nº 11.494/2007, tem natureza contábil e retine, no âmbito de cada Estado, recursos dos governos estaduais e municipais envolvendo, dos 25% obrigatórios, uma parcela de 80%, dos principais tributos (ICMS, IPI exportação, IPVA, ITCMD e ITR) e das transferências envolvendo os Fundos de Participação dos Estados e dos Municípios. Esses recursos são redistribuídos aos governos estaduais e municipais, tendo por base o número de alunos matriculados em cada rede. Se dentro de determinados Estados, a razão entre o total de recursos do Fundo e o total de alunos matriculados nas duas redes for inferior a um determinado valor, a União adiciona recursos ao Fundo desses Estados. Estima-se que, em 2013, no total, o Fundeb movimentou R\$ 104 bilhões/ano, cabendo à União 10% desse valor. Por ser um fundo de caráter transitório, com o objetivo de incentivar a melhoria educacional, o Fundeb tem vigência até 2020.

- Quais as condições de trabalho e a competência pedagógica da professora?
- Em que medida as políticas adotadas estão possibilitando o acesso das crianças e a melhoria da qualidade do ensino ministrado?

É interessante notar, comparando com os dias de hoje, em que o ranqueamento ganha tamanha importância e assume ares de naturalidade, que as preocupações nesses anos iniciais da avaliação no Brasil eram voltadas para obter informações sobre o sistema educacional em que as escolas estavam inseridas.

Estabeleceu-se que seriam realizadas medidas a cada dois anos, provavelmente por causa do tempo necessário para a preparação, a aplicação e o processamento dos dados. Com base na metodologia dos três estudos referenciados anteriormente, estabeleceu-se aquela que seria utilizada pelo Saep. Da primeira aplicação, em 1990, participaram quase todos os Estados (devido a razões desconhecidas, o Piauí e Mato Grosso do Sul não participaram) e, com base na amostra estabelecida, testaram-se 108.875 alunos. Para as 1ª e 3ª séries, foram aplicados testes de Português e Matemática envolvendo 30 itens semiobjetivos e para as 5ª e 7ª séries, testes de Português, inclusive redação, Matemática e Ciências, envolvendo 30 itens de múltipla escolha com quatro alternativas. Cada aluno respondia a apenas um teste, com exceção daqueles das 5ª e 7ª séries que faziam também a redação. Em 1991, ainda na fase de análise de dados, o sistema ganha seu nome atual, Sistema Nacional de Avaliação da Educação Básica (Saeb).

Enquanto o relatório do estudo que antecede o Saeb, coordenado por Vianna (1990), é mais descritivo procurando apontar, com exemplos, os possíveis problemas ocorridos com a aprendizagem dos alunos, o relatório do Inep (1992) é mais geral, pois mostra apenas a metodologia utilizada e os resultados numéricos obtidos, indicando a média das notas obtidas por regiões e redes de ensino. Começa, pois, a transição de um instrumento com visível preocupação com a aprendizagem dos alunos e com os processos da escola, ou seja, com o processo educacional, para outro focado nos resultados. Em suas conclusões, o relatório do Inep mostra que o desempenho geral não se mostrou tão mal quanto em geral se supunha, pois as linhas médias de domínio de conteúdos se situam num patamar que não se pode considerar tão baixo, apesar de distanciarem-se do mínimo desejado e de contarem diferenciais muito significativos (INEP, 1992, p. 95). O relatório aponta também para um sério problema, ao afirmar que os sistemas de ensino mostram em geral níveis declinantes de rendimento em todas as matérias examinadas, à medida que avança a seriação (INEP, 1992, p.

95). Esses tipos de informações mais gerais adotadas no relatório produzido pelo Inep passam a servir de modelo para a produção dos futuros relatórios.

A expressão dos resultados utilizou uma escala de 0 a 100, em que 100 significava que os conteúdos mínimos foram apreendidos e que todos os itens testados foram marcados corretamente. Quanto mais distante desse valor, mais distante desses conteúdos estariam os alunos. Esses resultados podem ser observados no Quadro 2.2.

**Quadro 2.2 – Média das provas de Português, Matemática e Ciências, aplicadas a uma amostra de alunos de 1ª, 3ª, 5ª e 7ª séries do 1º grau da rede oficial de ensino, em 25 unidades da Federação – Saeb 1990**

Série	Matéria	Média
1ª	Português	56,8
	Matemática	52,8
3ª	Português	62,9
	Matemática	50,2
5ª	Português	51,9
	Redação	49,5
	Matemática	31,6
	Ciências	41,7
7ª	Português	48,6
	Redação	58,4
	Matemática	30,0
	Ciências	43,0

Fonte: INEP, 1992, com elaboração do autor

Apesar de não ser possível comparar os Quadros 2.1 e 2.2, já que o Quadro 2.1 mostra a média de escores e o Quadro 2.2 a média de acertos expressa em uma escala de 0 a 100, e pelo fato de a metodologia utilizada, TCT, não permitir comparar testes diferentes realizados por populações distintas, observam-se algumas tendências comuns. Nos dois estudos, os alunos apresentaram um rendimento mais baixo em Matemática, quando comparado com as outras áreas. Além disso, observa-se uma queda de desempenho nas 5ª e 7ª séries, quando comparado com o desempenho nas 1ª e 3ª séries, com uma queda acentuada em Matemática.

Para tentar explicar esse fato para o ciclo de 1990, apresentou-se a seguinte hipótese, levantada pelo então gestor do Inep e sua equipe, sem que nenhum estudo fosse realizado para tentar comprová-la:

A tendência a rendimentos decrescentes pode ser vinculada ao fato de que o aluno com problemas já nas primeiras séries (onde deixa de aprender 50% do mínimo necessário), progressivamente, vê estreitarem-se as possibilidades de assimilação de novos conhecimentos. Por outro lado, também é possível que as medidas adotadas por muitas unidades da Federação, para melhorar o ensino nas primeiras séries (capacitação de professores, ciclo básico, etc.) e que já provocaram modificações na eficiência do fluxo das primeiras séries, tenham repercutido no desempenho qualitativo dessas séries, melhorando os níveis de aprendizagem; dado o caráter recente das medidas, essas camadas ainda não ingressaram nas séries superiores que, por isso, continuam apresentando baixo desempenho (PILATTI, 1995, p. 17-18).

O relatório, além de mostrar a média de desempenho, mostra também a percentagem de alunos por faixas de desempenho, variando de 10 em 10 pontos. Em geral, para a última faixa, acima de 90 pontos, independentemente da área testada, o percentual da população que a atinge é inferior a 3%. Essa forma de enquadrar os desempenhos por faixas, longe de indicar uma comparação com o sentido de classificar, tenta exprimir a eficiência do sistema educacional, tendo por base que o ponto 100 indicaria a situação em que os conhecimentos mínimos, definidos depois de consulta a especialistas, a professoras do 1º grau e a currículos escolares dos Estados e de algumas capitais desses Estados, estariam sendo atingidos.

No relatório do ciclo de 1990, do Inep, não são exibidos o desvio-padrão e a discriminação média, fatores fundamentais para compreender os resultados obtidos, diferentemente do que ocorreu com o relatório de Vianna (1990).

### **Ciclo de 1993**

O próximo ciclo de avaliação, que deveria ocorrer em 1992, só ocorreu no ano seguinte, devido a problemas orçamentários. Além disso, o Inep passa a coordenar e administrar o Saeb, algo que no ano anterior era feito por equipe do MEC. Dentro dessas novas atribuições, o órgão, além de revisar suas estratégias, métodos e tecnologias, recuperou e ampliou sua infraestrutura, incentivando a constituição de equipes técnicas nas secretarias estaduais, responsáveis pelo trabalho de campo, processamento, análise, disseminação e utilização dos resultados no seu âmbito (PILATTI, 1995, p. 7). As três perguntas de pesquisa que orientaram o primeiro ciclo, de 1990, sofreram pequenas alterações, mas se mantiveram como eixos centrais para o ciclo de 1993 e, em termos de

objetivo estratégico, o Saeb, não contrário de pretender julgar ou classificar estudantes, unidades escolares ou federativas, se propôs a ser um instrumento político para o desenvolvimento e o aperfeiçoamento do sistema educacional público brasileiro (PILATTI, 1995, p. 101). Portanto, nesses primeiros ciclos do Saeb, não havia nenhuma preocupação com ranqueamento. No entanto, é interessante conceituar o Saeb como instrumento político, o que reforça a ênfase maior dada ao resultado que ao processo educacional. Essa ênfase nos resultados se transforma, paulatinamente, em ênfase na classificação, e os processos de aprendizagem desenvolvidos pela escola ficam definitivamente fora das preocupações dos relatórios do Inep.

O desenvolvimento dos testes contou com a colaboração de mais de 40 profissionais, quase todos ligados a universidades, e as professoras do antigo 1º grau ficaram com a responsabilidade de construir os itens. Nessa construção, buscou-se melhorar a qualidade dos itens, com base nas críticas do ciclo anterior, que apontaram itens inadequados/mecanizados, com vocabulário que propiciava entendimento diversificado nas diferentes regiões e ilustrações que davam margem a diversas interpretações. Houve três alterações nos testes, com relação ao ciclo anterior: a diminuição da quantidade de itens para as 1ª e 3ª séries, que passaram de 30 para 20 itens semiobjetivos; aumento no número de alternativas dos itens de múltipla escolha, aplicados aos alunos das 5ª e 7ª séries, que passaram de quatro para cinco alternativas; eliminação do teste de redação, devido à dificuldade em se garantir a unidade dos procedimentos para a correção efetivamente adotados pelas equipes de professoras das diferentes regiões do País, agravada pelo grande volume de trabalho (PILATTI, 1995, p. 11). Desde então, nunca mais se aplicou redação no Saeb.

Os resultados obtidos nesse ciclo constam do Quadro 2.3.

**Quadro 2.3 – Média das provas de Português, Matemática e Ciências, aplicadas a uma amostra de alunos de 1<sup>a</sup>, 3<sup>a</sup>, 5<sup>a</sup> e 7<sup>a</sup> séries do 1º grau da rede oficial de ensino, em 27 unidades da Federação – Saeb 1993**

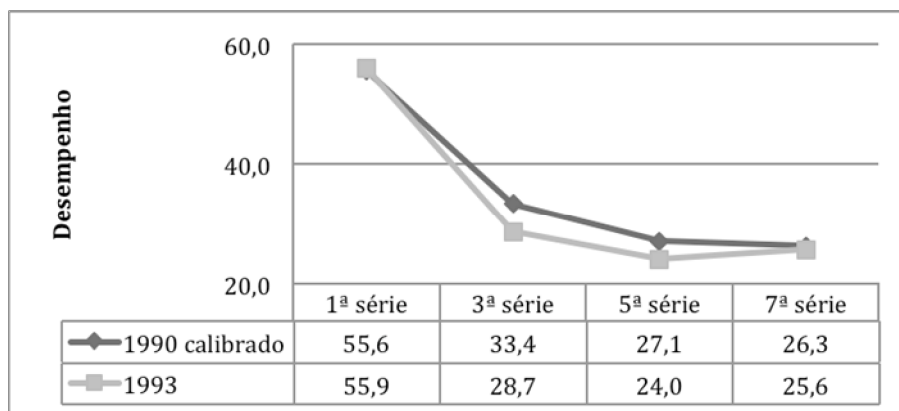
Série	Matéria	Média
1 <sup>a</sup>	Português	60,9
	Matemática	55,9
3 <sup>a</sup>	Português	52,9
	Matemática	28,7
5 <sup>a</sup>	Português	47,9
	Matemática	24,0
	Ciências	39,2
7 <sup>a</sup>	Português	55,3
	Matemática	25,6
	Ciências	41,9

Fonte: INEP, 1995, com elaboração do autor

O quadro mostra que, com exceção de Ciências, para as outras áreas com o avanço nos anos escolares, havia uma queda na pontuação média dos testes.

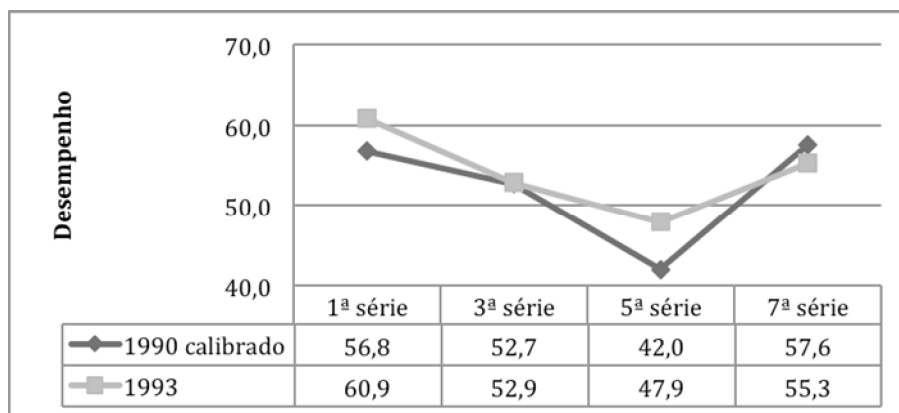
O relatório desse ciclo denota grande preocupação com a descrição da metodologia utilizada, destacando o aprimoramento da técnica de pesquisa utilizada, quando comparado com o ciclo anterior. Um desses aprimoramentos refere-se à calibragem dos testes do 1º e do 2º ciclo, que, depois de 1993, submeteram grupos de alunos de diversos Estados aos testes dos dois anos. Os resultados foram analisados utilizando-se a TRI de um parâmetro, conhecida como modelo de Rasch, em que a habilidade seria constante, já que o mesmo aluno respondeu aos dois testes e não seria levado em conta o *öchuteö*. Outro avanço foi a descrição dos conteúdos curriculares testados nas três áreas.

A partir do processo de calibração, foram geradas novas médias de desempenho para o ciclo de 1990, de forma que fosse possível comparar os desempenhos desse ciclo com o de 1993. Essas comparações deram origem aos Gráficos 2.1, 2.2 e 2.3, que mostram os desempenhos de Português, Matemática e Ciências, respectivamente.



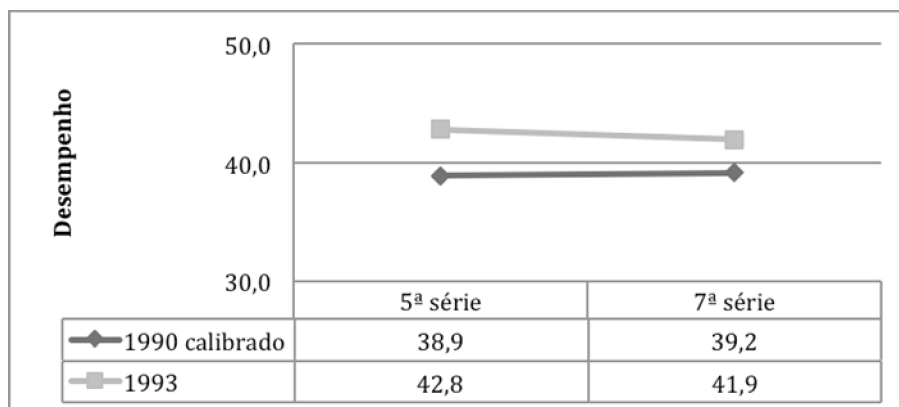
**Gráfico 2.1 – Desempenho nas 1ª, 3ª, 5ª e 7ª séries do 1º grau em Matemática: ciclo de 1990 calibrado e 1993**

Fonte: INEP, 1995.



**Gráfico 2.2 – Desempenho nas 1ª, 3ª, 5ª e 7ª séries do 1º grau em Português: ciclo de 1990 calibrado e 1993**

Fonte: INEP, 1995.



**Gráfico 2.3 – Desempenho de português nas 5ª e 7ª séries do 1º grau em Ciências: ciclo de 1990 calibrado e 1993**

Fonte: INEP, 1995.

Antes de analisar os Gráficos 2.1, 2.2 e 2.3, são necessárias algumas discussões preliminares.

Primeiramente, é necessário ressaltar uma limitação dos dados fornecidos pelo Inep, que infelizmente se repete em muitos outros relatórios do Saeb: a falta de informação sobre os erros dificulta uma análise mais precisa do resultado. Como toda medida é sujeita a erros, o fato de estes não serem explicitados deixa a falsa impressão de que o número obtido é absolutamente verdadeiro. Além disso, os relatórios não explicitam o que se espera das crianças, enquanto elas progridem em sua vida escolar, não mostrando, por exemplo, que a falta de domínio de atividades mais simples dificulta o domínio de habilidades mais complexas. Para o caso da Matemática, isso é bem claro: determinado tema depende muito do domínio de temas anteriores e sem esse domínio é muito difícil dominar as habilidades seguintes. No caso de Português, no caso dos testes cujo foco é a leitura, talvez nem tanto, dependendo muito da forma como os conhecimentos foram tratados em sala de aula. Finalmente, no caso de Ciências, existe relativa independência com relação à ordem em que os temas podem ser tratados em sala de aula. Mesmo sabendo que os testes foram elaborados com base em um currículo ideal, pois não existe um currículo nacional unificado, e ainda que se tenha ouvido especialistas, sobre o que era possível esperar dos alunos em termos de desenvolvimento cognitivo, e as professoras, sobre a realidade praticada na escola, seria necessário exemplificar melhor as informações fornecidas pelos relatórios do Saeb, relacionando-as com a realidade da escola. A ausência desses procedimentos pode dificultar muito a compreensão dos resultados pelas equipes das secretarias e das escolas.

É importante destacar, também, que os dois ciclos, 1990 e 1993, trabalharam com a testagem de conhecimentos considerados mínimos para cada ano escolar. Por isso, era de se esperar que os desempenhos dos alunos, independentemente da área, fossem próximos uns dos outros. Como não é isso o que ocorre, pode-se inferir que a referência utilizada não era a mais adequada, ou que, com o passar dos anos escolares, a escola não estaria conseguindo garantir a evolução da aprendizagem. Para nenhum dos dois casos foram realizadas pesquisas que explicassem e iluminassem o fenômeno, que, aliás, com relação à queda de desempenho com o avanço da escolaridade, aparecerá nos ciclos seguintes.

Outro fator que deve ser levado em conta é o fato de não estarem claras, nos relatórios dos dois ciclos, quais habilidades foram testadas nas três áreas e se elas variaram entre um ciclo e outro, o que significaria uma comparação espúria.

Feitas essas considerações, percebe-se no Gráfico 2.1, relativo ao desempenho de Matemática, uma queda constante e acentuada a partir da 1ª série, indicando que a escola não



está conseguindo contribuir para a superação das dificuldades dos alunos. No caso de Português, descrito no Gráfico 2.2, a queda é constante até a 5ª série, com um aumento do desempenho na 7ª série, que se aproxima do desempenho da 1ª série. No caso de Ciências, descrito no Gráfico 2.3, não é possível observar grandes variações entre os dois anos escolares.

Com relação aos dois casos tratados aqui, algumas discussões podem ser feitas. Para nenhum deles, é possível afirmar que um aluno da 3ª série, com um desempenho menor, sabe menos, ou tem menos habilidades, que um aluno da 1ª, que apresenta um desempenho maior, pois o que foi testado em uma série não foi testado em outra. Da mesma forma, esse aluno, apesar de ter sido testado em função de conhecimentos considerados mínimos, na escola também é apresentado a outros conhecimentos que engendrarão habilidades úteis em sua preparação para a vida em sociedade, função principal da escola. Só que essas habilidades, apesar de importantes para o aluno, não são testadas. Assim, essas comparações, que têm em vista apenas os números que indicam as proficiências médias, devem ser feitas com muito cuidado. Outrossim, é possível afirmar que a escola e o Estado que a mantém estão falhando, ao deixarem de assegurar as aprendizagens necessárias e em não propiciarem os avanços aos alunos, nem a correção de suas defasagens, no sentido de possibilitar a eles melhor compreensão do mundo em que vivem, acarretando graves prejuízos para a vida pessoal e profissional deles.

Muitas podem ser as hipóteses para o problema da queda de desempenho ao longo dos anos escolares, e muitas também podem ser as possíveis soluções. Para melhor enfrentar esse dilema, seriam necessários outros estudos para se obter respostas claras, algo que somente a avaliação externa não tem condições de prover, pois se limita a mostrar o efeito e não as causas. Essa discussão é importante para sugerir aos órgãos responsáveis pelas avaliações, de qualquer âmbito de atuação, que contextualizem com precisão as informações que eles disponibilizam. Um exemplo a ser evitado é a justificativa dada pela direção do Inep para a queda de desempenho entre as séries do ciclo de 1990, já tratado nesse texto e aqui reforçado:

[...] também é possível que as medidas adotadas por muitas unidades da Federação para melhorar o ensino nas primeiras séries (capacitação de professores, ciclo básico, etc.), que já provocaram modificações na eficiência do fluxo das primeiras séries, tenham repercutido no desempenho qualitativo destas séries, melhorando os níveis de aprendizagem; dado o caráter recente das medidas, essas camadas ainda não ingressaram nas séries superiores, que, por isso, continuam apresentando baixo desempenho (PILATTI, 1995, p. 17-18).

O tempo decorrido de três anos entre o ciclo de 1990 e o ciclo de 1993 teria permitido que as crianças avançassem nas séries, mas o mesmo efeito do ciclo de 1990 se repetiu em 1993, sem que nenhum comentário fosse feito, comparando-se a conclusão de 1990 com os resultados do ciclo seguinte.

Outro ponto importante a se destacar com relação ao ciclo de 1993 é que, diferentemente do anterior, foram utilizados seis questionários para levantar os dados do contexto em que ocorria a aprendizagem: informações da escola e do alunado; custo-aluno direto; custo-aluno indireto (um específico para a rede estadual ou de município de grande porte e outro para municípios de pequeno porte); aspectos da gestão; práticas das professoras. Como resultado desse levantamento, o relatório mostra algumas evidências para justificar o maior desempenho dos alunos: ter à disposição livros didáticos; desenvolvimento de todo o conteúdo previsto para o ano letivo; bom aproveitamento do tempo pedagógico; formas de trabalho coletivo; incentivo da escola à participação dos pais e o interesse deles em participar. Por outro lado, mostra, também, que não foi possível associar os melhores resultados dos alunos ao fato de as professoras e diretores terem participado de capacitações. O relatório, no entanto, assinala o limite dessas associações, informando que foram realizadas análises simples, que precisariam ser mais bem estudadas com análises multiníveis, mais adequadas ao processo educacional, que é multideterminado.

### *Segunda fase: de 1995 a 2003*

Esta segunda fase reflete grande mudança nos procedimentos adotados até então pelo Saeb. Boa parte desse período coincide com os dois governos de Fernando Henrique Cardoso, entre 1994 e 2002.

### **Ciclo de 1995**

Por um curto período de tempo, entre a posse do novo governo e a divulgação dos dados preliminares, a execução do Saeb voltou a ser atribuição do MEC.

O ciclo de 1995 só apresentou seus resultados finais em 1998 (INEP, 1998), apesar de ter havido um documento com dados preliminares, divulgado em abril de 1996, cinco meses

depois da aplicação do teste, e produzido pelo Departamento de Avaliação da Educação Básica (Daeb), ligado à Secretaria de Informação e Avaliação Educacional do MEC (Sediae) (MEC, 1996).

Algumas mudanças introduzidas no ciclo de 1995 foram marcantes: no ensino fundamental, o teste foi direcionado para as 4<sup>a</sup> e 8<sup>a</sup> séries, com a justificativa de que havia, em nível nacional, grande disparidade entre os currículos das escolas nas séries testadas nos ciclos anteriores (1<sup>a</sup>, 3<sup>a</sup>, 5<sup>a</sup> e 7<sup>a</sup> séries), algo que diminuía sensivelmente nos finais de ciclo; testar não apenas os conteúdos mínimos, mas toda a gama de habilidades que cada ciclo deveria desenvolver; testar também as escolas privadas, de forma a abarcar toda a educação escolar; inclusão do ensino médio nos testes, de forma a possibilitar um panorama integral de toda a educação básica. Com relação ao ensino médio<sup>45</sup>, testaram-se as 2<sup>a</sup> e 3<sup>a</sup> séries, sem que houvesse uma justificativa para isso. O fato é que devido a problemas de amostragem, os resultados da 2<sup>a</sup> série do ensino médio não foram divulgados. Em 1995, testaram-se apenas Português e Matemática, com itens de múltipla escolha e, no ano seguinte, em 1996, novos testes de Ciências e habilidades de escrita estavam previstos, fato que não ocorreu.

O Inep encomendou um estudo a dois pesquisadores americanos ligados à American College Testing (ACT), organização que desenvolve testes de seleção de estudantes para o ensino superior americano, com o objetivo de propor melhorias ao Saeb. Não existem dados para avaliar qual o impacto do documento produzido por eles, mas o fato é que quase todas as recomendações propostas, envolvendo as revisões dos questionários, da amostragem e dos testes, foram incorporadas aos procedimentos utilizados no terceiro ciclo do Saeb. Especificamente, com relação aos testes, foram propostas, e de fato implementadas, as seguintes ações: ampliação da quantidade de habilidades medidas pelo teste, que foram consideradas muito poucas; número de itens por habilidade, que permitisse medir a habilidade, tipicamente por volta de cinco; recrutamento de professoras familiarizadas com os testes, para criar itens que garantissem que a linguagem, a dificuldade e o conteúdo dos itens fossem apropriados àqueles que farão o teste; pré-testagem de todos os itens; os itens aprovados no pré-teste, conjuntamente com suas estatísticas, devem fazer parte de um banco; realização de comparações entre testes apenas quando apresentarem alta fidedignidade e quando tiverem as mesmas referências em termos de habilidades testadas (KORTE; NASH, 1995). Nesse último caso, o da comparação entre os testes, os autores afirmam que os testes de 1990 e 1993 apresentavam graus diferentes de dificuldade e que não foram construídos

---

<sup>45</sup> Apesar da ênfase da pesquisa ser o ensino fundamental, os quadros e tabelas farão também referência ao ensino médio com o sentido de preservar a organicidade histórica.

observando o mesmo conjunto de habilidades; portanto, não poderiam ser comparados. Como não estão disponíveis para pesquisa os materiais que embasaram as conclusões desses pesquisadores, não é possível discutir essa afirmação.

Para que muitas dessas sugestões pudessem ser colocadas em prática, a TRI passou a ser utilizada para a análise dos dados. Com isso, como mostra o relatório do ciclo de 1995 (INEP, 1998), foi possível ampliar a quantidade de itens testados, que passam dos antigos 60 itens, 30 por área, para 1.236 itens, uma média de 155 itens por área. Ou seja, em vez de um teste único, havia 208 cadernos distintos de teste, sendo que cada aluno respondia a um único teste, de Português ou de Matemática, e o número de itens por teste foi de 30 para a 4ª série, de 39 para a 8ª série do ensino fundamental e a 3ª série do ensino médio. Dessa forma, ampliava-se a possibilidade de se testar um conjunto maior de habilidades, e cada uma delas em diferentes contextos, com o objetivo de ampliar o foco de análise, antes considerado restrito. Essa alteração, se por um lado traz uma grande complicação para o processo, pois envolve a elaboração de um número muito grande de itens, por outro, amplia o olhar sobre o processo educacional. No entanto, caso não sejam tomados cuidados na definição das referências para a construção dos itens, essa ampliação da quantidade de itens é inócua. O fato é que, tanto em um como em outro caso, o mais importante é a explicitação clara dos critérios que estão sendo utilizados para o levantamento dos dados da pesquisa, e em que contexto os resultados devem ser interpretados.

Os itens foram construídos obedecendo-se às Matrizes de Especificações de conhecimentos e habilidades previamente elaboradas por série e disciplina. Nessas matrizes, indicavam-se quantos itens seriam utilizados, a quais temas deveriam se referir e quais habilidades testariam. Do total de itens utilizados, 33 deles solicitaram que os alunos da 8ª série do ensino fundamental e da 3ª série do ensino médio construíssem e escrevessem sua resposta. Talvez isso tenha sido o motivo de o relatório do ciclo de 1995 só ter sido publicado três anos depois. O restante dos itens eram de múltipla escolha, com quatro alternativas para a 4ª série e 5 alternativas para as demais.

Para visualizar o conteúdo das Matrizes de Especificações, montaram-se os Quadros 2.4 e 2.5, para as áreas de Língua Portuguesa e Matemática e para os três anos escolares testados em 1995.

**Quadro 2.4 – Matriz de especificação de itens de Língua Portuguesa e sua quantidade para as 4ª e 8ª séries do 1º grau e para a 3ª série do 2º grau – Ciclo 1995**

<b>Habilidades – 4ª série</b>				
	Compreensão inicial e estabelecimento de significado	Reflexão e extensão do significado	Exame crítico dos significados	Totais
Textos literários	21	12	3	36
Textos informativos	62	18	5	85
Textos práticos ou funcionais	7	2	0	9
<b>Totais</b>	<b>90</b>	<b>32</b>	<b>8</b>	<b>130</b>
<b>Habilidades – 8ª série</b>				
	Compreensão inicial e estabelecimento de significado	Reflexão e extensão do significado	Exame crítico dos significados	Totais
Textos literários	65	16	5	86
Textos informativos	31	17	3	51
Textos práticos ou funcionais	26	3	0	29
<b>Totais</b>	<b>122</b>	<b>36</b>	<b>8</b>	<b>166</b>
<b>Habilidades – 3ª série</b>				
	Compreensão inicial e estabelecimento de significado	Reflexão e extensão do significado	Exame crítico dos significados	Totais
Textos literários	34	45	10	89
Textos informativos	36	15	2	53
Textos práticos ou funcionais	0	16	7	23
Totais	70	76	19	165
<b>Total geral para os três anos testados</b>	<b>282</b>	<b>144</b>	<b>35</b>	<b>461</b>

Fonte: INEP, 1998, elaboração do autor.

**Quadro 2.5 – Matriz de especificação de itens de Matemática e sua quantidade para as 4ª e 8ª séries do 1º grau e para a 3ª série do 2º grau – Ciclo 1995**

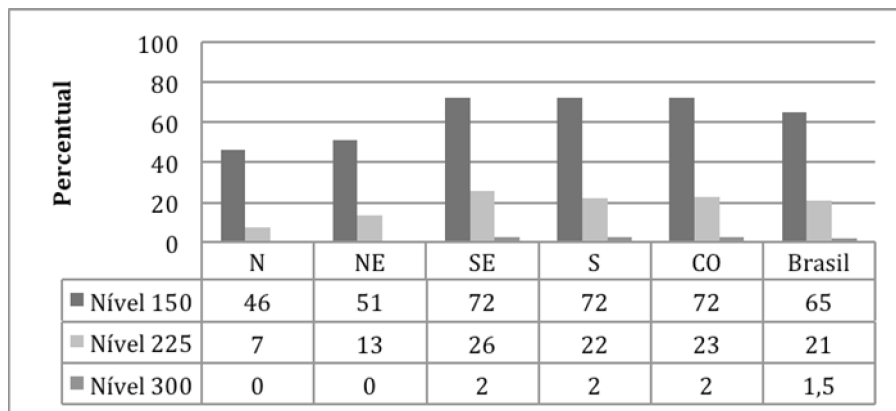
<b>Habilidades – 4ª série</b>				
	Compreensão de conceitos	Conhecimento de procedimentos	Aplicação ou resolução de problemas	Totais
Número e operações	55	37	16	108
Medidas	8	5	3	16
Geometria	5	1	0	6
Análise de dados, estatística e probabilidade	0	0	0	0
Álgebra e funções	0	0	0	0
<b>Totais</b>	<b>68</b>	<b>43</b>	<b>19</b>	<b>130</b>
<b>Habilidades – 8ª série</b>				
	Compreensão de conceitos	Conhecimento de procedimentos	Aplicação ou resolução de problemas	Totais
Número e operações	17	37	10	64
Medidas	5	9	4	18
Geometria	5	12	5	22
Análise de dados, estatística e probabilidade	8	5	7	20
Álgebra e funções	6	22	1	29
<b>Totais</b>	<b>41</b>	<b>85</b>	<b>27</b>	<b>153</b>
<b>Habilidades – 3ª série</b>				
	Compreensão de conceitos	Conhecimento de procedimentos	Aplicação ou resolução de problemas	Totais
Número e operações	5	21	6	32
Medidas	5	4	2	11
Geometria	2	13	22	37
Análise de dados, estatística e probabilidade	5	10	10	25
Álgebra e funções	9	32	4	45
Totais	26	80	44	150
<b>Total geral para os três anos testados</b>	<b>135</b>	<b>208</b>	<b>90</b>	<b>433</b>

Fonte: INEP, 1998, elaboração do autor.

Pelos quadros, percebe-se que para cada uma das duas áreas foram escolhidos diferentes conhecimentos e três grandes habilidades para serem testadas. A quantidade de itens para cada uma delas variava de acordo com o ano em que o teste seria aplicado.

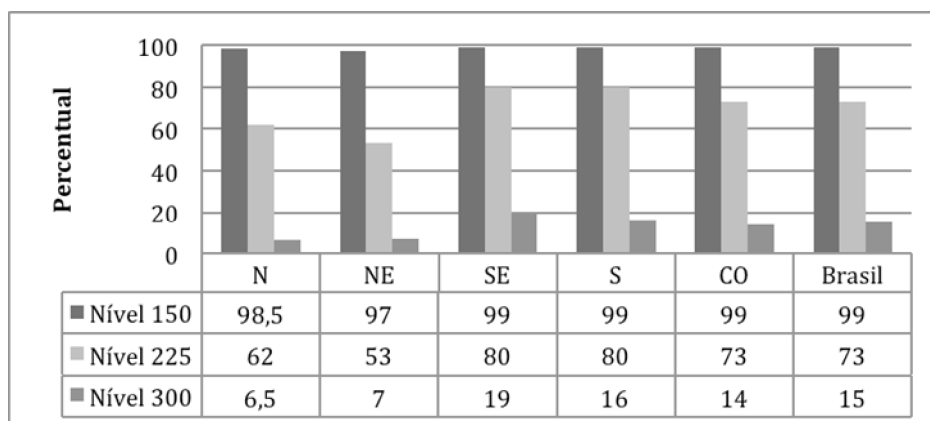
Os resultados foram agregados em nível nacional, estadual, regional e por rede federal, estadual, municipal e privada, e expressos, para cada área avaliada, em escalas que variam de 0 a 500. Os desempenhos dos três anos escolares testados de uma mesma área são colocados na mesma escala. Assim como na metodologia anterior, utilizada nos ciclos de 1990 e 1993, não é possível comparar desempenhos de alunos em testes de áreas diferentes, ou seja, 200 pontos na escala de Matemática não podem ser comparados com 200 pontos na escala de Português.

Para exemplificar como foram divulgados os relatórios, apresentam-se os Gráficos 2.4, 2.5 e 2.6, que indicam, para Matemática, o percentual da população testada que se encontra acima de cada nível de proficiência indicado, para os três anos escolares.



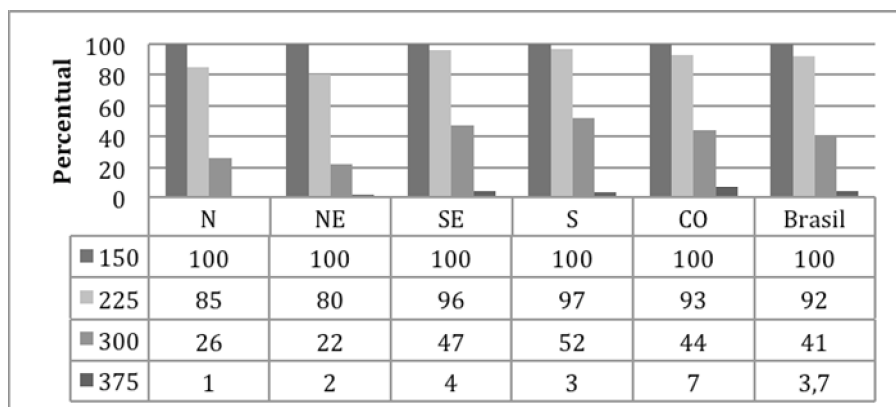
**Gráfico 2.4 – 4ª série do 1º grau, Matemática: percentual de alunos que se encontram acima dos níveis de proficiência 150, 225 e 300 – Regiões e Brasil – 1995**

Fonte: INEP, 1998, elaboração do autor.



**Gráfico 2.5 – 8ª série do 1º grau, Matemática: percentual de alunos que se encontram acima dos níveis de proficiência 150, 225 e 300 – Regiões e Brasil – 1995**

Fonte: INEP, 1998, elaboração do autor.



**Gráfico 2.6 – 3ª série do 2º grau, Matemática: percentual de alunos que se encontra acima dos níveis de proficiência 150, 225, 300 e 375 – Regiões e Brasil – 1995**

Fonte: INEP, 1998, elaboração do autor.

Analisando-se os Gráficos 2.4, 2.5 e 2.6, nota-se a diversidade existente entre as regiões, com um maior percentual de alunos agrupados nos níveis de mais alto desempenho nas Regiões Sul, Sudeste e Centro-Oeste, se comparado às Regiões Norte e Nordeste.

Essa forma de apresentar os resultados, um pouco confusa, procurou mostrar que, com o progresso escolar, aumenta o percentual de alunos que estão acima de determinado nível. Não se deve esquecer que a escala do Saeb varia de 0 a 500 pontos. Note-se que o nível 375 só aparece na 3ª série do 2º grau, conforme indicado no Gráfico 2.6.

Percebe-se, pelos gráficos, que, conforme se avança nos anos escolares, aumenta o percentual de alunos distribuídos em cada nível de proficiência, e isso faz todo sentido: espera-se que a porcentagem dos alunos dos anos finais da educação básica que tenham desenvolvido as habilidades mais simples, representadas pelo nível de proficiência 150, seja superior à dos alunos dos anos iniciais do ensino fundamental, para o mesmo nível 150. É de se esperar que ao final da educação básica, conforme aumentem os anos de escolaridade e com o maior desenvolvimento cognitivo dos alunos, os percentuais classificados nos níveis mais altos de proficiência também aumentem. Isso se justifica em razão da escala ser construída com base no pressuposto de que os níveis de complexidade são crescentes. Com o passar do tempo, como resultado das melhorias dos processos educacionais, espera-se que dentro de cada sistema educacional aumente o percentual de alunos classificados nos níveis mais altos de desempenho. No entanto, é muito complexo definir metas em termos de percentual de alunos que devem estar classificados em cada nível de desempenho, assim como é difícil para uma professora estabelecer como meta que, nos próximos anos, a quantidade de alunos que alcancem, no mínimo, a média final 7 seja de 80% da turma. Mesmo que se encontre uma forma razoável de fazer isso, qual o significado de uma meta como essa, lembrando, novamente, que apenas parte das habilidades que são trabalhadas nas escolas estão sendo medidas e que a educação escolar é muito maior do que a área de Matemática?

Com relação a dar um significado pedagógico aos níveis de proficiência, o relatório assinala que a interpretação dos níveis de proficiência ainda não foi realizada, por se tratar de uma técnica em desenvolvimento que precisa de outros elementos, como, por exemplo, os parâmetros nacionais por série e disciplina, para realizar as análises comparativas correspondentes (INEP, 1998, p. 25).

Por outro lado, mesmo que houvesse um currículo nacional, relacionar um desempenho a um currículo não faz nenhum sentido, pois o currículo mostra quais temas, entre um conjunto amplo de assuntos, serão trabalhados em cada etapa de ensino, mas não é



seu papel recomendar qual deveria ser o desempenho que se espera dos alunos. Muito menos seria esse o papel que se esperaria dos testes avaliativos.

O relatório mostra como um sério problema (INEP, 1998, p. 15) o absenteísmo no teste, que varia entre 18% e 30%, dependendo da série. Conforme a região geográfica e a série, o documento mostra que esse absenteísmo pode ter ultrapassado 40%<sup>46</sup>. Houve menção a essa preocupação no relatório do ciclo de 1990 (INEP, 1992).

Não se apresentou nenhum item como exemplo daqueles que foram utilizados nos testes. É importante destacar que na metodologia utilizada nos testes, para que seja possível comparar o desempenho entre os anos escolares e entre um ciclo e outro, nem todos os itens do teste podem ser divulgados, pois alguns deles serão utilizados no ciclo seguinte, para calibrar os próximos itens de forma a estarem todos na mesma escala.

### **Ciclo de 1997**

Algumas modificações ocorreram neste ciclo: excluíram-se do universo testado os alunos das turmas multisseriadas e das escolas federais e os estudantes das escolas rurais da Região Norte e da 3ª série do ensino médio matriculados em cursos profissionalizantes. A área de Ciências voltou a ser testada no ensino fundamental, como ocorreu em 1990 e 1993, além de Física, Química e Biologia no ensino médio e, definitivamente, a 2ª série do ensino médio deixou de ser testada.

A principal das alterações foi substituir o que se chamou de Matriz de Especificações, em que genericamente se indicavam a quantidade de habilidades a serem testadas no ciclo de 1995, por Matrizes de Referência para o Saeb, descrevendo as habilidades que seriam testadas. Assim, no seu quarto ciclo, torna-se pública a referência a ser utilizada para as medidas. Na sua constituição, as matrizes utilizam os conceitos de *competências*<sup>47</sup> e *habilidades cognitivas*, conceito que também estava em discussão pelas equipes do MEC que estavam preparando, na mesma época, os Parâmetros Curriculares Nacionais.

---

<sup>46</sup> Avançando um pouco no tempo para debater esse ponto, o absenteísmo elevado foi tema de uma declaração feita em 2012 pelo ministro Aloísio Mercadante, ao afirmar que uma das saídas para controlar esse problema, pelo menos na 3ª série do ensino médio, seria substituir o Saeb, que tem como objetivo fornecer um retrato dos erros e acertos da educação básica, pelo Enem, que se transformou, em 2008, em exame de seleção para a universidade, ainda que também seja utilizado como etapa de certificação para a educação de jovens e adultos. Essa declaração levantou uma intensa discussão na imprensa, com diferentes posições. O mais importante é destacar que, para a direção do MEC, se trata apenas de substituir um instrumento por outro, de forma a ter números mais precisos, independentemente de esses números terem significados totalmente distintos, evidenciando-se, assim, que o processo educacional saiu definitivamente das preocupações, importando apenas os resultados do desempenho dos alunos.

<sup>47</sup> O termo *competência* trata de um conceito complexo e com diferentes significados, dependendo da cultura, do contexto em que está inserido e da perspectiva de análise. Para aprofundar essa discussão, recomenda-se Rychen e Salganick (2001) e Sacristán (2011).

Cada matriz relacionava três níveis de competências cognitivas: básica ou presentativa, na qual se encontrariam as ações que tornariam presente o objeto do conhecimento para o sujeito; operacional ou procedural, reservada para as ações e operações que pressupunham o estabelecimento de relação com e entre os objetos; global ou operatório, que se referia às ações e operações mais complexas que envolveriam a aplicação de conhecimentos e a resolução de problemas inéditos (INEP, 1998a). No total, foram construídos 319 descritores para Língua Portuguesa (em 1995, eram 461, uma redução de 31%), 239 para Matemática (em 1995, eram 433, uma redução de 45%), 210 para Ciências, 133 para Física, 151 para Química e 140 para Biologia. A redução no número de descritores não foi justificada, mas isso significou uma restrição no número de habilidades testadas, diminuindo a amplitude da medida e essa ação pode ter sido utilizada como uma forma de reduzir custos.

O relatório, por sua vez, não informou a quantidade de itens utilizados, nem a quantidade de blocos e cadernos de provas gerados, representando um problema sério para uma pesquisa desse tipo, pois priva a sociedade e os pesquisadores da área, em particular, de informações para analisar e julgar o processo desenvolvido. Aliás, a não divulgação completa dos procedimentos utilizados tem se revelado uma prática comum em todos os relatórios do Saeb divulgados.

Os Quadros 2.6, 2.7 e 2.8 contêm a síntese das Matrizes de Língua Portuguesa e o Quadro 2.9 a síntese de Matemática para as 4<sup>a</sup> e 8<sup>a</sup> séries do ensino fundamental e para a 3<sup>a</sup> série do ensino médio.

**Quadro 2.6 – Matriz de Referência de itens de Língua Portuguesa e sua quantidade para a 4ª série do ensino fundamental, em 1997**

Conteúdos		Níveis de competências			
Temas	Tópicos.	Nível presentativo	Nível operacional	Nível operatório	Total
		(Quantidade de descritores)			
Práticas de leitura de textos	Estratégias de leitura.	4	7	4	15
	Articulação texto e conteúdo.	1	1	1	3
	Utilização de mecanismos básicos de coesão no processamento do texto.	2	3	1	6
	Relações na progressão temática do texto.	-	2	2	4
	Relações entre recursos expressivos e efeitos de sentido.	1	4	3	8
Análise e reflexão linguística	Texto e gênero.	3	2	2	7
	Variação linguística.	1	3	0	4
	Operações linguísticas de estabelecimento de relações entre forma e sentido.	2	2	2	6
	Paradigmas linguísticos (regras, classes).	-	5	1	6
<b>Totais</b>		<b>14</b>	<b>29</b>	<b>16</b>	<b>59</b>

Fonte: INEP, 1998a, elaboração do autor.

**Quadro 2.7 – Matriz de Referência de itens de Língua Portuguesa e sua quantidade para a 8ª série do ensino fundamental, em 1997**

Conteúdos		Níveis de competências			
Temas	Tópicos	Nível presentativo	Nível operacional	Nível operatório	Total
		(Quantidade de descritores)			
Práticas de leitura de textos	Gêneros e suas características.	2	8	5	15
	Sentidos constituídos em função das características do portador, do gênero e do enunciador.	-	4	4	8
	Estratégias de leitura.	4	11	5	20
	Relações entre textos.	1	2	2	5
	Mecanismos de coesão referencial no processamento do texto.	3	2	-	5
	Progressão temática de um texto.	-	4	4	8
Análise e reflexão linguística	Recursos expressivos recorrentes, em função da estratégia de construção textual ou do gênero em que o texto se insere.	1	1	1	3
	Relações texto, gênero e contexto.	2	2	2	6
	Operações linguísticas de estabelecimento de relações entre forma e sentido.	3	3	-	6
	Variação linguística.	4	3	2	9
<b>Totais</b>		<b>20</b>	<b>40</b>	<b>25</b>	<b>85</b>

Fonte: INEP, 1998a, elaboração do autor.

**Quadro 2.8 – Matriz de Referência de itens de Língua Portuguesa e sua quantidade para a 3ª série do ensino médio, em 1997**

Conteúdos		Níveis de competências			
Temas	Tópicos	Nível presentativo	Nível operacional	Nível operatório	Total
		(Quantidade de descritores)			
Práticas de leitura de textos	Características e estratégias próprias de cada gênero.	3	6	5	14
	Valores e sentidos constituídos pelo texto em função das características do portador e/ou enunciador.	6	4	4	8
	Estratégias de leitura.	6	11	8	25
	Relação entre texto e contexto e entre texto e outros textos implicados pelo primeiro.	1	6	4	11
	Mecanismos de coesão referencial no processamento do texto.	2	3	1	6
	Índices formais de estabelecimento de relação sintático-semântica na progressão temática.	6	9	18	27
	Progressão temática e estratégia argumentativa.	1	3	3	7
	Contexto histórico social na análise textual.	6	1	2	3
	Estratégias de construção da verossimilhança do universo ficcional.	6	3	2	5
	Recursos expressivos utilizados na produção de um texto.	5	5	5	15
Análise e reflexão linguística	Recursos expressivos recorrentes em função da estratégia de construção textual ou do gênero em que o texto se insere.	1	1	1	3
	Construção da imagem do locutor e do elocutório.	1	1	1	3
	Relações de intertextualidade entre dois enunciados.	1	1	1	3
	Operações linguísticas de estabelecimento de relações entre forma e construção de sentido.	11	11	1	23
	Variação linguística.	4	5	1	10
	O texto enquanto objeto sociologicamente construído.	1	4	3	8
	Funções da linguagem e atos de fala.	6	2	2	4
<b>Totais</b>		<b>24</b>	<b>76</b>	<b>62</b>	<b>175</b>

Fonte: INEP, 1998a, elaboração do autor.

**Quadro 2.9 – Matriz de Referência para o Saeb de Matemática e sua quantidade para as 4ª e 8ª séries do ensino fundamental e para a 3ª série do ensino médio, em 1997**

Conteúdos		Competências									
Temas	Tópicos	Básicas			Operacionais			Globais			Total
		4ª	8ª	3ª	4ª	8ª	3ª	4ª	8ª	3ª	
Geometria	Espaço	7			5		2				14
	Formas	5	6	6	11	6	2	3	6	6	21
	Curvas	2	6	2	2	6	2	2	6	6	10
	Retas	6	2	6	6	2	6	6	6	6	4
	Ângulos	6	2	6	6	2	6	6	1	6	5
	Figuras planas	6	3	6	6	7	6	6	3	6	13
	Figuras espaciais	6	6	6	6	8	6	6	6	6	8
	Ponto e reta	6	6	6	6	6	8	6	6	6	8
	Circunferência	6	6	1	6	6	4	6	6	6	5
Cônicas	6	6	1	6	6	1	6	6	6	2	
Números	Números naturais	1	6	6	2	6	6	1	6	6	4
	Sistemas de numeração	1	6	6	6	6	6	1	6	6	2
	Números racionais positivos	1	6	6	2	6	6	1	6	6	4
	Números naturais e operações	6	6	6	6	2	6	6	1	6	3
	Números inteiros e operações	6	2		6	5	6	6	1	6	8
	Números racionais e operações	6	2	6	6	8	6	6	6	6	10
	Operações algébricas	6	6	6	6	2	6	6	3	6	5
Operações	Significado das operações	2	6	6	2	6	6	2	6	6	6
	Aplicação de propriedades	6	6	6	6	6	6	11	6	6	11
	Cálculo mental	1	6	6	2	6	6	6	6	6	3
Medidas	Cumprimento, superfície, capacidade e massa	6	6	6	8	6	6	6	6	6	8
	Tempo	1	6	6	2	6	6	6	6	6	3
	Temperatura	1	6	6	1	6	6	6	6	6	2
	Sistema monetário	1	6	6	3	6	6	6	6	6	4
Estatística	Lista, tabela e gráfico	6	6	6	2	6	6	1	6	6	3
	Noções de proporcionalidade, porcentagem e juros	6	6	6	6	8	6	6	3	6	11
	Estatística	6	6	6	6	6	2	6	6	6	2
Trigonometria						6			3		
Relações e funções	Conceituação e propriedades	6	6	2	6	6	3	6	6	1	6
	Funções, equações, inequações de 1º grau	6	6	2	6	6	6	6	6	6	8
	Módulo	6	6		6	6	2	6	6	6	2
	Exponencial e logarítmico	6	6	6	6	6	5	6	6	6	5
	Funções trigonométricas	6	6	6	6	6	1	6	6	6	6
Matrizes, determinantes e sistemas lineares	Matrizes	6	6	6	6	6	2	6	6	6	6
	Determinantes	6	6	6	6	6	1	6	6	1	2
	Sistemas lineares	6	6	6	6	6	2	6	6	6	2
Análise combinatória e probabilidade	6	6	6	6	6	3	6	6	2	4	
Números complexos	6	6	6	6	6	8	6	6	1	3	
Polinômios e equações polinomiais	6	6	6	6	6	8	6	6	6	4	
<b>Totais</b>		<b>23</b>	<b>11</b>	<b>8</b>	<b>42</b>	<b>44</b>	<b>68</b>	<b>22</b>	<b>12</b>	<b>9</b>	<b>239</b>

Fonte: INEP, 1998a, elaboração do autor.

Pelos quadros, percebe-se que os conteúdos de cada área foram agrupados em temas e tópicos dentro da especificidade de cada uma. Ao mesmo tempo, foram estabelecidas três competências comuns para as duas áreas. A quantidade de itens que deveriam fazer parte de

cada teste surge do cruzamento entre conteúdos e competências, variando em função do ano escolar em que o teste foi aplicado.

Da mesma forma que nos ciclos anteriores, cada aluno foi submetido a apenas um dos testes, mas o relatório não informa quantos itens foram utilizados. Só existe a informação da utilização de 691 itens para a interpretação da escala.

Outra novidade foi a discussão com especialistas para estabelecer associações ou relações entre os momentos dos ciclos escolares (e os desempenhos mínimos ou básicos que a eles correspondiam) e os níveis de proficiência da escala (que retratam o desempenho real dos alunos) (INEP, 1997a, p. 11). Essa parece ter sido a primeira tentativa para classificar os sistemas educacionais a partir do desempenho dos alunos.

Do conjunto de itens testados, alguns que apresentam determinadas características<sup>48</sup> são escolhidos para serem colocados no respectivo nível e, com base neles, um grupo de especialistas os analisa, buscando fornecer respostas sobre o que o aluno demonstrou saber ao assinalar a resposta correta. O relatório não mostra quantos itens, do total de 1.236 utilizados nos testes, foram considerados âncora. A descrição dos diferentes itens de um determinado nível são agrupados para descrever o significado daquele nível de proficiência, mostrando o que os alunos com a proficiência igual ao do nível são capazes de fazer.

A classificação obtida, relacionando o nível de proficiência e a etapa de ensino, é apresentada no Quadro 2.10.

**Quadro 2.10 – Relação entre nível de proficiência na escala do Saeb/97 e etapa de ensino**

Nível de proficiência (escala Saeb/97)	Ciclo e etapa de ensino		
	Matemática	Língua Portuguesa	Ciências (Física, Química e Biologia)
100	Não significativo.	Até a metade do 1º ciclo do ensino fundamental.	Até a metade do 1º ciclo do ensino fundamental.
175	Até a metade do 1º ciclo do ensino fundamental.	Até o final do 1º ciclo do ensino fundamental.	Até o final do 1º ciclo do ensino fundamental.
250	Até o final do 1º ciclo do ensino fundamental.	Até o final do 2º ciclo do ensino fundamental.	Até a metade do 2º ciclo do ensino fundamental.
325	Até o final do 2º ciclo do ensino fundamental.	Até o final do ensino médio.	Até o final do 2º ciclo do ensino fundamental.
400	Até o final do ensino médio.	Além do final do ensino médio	Até o final do ensino médio.

Fonte: INEP, 1997a, p. 11.

<sup>48</sup> Devem ter as seguintes características: a maioria (65%) dos alunos com aquele nível de proficiência deve ter acertado o item. Ao mesmo tempo, devem ter acertado o item a maioria (95%) dos alunos com nível de proficiência superior àquele nível e no máximo (25%) dos alunos do nível anterior. Nesse caso o item é chamado de âncora (INEP, 1998).

De acordo com o quadro, ao final dos anos iniciais do ensino fundamental, esperava-se que os alunos dos diferentes sistemas educacionais apresentassem proficiências de 250 em Matemática, 175 em Língua Portuguesa e também 175 em Ciências.

Tomando-se o caso da Matemática, apenas para exemplificar, construiu-se o Quadro 2.11, que mostra a descrição dos dois níveis de proficiência, o 175 e o 250. Como, pelo modelo utilizado, o conhecimento é cumulativo, os alunos com desempenho localizado no nível 250, além de terem grande probabilidade de dominar as habilidades desse nível, têm grande probabilidade de dominar aquelas do nível anterior. Os especialistas, referidos pelo relatório, avaliaram que essas habilidades, contidas na descrição dos níveis 175 e 250, seriam aquelas que deveriam ser desenvolvidas até o final da 4ª série do ensino fundamental.

**Quadro 2.11 – Descrição dos níveis de proficiência 175 e 250 – Ciclo de 1997**

Nível 175	Nível 250
<ul style="list-style-type: none"> <li>Ó Demonstram conhecimentos elementares de geometria. São capazes de localizar a posição dos objetos tendo como referencia o próprio corpo e de reconhecer figuras geométricas simples, desde que representadas na forma usual.</li> <li>Ó Reconhecem o valor de cédulas e moedas.</li> <li>Ó Leem horas em relógios digitais e analógicos e sabem que a hora tem 60 minutos.</li> <li>Ó Leem e escrevem números de pouco dígitos.</li> <li>Ó Resolvem problemas simples de adição e subtração com números naturais.</li> </ul>	<ul style="list-style-type: none"> <li>Ó Os alunos possuem conhecimento de geometria, que possibilita a descrição da movimentação de objetos, tendo como referencia o próprio corpo.</li> <li>Ó Reconhecem polígonos e quadriláteros.</li> <li>Ó Estabeleceram relações entre os valores de cédulas e moedas e resolvem situações de pagamento e troco.</li> <li>Ó Relacionam diferentes unidades de medida de tempo e sabem selecionar a unidade de medida mais adequada para fazer determinada medição.</li> <li>Ó São capazes de multiplicar e dividir e de identificar unidades, dezenas, centenas, etc.</li> <li>Ó Resolvem problemas envolvendo mais de uma operação.</li> <li>Ó Adicionam e subtraem frações de mesmo denominador e conhecem números naturais na forma fracionária.</li> <li>Ó Interpretam gráficos de barra e de setor e identificam o gráfico mais adequado para representar uma dada situação.</li> <li>Ó Expressam generalizações observadas em sequências numéricas por meio de representações algébricas.</li> </ul>

Fonte: INEP, 1997a, elaborado pelo autor.

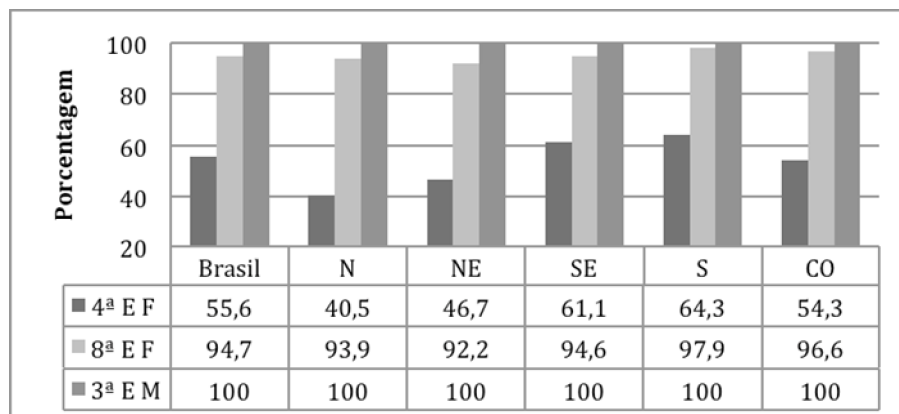
Observa-se no quadro que a descrição dos níveis é feita a partir de frases que se referem às habilidades que foram testadas pelos itens.

Mesmo que tecnicamente, com base no consenso dentro de um pequeno grupo de especialistas, faça sentido associar um grupo de habilidades a uma etapa da educação básica, o fato de essa associação ter sido realizada sem que se envolvessem grupos bem maiores de educadores discutindo em diversas instâncias a adequação ou não da referência só serve para

trazer ainda mais desconfianças aos testes avaliativos. No fundo, esse pequeno grupo de especialistas tomou decisões ignorando a existência de um debate que dura há muitos anos, não só aqui no Brasil: afinal deve haver ou não referências nacionais para a aprendizagem dos alunos? Se a resposta for positiva, essas referências devem ser por anos escolares e qual deve ser o detalhamento dessas referências? E ainda, qual deve ser o processo de discussão e quais instituições devem direcionar o processo? Essa é uma questão importante e qualquer que seja o caminho a seguir, deve ser respeitado o pacto federativo e as atribuições legais de cada Ente da Federação. É importante recordar que nos primórdios do uso das avaliações, na época de Tyler, elas estiveram ligadas ao aprimoramento do currículo. Sendo assim, era o currículo a referência para os testes e para a análise de seus resultados. No momento, os testes avaliativos, seja no Brasil, seja em outros países, ou os testes produzidos por organismos internacionais, utilizam como referência aquela que um grupo de especialistas considera como adequada ao desenvolvimento cognitivo dos alunos e a uma pretensa exigência de um mundo em constante transformação. Dessa forma, os testes avaliativos deixam de servir como instrumentos para avaliar as dificuldades dos alunos, as estratégias de ensino escolhidas pela professora ou mesmo as políticas educacionais e transformam-se na própria política educacional, impondo que todos os esforços sejam dirigidos no sentido de buscar o aumento das pontuações. Justifica-se esse fetiche pelos números como a forma de garantir um pretense direito à aprendizagem dos alunos ou de melhorar a qualidade da educação, qualidade essa limitada, pois se baseia apenas em testes com limitada capacidade para captar a complexa realidade do processo educacional.

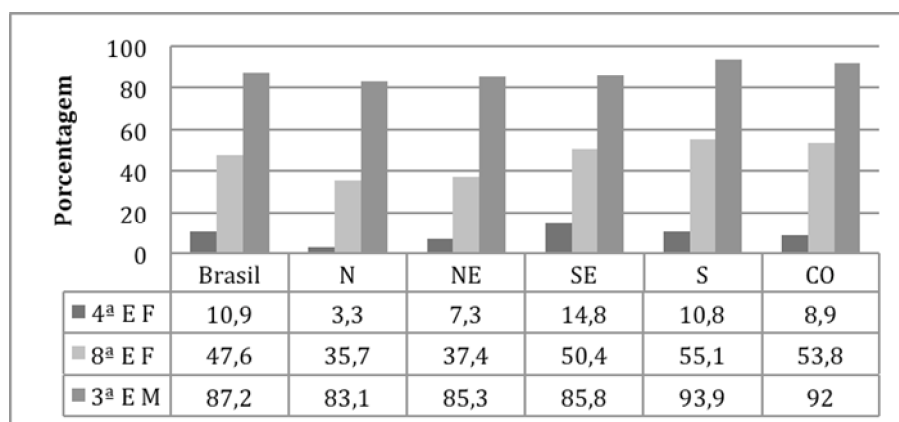
Mantendo o mesmo procedimento utilizado no ciclo de 1995, o relatório apresenta o desempenho dos estudantes como função da porcentagem de alunos em cada nível. Para exemplificar como esse resultado foi apresentado, são transcritos os Gráficos 2.7, 2.8 e 2.9, que mostram o desempenho em Matemática nos três anos escolares testados.





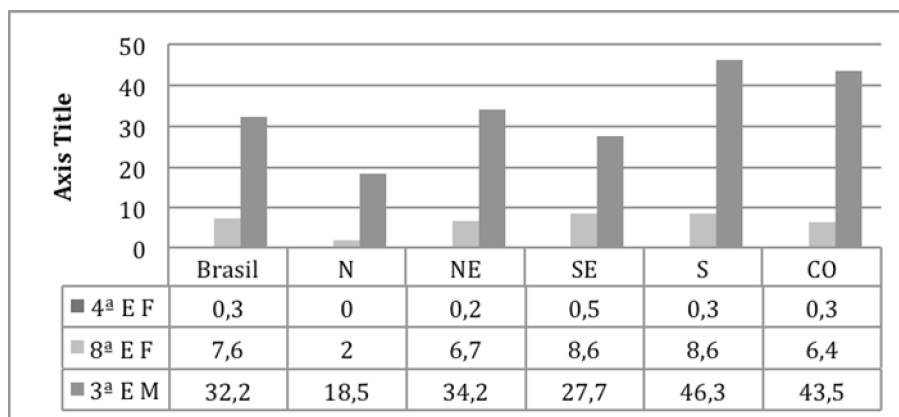
**Gráfico 2.7 – Percentagem de alunos acima do nível 175 na escala de Matemática  
Brasil e regiões – Ciclo de 1997**

Fonte: INEP, 1997a, p. 12.



**Gráfico 2.8 – Percentagem de alunos acima do nível 250 na escala de Matemática  
Brasil e regiões – Ciclo de 1997**

Fonte: INEP, 1997a, p. 12.



**Gráfico 2.9 – Percentagem de alunos acima do nível 325 na escala de Matemática  
Brasil e regiões – Ciclo de 1997**

Analisando os Gráficos 2.7, 3.8 e 3.9, percebe-se que, tal qual a discussão acerca do relatório de 1995, conforme se avança na escolaridade, aumenta o percentual de alunos distribuídos nos níveis mais altos de proficiência, revelando um aumento progressivo no domínio das habilidades mais complexas.

Por sua vez, o relatório dá o primeiro passo no sentido de classificar os sistemas educacionais, ao divulgar a posição relativa dos Estados com relação à média de desempenho nacional, sem ainda classificá-los individualmente por suas próprias médias. Um exemplo dessa forma de classificação é mostrado no Quadro 2.12, que apresenta a média de Matemática.

**Quadro 2.12 – Comparação das médias de proficiência em Matemática das unidades da Federação e das regiões com a média nacional, segundo a série – Ciclo de 1997**

Série	Abaixo da média nacional	No mesmo nível da média nacional	Acima da média nacional
4ª do ensino fundamental	N, NE RO, AC, AM, RR, PA, AP, TO, MA, PI, PE, AL	SE, S, CO CE, RN, PB, SE, BA, ES, RJ, SP, PR, SC, RS, MS, MT, GO, DF	MG
8ª do ensino fundamental	N, NE RO, AC, RR, AP, MA, PB, PE, AL, SE, MT	SE, CO AM, PA, TO, PI, CE, RN, BA, MG, ES, RJ, SP, RS, MS, GO, DF	S PR, SC
3ª do ensino médio	N RO, AC, AM, RR, PB	NE, SE PA, AP, TO, MA, CE, RN, PE, AL, BA, ES, RJ, SP, PR, MS, MT, GO	S, CO PI, SE, MG, SC, RS, DF

Fonte: INEP, 1997a, p. 19.

Analisando-se o quadro, observa-se que, mesmo não ranqueando os Estados, a forma de divulgar os resultados apresenta um juízo de valor com relação à média nacional, mostrando aqueles que estão abaixo da média, nela, ou acima dela. Que tipo de conclusões pode ser tirado a partir dessa informação? Nenhuma. São informações que apenas posicionam os Estados e as regiões em relação a uma média. Nada mais.

O relatório também apresenta um dado muito importante para se compreender melhor o que se encontra por trás das médias de proficiência, ao mostrar sua dispersão em cada Estado. Com esses dados, caracteriza-se um dado preocupante da realidade educacional brasileira: a extrema desigualdade em termos de desempenho nos testes existentes dentro de cada Estado. A título de exemplo, apresenta-se o Quadro 2.13, relativo à proficiência em Matemática para a 4ª série do ensino fundamental, em que esses dados podem ser visualizados.

**Quadro 2.13 – Percentis de distribuição, dispersão e erro-padrão da proficiência em Matemática para a 4ª série do ensino fundamental – Brasil, regiões e Estados – Ciclo de 1997**

Área geográfica	Proficiência							
	5%	25%	50%	Média	75%	95%	$\Delta$ (95%-5%)	Erro-padrão
<b>Brasil</b>	<b>117</b>	<b>150</b>	<b>183</b>	<b>187</b>	<b>220</b>	<b>272</b>	<b>155</b>	<b>1,8</b>
RO	112	145	173	175	202	245	133	3,8
AC	112	133	155	161	184	228	116	2,4
AM	113	140	165	172	198	249	136	3,0
RR	105	132	155	159	181	228	123	1,5
PA	111	137	164	168	194	236	125	4,6
AP	111	136	163	165	188	233	122	1,9
TO	112	139	167	171	195	248	136	4,6
<b>N</b>	<b>112</b>	<b>138</b>	<b>165</b>	<b>169</b>	<b>195</b>	<b>240</b>	<b>128</b>	<b>2,1</b>
MA	111	137	162	169	195	251	140	5,5
PI	114	141	167	173	201	253	139	3,7
CE	117	144	174	180	210	270	153	3,6
RN	113	141	169	174	200	252	139	4,2
PB	112	147	178	178	206	252	140	4,1
PE	111	138	164	173	201	264	153	4,1
AL	109	136	161	167	193	241	132	4,9
SE	116	145	171	176	200	253	137	4,2
BA	111	146	178	182	209	269	158	5,1
<b>NE</b>	<b>112</b>	<b>142</b>	<b>171</b>	<b>176</b>	<b>204</b>	<b>261</b>	<b>149</b>	<b>1,9</b>
MG	123	168	214	209	247	288	165	5,3
ES	113	143	171	177	205	267	154	5,1
RJ	117	150	180	185	216	265	148	4,4
SP	118	158	189	193	227	285	167	6,3
<b>SE</b>	<b>118</b>	<b>157</b>	<b>192</b>	<b>195</b>	<b>230</b>	<b>282</b>	<b>164</b>	<b>3,5</b>
PR	122	166	197	197	227	270	148	6,0
SC	130	171	202	203	235	277	147	8,6
RS	123	153	181	186	218	259	136	5,8
<b>S</b>	<b>124</b>	<b>162</b>	<b>192</b>	<b>194</b>	<b>225</b>	<b>267</b>	<b>143</b>	<b>3,8</b>
MS	123	151	182	187	218	268	145	9,5
MT	112	141	167	172	198	245	133	9,2
GO	123	153	187	189	220	268	145	9,3
DF	122	149	180	187	214	280	158	3,1
<b>CO</b>	<b>120</b>	<b>150</b>	<b>180</b>	<b>185</b>	<b>214</b>	<b>266</b>	<b>146</b>	<b>4,8</b>

Fonte: INEP, 1997a, p. 56.

Pelo quadro, é possível perceber quão grande é a dispersão dos resultados, em média 155 pontos. Esse valor é maior do que 3,0 desvios-padrão, abrangendo mais do que um nível de proficiência e mostrando, assim, a grande desigualdade em um mesmo Estado. Fato semelhante ao da 4ª série ocorreu nos outros dois anos escolares.

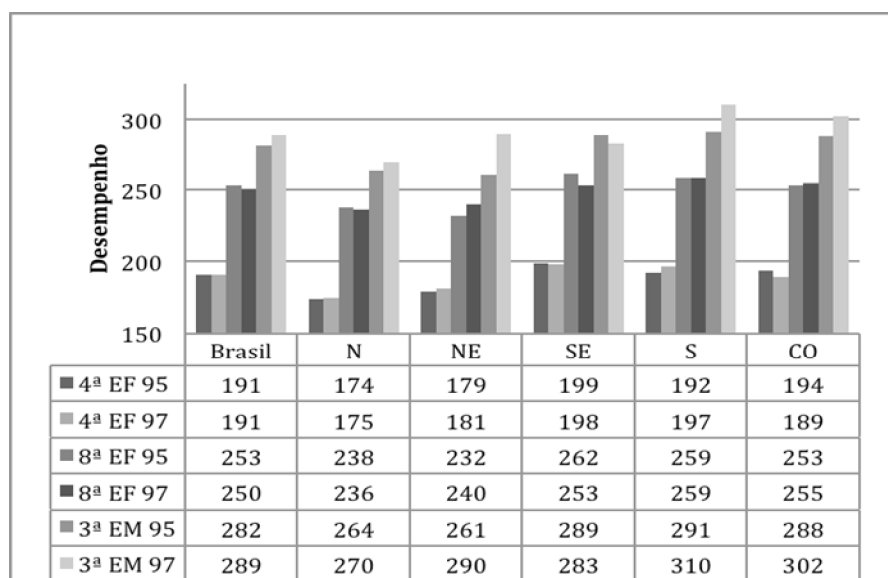
O relatório também apresentou diversos cruzamentos entre os desempenhos observados nas áreas testadas e nos três anos escolares e dados dos questionários respondidos pelos alunos, professoras e diretores. Os efeitos dos cruzamentos realizados indicaram: o baixo desempenho dos alunos que estudam na área rural; o maior desempenho dos alunos que estudam nas escolas particulares; que o desempenho dos alunos aumenta tanto com o aumento

da escolarização do pai como da mãe; o desempenho maior dos alunos que estudam na capital em relação àqueles que estudam no interior; a queda da proficiência com o aumento da idade do aluno, caracterizando o efeito negativo da repetência escolar; a variação por gênero, indicando que, em geral, as alunas apresentam maior desempenho em Língua Portuguesa e menor nas outras áreas, quando comparadas com os alunos.

No final do relatório, são apresentados, como exemplos dos itens utilizados nos testes, 36 itens âncora e suas estatísticas, juntamente com a descrição do nível de proficiência correspondente.

Pelos dados e pelas discussões apresentadas no relatório, percebe-se que, timidamente, surge, durante o ciclo de 1997, uma preocupação com a comparação de desempenhos entre os Estados.

Além do relatório, outro documento foi produzido, no qual se afirma que, pela primeira vez, seriam apresentados os resultados comparativos do Saeb (INEP, 1997), o que não corresponde à verdade, pois os dados de 1990 e 1993 também foram comparados, conforme foi discutido anteriormente neste texto. Destacou-se nesse documento a variação da proficiência média entre as regiões. Um exemplo de como essa informação foi apresentada pode ser observado no Gráfico 2.10, em que é apresentada a variação da proficiência em Matemática.



**Gráfico 2.10 – Proficiência média por série em Matemática. Brasil e regiões – 1995-1997**

Fonte: INEP, 1997, p. 5.

De acordo com o gráfico, observa-se que a maior variação ocorre na 3ª série do ensino médio, mas mesmo assim ela é pequena.

Nos resultados também foram expressos Estados e regiões, com relação à média de proficiência nacional. Nessa forma de expressão dos resultados, destaca-se a preocupação com a posição relativa entre os Estados, que aparece, inclusive, no título do Quadro 2.14, o qual reproduz o resultado da variação entre os dois ciclos para Matemática, contido no documento.

**Quadro 2.14 – Posição das unidades da Federação em relação à variação da média de proficiência em Matemática – 1995-1997**

Série	Variação das médias entre 1995 e 1997		
	Queda da média (diminuição estatisticamente significativa)	Estabilidade (diferença estatisticamente não significativa)	Aumento da média (aumento estatisticamente significativo)
4ª EF	RR, DF	<b>BRASIL, N, NE, SE, S, CO</b>  RO, AC, AM, PA, AP, TO, MA, PI, CE, RN, PB, PE, AL, SE, BA, MG, ES, RJ, SP, PR, SC, RS, RS, MS, MT, GO	
8ª EF	RR, SP, DF	<b>BRASIL, N, SE, S, CO</b>  RO, AC, AM, PA, AP, TO, MA, PI, CE, RN, PB, PE, AL, SE, BA, MG, ES, RJ, PR, RS, RS, MT, GO	<b>NE</b>  SC, MS
3ª EM		<b>BRASIL, N, SE, CO</b>  RO, AC, AM, RR, PA, AP, TO, MA, PB, AL, SE, ES, RJ, SP, PR, MT, GO, DF	<b>NE, S</b>  PI, CE, RN, PE, BA, MG, SC, RS, MS

Fonte: INEP, 1997, p. 6.

Analisando as informações do quadro, nota-se que, no geral, houve pouca alteração na educação básica como um todo. No ensino fundamental, nota-se, na 4ª série, queda em dois Estados; na 8ª série, queda de desempenho em três Estados e aumento em dois. Com relação ao ensino médio, observa-se aumento maior no desempenho, envolvendo duas regiões e nove Estados. Nenhuma hipótese foi lançada para explicar o resultado obtido, mesmo porque isso seria muito difícil, pois, como comentado antes, esse quadro não tem significado algum, a menos que seja utilizado para classificar os Estados.

### Ciclo de 1999

Nesse ciclo houve novas mudanças nas Matrizes de Referência do Saeb. Uma delas foi a introdução, para os três anos escolares, de Matrizes de História, com 127 descritores, e de

Geografia com 250. No entanto, apesar de a matriz ter sido desenvolvida, as duas áreas não foram testadas nesse ciclo e nem nos próximos. Além disso, foi feita uma nova redução na quantidade de descritores das outras áreas: Língua Portuguesa perdeu o tema Análise e Reflexão Linguística, ficando com 124 descritores, uma redução de 48%; Matemática, com um pequeno ajuste, ficou com 223 descritores, uma perda de 16 descritores ou menos de 1%; Física ficou com 100 descritores, uma perda de 25%; Química ficou com 101 descritores, uma perda de 33%; Biologia ficou com 100 descritores, uma perda de 29%.

A despeito das alterações, foram testadas apenas Língua Portuguesa e Matemática, e isso se manteve em todos os ciclos seguintes, sem qualquer explicação sobre por que se concentrar apenas nessas duas áreas, quer seja uma eventual dificuldade de elaboração de itens, por conta de problemas em alguns descritores, quer seja a necessidade de simplificar a aplicação do teste ou até mesmo dispender menos recursos com a aquisição de itens.

Os Quadros 2.15 e 2.16 apresentam uma síntese dos descritores das Matrizes de 1999.

**Quadro 2.15 – Matriz Curricular de Referência do Saeb para Língua Portuguesa e quantidade de descritores para as 4ª e 8ª séries do ensino fundamental e para a 3ª série do ensino médio – 1999**

Conteúdos		Competências									Total
		Básicas			Operacionais			Globais			
		4ª	8ª	3ª	4ª	8ª	3ª	4ª	8ª	3ª	
Práticas de leitura de texto	Procedimentos de leitura.	2	2	2	4	6	5	2	2	4	29
	Implicações do suporte do gênero e/ou do enunciador na compreensão do texto.	1	1	6	6	1	6	1	2	6	6
	Coerência e coesão no processamento do texto.	6	6	1	5	6	3	6	1	1	17
	Relações entre recursos expressivos e efeitos de sentido.	6	6	6	2	6	6	7	5	6	14
	Variações linguísticas.	4	3	6	6	6	1	6	1	3	12
	Relação entre textos: intertextualidade.	6	2	1	6	3	2	6	1	2	11
	Características dos gêneros.	6	6	3	6	6	2	6	6	4	9
	Características do suporte e/ou do enunciado na construção de valores e sentidos.	6	6	2	6	6	5	6	6	1	8
	Progressão temática e organização argumentativa e narrativa.	6	6	1	6	6	4	6	6	4	9
	Relações entre recursos expressivos e efeitos de sentido.	6	6	6	6	6	6	6	6	4	4
	Construção da imagem de locutor e de interlocutor.	6	6	1	6	6	6	6	6	2	3
	O texto enquanto objeto sócio historicamente construído.	6	6	4	6	6	2	6	6	6	6
<b>Totais</b>	<b>5</b>	<b>8</b>	<b>15</b>	<b>11</b>	<b>14</b>	<b>24</b>	<b>10</b>	<b>12</b>	<b>25</b>	<b>124</b>	

Fonte: INEP, 2000, elaboração do autor.

**Quadro 2.16 – Matriz Curricular de Referência do Saeb para Matemática e quantidade de descritores para as 4ª e 8ª séries do ensino fundamental e para a 3ª série do ensino médio, em 1999**

Conteúdos		Competências									
		Básicas			Operacionais			Globais			Tot
		4ª	8ª	3ª	4ª	8ª	3ª	4ª	8ª	3ª	
Temas/Tópicos		4ª	8ª	3ª	4ª	8ª	3ª	4ª	8ª	3ª	Tot
Geometria	Espaço ó localização, movimentação e representação (pontos de referência)	4	ó	ó	3	ó	ó	ó	ó	ó	7
	Formas bidimensionais e tridimensionais (elementos e propriedades)	7	ó	ó	11	ó	ó	5	ó	ó	23
	Retas no plano	ó	1	ó	ó	2	ó	ó	ó	ó	3
	Ângulos	ó	2	ó	ó	2	ó	ó	1	ó	5
	Figuras planas	ó	3	ó	ó	9	ó	ó	3	ó	15
	Figuras tridimensionais	ó	1	ó	ó	5	ó	ó	ó	ó	6
	Retas e planos no espaço	ó	ó	4	ó	ó	ó	ó	ó	ó	4
	Áreas e figuras planas	ó	ó	ó	ó	ó	3	ó	ó	1	4
	Sólidos	ó	ó	5	ó	ó	4	ó	ó	ó	9
	Ponto e reta	ó	ó	3	ó	ó	ó	ó	ó	1	10
	Circunferência	ó	ó	ó	ó	ó	3	ó	ó	ó	3
Cônicas	ó	ó	1	ó	ó	1	ó	ó	ó	2	
Referencial	ó	ó	1	ó	ó	ó	ó	ó	ó	1	
Números	Números naturais e sistema de numeração decimal	2	ó	ó	2	2	ó	ó	ó	ó	6
	Números naturais positivos: representação decimal e fracionaria	1	ó	ó	6	ó	ó	1	ó	ó	8
	Números naturais e operações ó inteiros racionais e reais	4	3	ó	ó	ó	1	ó	1	3	12
	Números inteiros e operações	ó	2	1	ó	3	2	ó	1	2	11
	Números racionais e irracionais e operações	ó	3	ó	ó	6	ó	ó	ó	4	13
	Operações algébricas	ó	ó	ó	ó	3	5	ó	1	ó	9
Operações	Significado das operações: adição, subtração, multiplicação e divisão.	ó	ó	ó	2	ó	ó	2	ó	ó	4
	Propriedades das operações	ó	ó	ó	ó	ó	ó	5	ó	ó	5
	Cálculo	ó	ó	ó	2	ó	ó	ó	ó	ó	2
Estatística	Lista, tabela simples e de dupla entrada e gráfico	ó	ó	ó	2	ó	ó	2	ó	ó	4
	Noções de proporcionalidade, probabilidade, porcentagem e juros	ó	ó	ó	ó	7	ó	ó	2	ó	9
	Estatística	ó	ó	ó	ó	ó	4	ó	ó	ó	4
Funções	Funções (conceitos e propriedades)	ó	ó	1	ó	ó	4	ó	ó	ó	5
	Funções polinomiais	ó	ó	1	ó	ó	4	ó	ó	ó	5
	Funções modulares	ó	ó	1	ó	ó	1	ó	ó	ó	2
	Funções exponenciais e logarítmicas	ó	ó	ó	ó	ó	5	ó	ó	ó	5
	Trigonometria	ó	ó	ó	ó	ó	6	ó	ó	3	9
Matrizes, determinantes e sistemas lineares	Matrizes e determinantes	ó	ó	ó	ó	ó	2	ó	ó	1	3
	Sistemas lineares	ó	ó	ó	ó	ó	2	ó	ó	ó	2
Sequências numéricas		ó	ó	ó	ó	ó	2	ó	ó	ó	2
Análise combinatória e probabilidade		ó	ó	ó	ó	ó	4	ó	ó	ó	4
Números complexos		ó	ó	ó	ó	ó	3	ó	ó	ó	3
Polinômios e equações polinomiais		ó	ó	ó	ó	ó	4	ó	ó	ó	4
Totais		18	15	18	28	39	66	15	9	15	223

Fonte: INEP, 2000, elaboração do autor.

Os quadros mostram a quantidade de itens que deveriam fazer parte de cada teste, que é determinada pelo cruzamento entre conteúdos e competências. Essa quantidade varia em função do ano escolar em que o teste é aplicado.

O relatório desse ciclo traz uma boa discussão sobre o panorama educacional daquela época e as diferentes políticas que vinham sendo desenvolvidas no período entre 1995 e 1999. Um destaque especial foi dado ao crescimento de 10% da matrícula no ensino fundamental e 45% no ensino médio e ao processo de municipalização do ensino fundamental. No entanto, a discussão sobre o resultado do desempenho dos alunos foi apresentado de forma bastante aligeirada, tomando apenas quatro das 79 páginas do relatório. Enquanto nos relatórios dos ciclos anteriores se apresentava uma farta quantidade de informações sobre as proficiências, no de 1999 elas eram esparsas. O Quadro 2.17 reproduz um dos dois quadros apresentados no relatório.



**Quadro 2.17 – Níveis de desempenho em Matemática – Saeb/99**

Série	Ciclo	Nível +160 até 175	Nível +175 até 225	Nível +225 até 275	Nível +275 até 325	Nível +325 até 375
4ª EF	1997	N  AC, RR, PA, AP, MA, AL	<b>BRASIL, NE, SE, S, CO</b>  RO, AC, AM, RR, PA, AP, TO, MA, PI, CE, RN, PB, PE, AL, SE, BA, MG, ES, RJ, SP, PR, SC, RS, MS, MT, GO, DF			
	1999	<b>BRASIL, N, NE, SE, S, CO</b>  RO, AC, AM, RR, PA, AP, TO, MA, PI, CE, RN, PB, PE, AL, SE, BA, MG, ES, RJ, SP, PR, SC, RS, MS, MT, GO, DF				
8ª EF	1997		MA	<b>BRASIL, NE, SE, S, CO</b>  RO, AC, AM, RR, PA, AP, TO, PI, CE, RN, PB, PE, AL, SE, BA, MG, ES, RJ, SP, PR, SC, RS, MS, MT, GO, DF		
	1999			<b>BRASIL, N, NE, SE, S, CO</b>  RO, AC, AM, RR, PA, AP, TO, MA, PI, CE, RN, PB, PE, AL, SE, BA, MG, ES, RJ, SP, PR, SC, RS, MS, MT, GO, DF		
3ª EM	1997			N  RO, AC, RR, PA, AP, TO, MA, RJ	<b>BRASIL, N, NE, SE, S, CO</b>  AM, PI, CE, RN, PB, PE, AL, SE, BA, MG, ES, SP, PR, SC, RS, MS, MT, GO, DF	MG, RS
	1999			N, NE  RO, AC, AM, RR, PA, AP, TO, MA, PI, CE, RN, PB, PE, AL, SE, BA	<b>BRASIL, N, NE, SE, S, CO</b>  MG, ES, RJ, SP, PR, SC, RS, MS, MT, GO, DF	

Fonte: INEP, 2000, p. 65.

Uma mudança que se observa na forma de comunicação dos resultados é o título do Quadro 2.17 referir-se a ãnível de desempenho, quando nos ciclos anteriores utilizava-se a

expressão ôníveis de proficiênciã. Pode parecer mero detalhe, mas denota mais um avanço em direção ao ranqueamento. Além disso, quando se comparam os Quadros 2.14 e 2.17, também se percebe que se informam duas coisas distintas como se fossem a mesma. No Quadro 2.14, informava-se para Matemática se havia ou não uma diferença estatisticamente significativa que implicasse queda, manutenção ou aumento da média de desempenho. No Quadro 2.17, informa-se apenas em que níveis de proficiência se encontravam os Estados e as regiões, não se tomando a média para fazer a comparação. Além disso, os intervalos dos níveis de desempenho utilizados foram alterados, algo que também ocorreu comparando-se 1995 e 1997.

Independentemente das razões que possam justificar as alterações na forma de apresentar os resultados, o fato é que a proficiência dos alunos apresentou sucessivas quedas em 1997 e 1999. Isso gerou um problema político grande para o governo da época, pois se considerava o fato como consequência da ineficiência da política federal na área educacional. Esta sempre foi a grande armadilha do Saeb para o governo federal: quando os resultados apresentam melhoria, rapidamente tenta-se associar o fato às medidas adotadas; quando o contrário acontece, e sob o ãfogo cerradoã da oposição, procura-se uma justificativa, de preferência, relacionando o fato a outra medida positiva. E foi o que aconteceu no período. Justificaram-se as quedas sucessivas pelo aumento das matrículas na educação básica, principalmente das camadas mais pobres da população. Essas camadas, por conta do grave problema social brasileiro, teriam mais dificuldades na escola e com isso o desempenho do sistema caiu. Que no período houve um aumento da matrícula, principalmente por causa de várias campanhas do MEC, é fato. Mas que a escola e suas professoras não estavam preparados para receber esse novo contingente também é verdade. Ataca-se um problema para propiciar a melhoria de alguns indicadores, mas se esquece que, na área social, muitos fatores estão interligados. Para minimizar o impacto na divulgação dessa queda, o relatório de 1999 talvez tenha dado pouco destaque aos resultados desse ciclo.

De qualquer forma, relatórios que se prestam a dar informações sobre resultados de um mesmo tipo de pesquisa, e que modificam a forma de apresentar seus resultados, trazem grandes dificuldades para os pesquisadores em busca de comparações e geram desconfianças quanto às metodologias adotadas. Seria muito importante que o Inep estabelecesse um modelo para os relatórios do Saeb e o seguisse fielmente, indicando claramente as alterações realizadas, suas razões e como comparar os novos formatos com os anteriores.

Assim como no relatório do ciclo de 1997, nesse não há a indicação de quantos itens foram utilizados, quantos blocos de itens foram gerados e nem quantos cadernos de provas foram construídos.

## **Ciclo 2001**

Um aspecto que marca esse ciclo é relatado por Franco (2001). O autor destaca que o ministro de Educação, na época Paulo Renato Souza, nomeou uma comissão de especialistas, por meio da Portaria MEC nº 243/2001, de 12 de fevereiro desse ano, com o objetivo de oferecer subsídios para aprimorar o Saeb e para refinar as contribuições que o sistema poderia oferecer à formulação de políticas educacionais. Além do texto do autor, nenhum outro documento foi encontrado pela pesquisa sobre as conclusões dessa comissão.

O fato é que nesse ciclo, não se sabe se por influência da comissão ou não, as Matrizes sofrem outra alteração, dessa vez com maior intensidade. Essa foi a última alteração e desde então nenhuma outra foi realizada nas Matrizes. Muitos descritores foram agrupados gerando novos e, no total, diminuiu-se significativamente seu número. As Matrizes de Matemática passaram a ter 110 descritores, uma redução de 51% e as de Língua Portuguesa passaram a ter 57 descritores, uma redução de 54%. Assim como nos ciclos anteriores, nenhuma justificativa foi apresentada.

Instituiu-se, também, provavelmente pelo fato de não haver equivalência entre o número de 169 itens utilizados no teste e o total de descritores nas Matrizes, diferentes graus de prioridade para cada descritor. Assim, por exemplo, para participar de um teste da 4ª série do ensino fundamental, os descritores associados ao tema Procedimentos de Leitura têm maior prioridade, no caso P1 e outro P5, do que um associado à Variação Linguística, que tem prioridade P5. O problema é que não existe um critério objetivo para isso e não se sabe qual o significado prático de um descritor ter prioridade P1. Provavelmente, por não terem sido mais utilizadas, as matrizes das outras áreas não sofreram alterações.

Os Quadros 2.18 e 2.19 apresentam as Matrizes de Língua Portuguesa e Matemática, a quantidade de descritores estabelecidos para os três anos escolares testados e as prioridades associadas a cada um.

**Quadro 2.18 – Matriz de Referência de Língua Portuguesa, quantidade de descritores para as 4ª e 8ª séries do ensino fundamental e para a 3ª série do ensino médio e grau de prioridade dos descritores, de P1 aP5, para sua inclusão no teste, em 2001**

<b>Matrizes de 2001 – Língua Portuguesa</b>			
<b>Temas</b>	<b>Descritores</b>		
	4ª série	8ª série	3ª série
Procedimentos de leitura.	5 (P1)	5 (P3)	5 (P3)
Implicações do suporte, do gênero e/ou do enunciador na compreensão do texto.	2 (P3)	2 (P5)	2 (P5)
Relação entre textos.	1 (P5)	2 (P4)	2 (P4)
Coerência e coesão no processamento do texto.	4 (P2)	7 (P1)	7 (P1)
Relação entre recursos expressivos e efeitos de sentido.	2 (P4)	4 (P2)	4 (P2)
Variação linguística.	1 (P5)	1 (P5)	1 (P5)
<b>Total</b>	<b>15</b>	<b>21</b>	<b>21</b>
<b>Total geral</b>	<b>57</b>		

Fonte: INEP, 2002, elaboração do autor.

**Quadro 2.19 – Matriz de Referência de Matemática, a quantidade de descritores para as 4ª e 8ª séries do ensino fundamental e para a 3ª série do ensino médio e o grau de prioridade dos descritores, de P1 aP5, para sua inclusão no teste, em 2001**

<b>Matrizes de 2001 – Matemática</b>			
<b>Tópicos</b>	<b>Descritores</b>		
	4ª série	8ª série	3ª série
Espaço e forma.	5 (P3)	11 (P2)	10 (P3)
Grandezas e medidas.	7 (P2)	4 (P3)	3 (P4)
Números e operações/álgebra e funções.	14 (P1)	20 (P1)	20 (P1)
Tratamento da informação.	2 (P4)	2 (P2)	2 (P2)
<b>Total</b>	<b>28</b>	<b>47</b>	<b>35</b>
<b>Total geral</b>	<b>110</b>		

Fonte: INEP, 2002, elaboração do autor.

Os quadros mostram a quantidade de itens que deveriam fazer parte de cada teste, e sua quantidade varia em função do ano escolar em que o teste será aplicado. Diferentemente dos anos anteriores, não foram estabelecidas habilidades ou competências gerais e cada item está relacionado a um determinado descritor.

Em termos gerais, o relatório do ciclo de 2001, que seria o último do Saeb durante o governo Fernando Henrique Cardoso, foi o mais bem elaborado, com uma descrição detalhada de todos os procedimentos técnicos utilizados, mostrando inclusive os intervalos de confiança para cada uma das médias apuradas, algo que esteve ausente nos relatórios anteriores e que não foram novamente utilizados nos relatórios seguintes. Se o Inep decidisse criar um padrão para os relatórios do Saeb, a base para essa discussão teria que, necessariamente, ser o desse ciclo. Além disso, foi o mais didático de todos. No total, foram produzidos quatro volumes: um descrevendo os pressupostos teóricos, as Matrizes, a organização dos testes e dos questionários e o plano amostral (INEP, 2002); outro mais geral,

descrevendo a metodologia utilizada, os dados de contexto e os resultados (INEP, 2002a); outro com o relatório de Matemática (INEP, 2002b) e o último com o de Língua Portuguesa (INEP, 2002c).

Quanto à forma de organizar os resultados, novas alterações foram efetuadas, uma constante a cada relatório apresentado. Apesar de os resultados ainda serem divulgados por níveis, eles agora passam a ser indicados não pelo ponto da escala (150, 225, etc.), mas por números, que variam de 1 a 8 para Língua Portuguesa e de 1 a 10 para Matemática. Os intervalos de pontos da escala para cada nível mudam novamente. Apresentam-se quadros com o percentual de alunos por nível de proficiência e uma descrição de habilidades por nível, a mais detalhada de todas já produzidas.

O relatório apresenta alguns estudos novos, como o ajuste da média de proficiência dos alunos pelo nível socioeconômico, procurando compensar as desigualdades sociais e seus impactos sobre o desempenho dos alunos. Um exemplo desse ajuste é apresentado no Quadro 2.20.

**Quadro 2.20 – Médias de desempenho do Saeb 2001 em Matemática, ajustadas pelo nível socioeconômico, para as 4ª e 8ª séries do ensino fundamental e para a 3ª série do ensino médio**

Série	Região	Proficiência	Proficiência ajustada
4ª EF	Norte	163,63	168,39
	Nordeste	158,70	167,21
	Sudeste	189,82	183,38
	Sul	188,15	183,17
	Centro-Oeste	175,73	175,31
8ª EF	Norte	231,86	235,49
	Nordeste	228,79	236,71
	Sudeste	249,72	244,95
	Sul	255,34	252,87
	Centro-Oeste	244,83	247,13
3ª EM	Norte	255,07	263,63
	Nordeste	264,12	270,98
	Sudeste	280,20	275,35
	Sul	292,99	295,15
	Centro-Oeste	285,06	282,91

Fonte: INEP, 2002a

Pelo quadro, observa-se que o desempenho das regiões mais pobres aumenta depois da aplicação do ajuste.

Como não foi apresentada a metodologia para a realização desse ajuste, não é possível discuti-la. O que se observa é que nas Regiões Norte e Nordeste, as mais pobres, o desempenho ajustado pelo nível socioeconômico dos alunos aumenta.

Quanto às proficiências, novamente observa-se uma queda, mas, diferentemente do relatório do ciclo passado, existem diversas tabelas que mostram essa queda, talvez porque a queda tenha sido pequena. Essas tabelas também comparam os resultados desse ciclo com os resultados dos outros ciclos, mostrando inclusive os erros-padrão associados a cada medida.

### **Ciclo de 2003**

Esse ciclo ocorreu durante o primeiro governo de Luiz Inácio Lula da Silva, entre 2003 e 2007. Nessa época, havia uma grande expectativa sobre quais seriam os rumos que os testes iriam tomar, pois durante todo governo anterior setores expressivos do Partido dos Trabalhadores, principal partido da coalisão que elegeu o novo presidente, criticaram duramente a realização dos testes.

Dois fatos marcaram o início do governo. Primeiro a realização de um seminário no Inep, entre os dias 7 e 8 de abril de 2003, intitulado "Avaliar para quê? Avaliando as políticas de avaliação educacional", com a participação de diversos pesquisadores brasileiros, entidades de classe, como a Confederação Nacional dos Trabalhadores em Educação (CNTE) e a União Nacional dos Estudantes (UNE), e dirigentes educacionais, para discutir as mudanças nas avaliações educacionais. Segundo o presidente do Inep na época, Otávio Helene, o objetivo do evento era "debater os processos de avaliação e adequá-los à situação de um novo governo que tem um compromisso social com a área da Educação, diferente do governo anterior, que tinha compromisso com um perfil liberalizante ou mercantilista de educação" (MEC, 2003). Pelo tom da declaração, era de se esperar grandes alterações. Infelizmente, não foi possível encontrar documentos resultantes desse seminário que possibilitassem uma visão clara sobre as propostas ali apresentadas.

O segundo fato foi a publicação, em junho daquele ano, de um documento produzido pelo MEC, à época, comandado por Cristovam Buarque, intitulado *Toda Criança Aprendendo* (MEC, 2003a), em que são traçadas quatro diretrizes a serem adotadas pelo Ministério, relativas à educação básica: 1) Política Nacional de Valorização e Formação de Professores,

envolvendo o piso salarial, a Bolsa Federal de Incentivo à Formação Continuada e o Exame Nacional de Certificação de Professores;<sup>49</sup> 2) ampliação do atendimento escolar, com o objetivo de incluir as crianças de 6 anos de idade no ensino fundamental; 3) programas de apoio ao letramento, envolvendo ações de aceleração de aprendizado, aliás uma política muito defendida no governo anterior; e 4) a gratificação de incentivo ao letramento, instituída pelo MEC e que seria

[...] calculada com base nos padrões de desempenho alcançados pelos alunos das séries ou ciclos iniciais das unidades escolares. Os recursos correspondentes a essa gratificação serão concedidos à escola, podendo ser utilizados para remunerar os seus professores e demais servidores. Competirá aos Estados e municípios aferir os resultados alcançados, utilizando, para tanto, procedimentos de avaliação externa do desempenho do alunado (MEC, 2003a, p. 201).

Pelo que se percebe até aqui, havia uma divergência entre o MEC e sua autarquia, o Inep. Ao que parece, o ministério queria aumentar a quantidade de testes avaliativos, ampliando sua abrangência, agregando, além de testes para professoras, testes para aferir o letramento, testes esses que poderiam significar o recebimento de uma gratificação diretamente do governo federal, a ser paga às professoras e aos servidores de outros entes federados. No caso do Inep, pelas declarações de seu presidente, parece que o objetivo era restringir o uso dos testes, inclusive modificando seu enfoque.

A última das quatro diretrizes do programa *Toda Criança Aprendendo* previa o apoio à construção de sistemas estaduais de avaliação da educação básica, pois, segundo o documento, o sistema atual não estaria conseguindo impactar o cotidiano da escola. Portanto, incentivar os Estados a desenvolver sistemas próprios seria um caminho para influenciar mais diretamente a escola, no sentido de melhorar o desempenho dos alunos.

O fato é que nem o ministro nem o presidente do Inep permaneceram em seus cargos até fins de 2003. No entanto, parte das ideias de Cristovam Buarque foi aproveitada pelos dirigentes que o seguiram, no sentido de gerar mais pressão sobre a escola, como estratégia para garantir melhores resultados, estratégia muito distinta daquela defendida por Otaviano Helene e que caminhava mais ainda em direção a um aprofundamento das políticas neoliberais que se queria combater inicialmente.

---

<sup>49</sup> O MEC chegou a instituir o Sistema Nacional de Formação Continuada e Certificação de Professores, instituído pela Portaria MEC nº 1.043/2003, revogada no início de 2004, quando Tarso Genro assume o Ministério. Mais informações, inclusive a Matriz de Referência para o Exame, podem, ser obtidas em MEC, 2003b.

O relatório técnico com os resultados do Ciclo de 2003 do Saeb inova ao discutir alterações nos questionários contextuais, pois surgiu uma série de questionamentos acerca da validade dos dados coletados e ainda do grau de adequação de cada um dos questionários à realidade brasileira (INEP, 2006, p. 172), e apresenta 12 quadros nos quais justifica cada uma das alterações realizadas.

Como nesse ciclo começa uma melhoria das médias de proficiência, o relatório ressalta isso como um fato importante, dando grande destaque à queda sofrida nos ciclos anteriores, utilizando, para tanto, gráficos com escalas, em que as quedas dos anos anteriores estavam bem destacadas.

A última e mais polêmica inovação trazida pelo relatório foi a criação de níveis adjetivados como muito crítico, crítico, intermediário e adequado, para classificar o desempenho dos alunos de acordo com os intervalos da escala. Para cada nível, foram estabelecidas quais seriam as habilidades típicas dos alunos. O problema é que o relatório não apresenta uma única justificativa para as escolhas dos intervalos, nem para a descrição utilizada. Além disso, vários erros podem ser observados no documento, demonstrando falta de cuidado técnico com uma publicação de tamanha importância.

### **3. Terceira fase: a partir de 2005**

A Portaria MEC nº 931, de 21 de março de 2005, regulamenta o Saeb, que passaria a ser composto pela Avaliação do Rendimento Escolar (Anresc), conhecida por Prova Brasil, e pela Avaliação Nacional da Educação Básica (Aneb), que mantém as características do Saeb, de medir o desempenho a partir de uma amostra de estudantes, mais conhecida como Saeb amostral. A Prova Brasil passa a ser aplicada em todas as escolas públicas urbanas com mais de 30 alunos matriculados nas 4ª e 8ª séries do ensino fundamental, e o Saeb amostral continua coletando informações sobre as escolas públicas e privadas, urbanas e rurais, das 4ª e 8ª séries do ensino fundamental e da 3ª série do ensino médio.

Antes de 2005, cada aluno respondia a um teste que envolvia uma única área do conhecimento. A partir desse ciclo, cada aluno responde a um teste com itens das duas áreas, Língua Portuguesa e Matemática. Não foi feito nenhum estudo para verificar se essa alteração trouxe algum tipo de variação nos resultados, já que todo o restante permaneceu inalterado. Apenas para recordar algo já comentado neste texto, Heraldo Vianna afirmou que para não cansar os alunos, cada um deles responderia ao teste de uma única área (VIANNA, 1990).



Desde o ciclo de 2005, não são mais publicados os relatórios técnicos, e as informações relativas às proficiências das escolas passaram a ser divulgadas no *site* do Inep. Um fato como esse fato prejudica bastante a transparência de um procedimento de tamanha relevância, que envolve tantas pessoas. Qualquer que seja a pesquisa, uma das regras básicas é o acesso à metodologia utilizada e às informações que deram origem às medidas divulgadas. Portanto, seria importante que o Inep voltasse a produzir e a divulgar os relatórios técnicos.

A partir desse ciclo, passaram a ser corriqueiras, a partir do *ranking* fornecido pelo Inep, a classificação e a identificação das melhores escolas e também das piores, noticiadas pela imprensa. As melhores escolas têm direito a matérias elogiosas nos meios de comunicação. Além disso, se transformam em estudos de caso que dão origem a pesquisas, com o objetivo de identificar as boas práticas que resultam em bons resultados na Prova Brasil (MEC; INEP; UNICEF, 2007). Quanto às piores, a exposição pública.

Com relação aos custos de aplicação, a pesquisa de Oliveira (2011) encontrou os valores de R\$ 10,60 por aluno e um total de quase R\$ 55 milhões, em 2005; R\$ 13,14 por aluno e um total de R\$ 71 milhões, em 2007; e R\$ 9,00 por aluno e um total de R\$ 48 milhões, em 2009.

Um fato importante é que, a partir de 2011, o Inep passa a ter o controle sobre a construção dos itens que usa em seus testes. Até então, usava-se como estratégia comprá-los em lotes, mediante processos licitatórios nos quais concorriam instituições com experiência em avaliação educacional. A partir de 2011, o processo envolve o treinamento, pelos pesquisadores do Inep, de equipes de professoras universitárias de instituições públicas, para a construção de itens, que, depois de construídos e de passarem por revisão baseada em critérios também definidos pelo Inep, são validados pelo Instituto. Com isso, abre-se a possibilidade de melhorar a qualidade dos itens dos testes preparados pelo Inep. Para verificar essa melhoria, basta comparar a qualidade pedagógica dos itens constantes do Enem de 2009, por exemplo, com os itens do exame em 2012. A mudança foi significativa.

#### **4. Mudanças na composição dos testes entre 1995 e 2011**

Neste tópico, pretende-se dar destaque às mudanças ocorridas na composição dos testes. Para isso, comparou-se o teste aplicado em 1995, primeiro a utilizar a análise por meio da TRI, que garantiu a comparação constante entre os resultados, com o de 2011, o último aplicado até o momento desta pesquisa.

Assim como houve mudança nas matrizes, também houve mudança no formato do teste, para adequá-lo às variações na quantidade dos descritores, e também no formato da aplicação amostral para quase censitária, característica da Prova Brasil. O que permaneceu foi a forma de dividir o teste em blocos e cadernos, a utilização de itens comuns entre séries e ciclos e a divisão de tempos, existindo um tempo máximo para a conclusão de cada bloco e a determinação de os alunos só avançarem para o bloco seguinte quando esse tempo máximo for atingido.

O teste de 1995, além das tradicionais 4<sup>a</sup> e 8<sup>a</sup> séries do ensino fundamental e da 3<sup>a</sup> série do ensino médio, foi aplicado também à 2<sup>a</sup> série do ensino médio. Utilizaram-se, no total, 1.236 itens (338 apenas para a 2<sup>a</sup> série do ensino médio), reunidos em 13 blocos e em 26 cadernos por série e disciplina, agrupados em 208 cadernos. Cada caderno foi composto por três blocos contendo entre 10 e 13 itens cada. Cada aluno deveria responder entre 30 e 39 itens, dependendo da série, sendo que uma parte da turma respondia a cadernos de Língua Portuguesa e a outra parte a cadernos de Matemática. Quanto às questões comuns, utilizaram-se: 10 itens de Língua Portuguesa e 10 de Matemática para a 4<sup>a</sup> e a 8<sup>a</sup> série do ensino fundamental e a 3<sup>a</sup> série de ensino médio; 26 de Língua Portuguesa e 22 de Matemática para a 8<sup>a</sup> série do ensino fundamental e a 3<sup>a</sup> série do ensino médio; 30 de Língua Portuguesa e 30 de Matemática para a 4<sup>a</sup> e a 8<sup>a</sup> série do ensino fundamental; 13 de Língua Portuguesa e 13 de Matemática para a 8<sup>a</sup> série do ensino fundamental e para a 2<sup>a</sup> e a 3<sup>a</sup> série do ensino médio; 39 de Língua Portuguesa e 39 de Matemática para a 2<sup>a</sup> e a 3<sup>a</sup> série do ensino médio. Alguns dos itens foram do tipo resposta construída, apesar de não haver indicação de qual a quantidade utilizada. Esses itens foram corrigidos por bancas formadas por quatro professoras, cada uma para Língua Portuguesa, e por outras quatro para cada banca de Matemática. O relatório técnico (BRASIL, 1998a) explica como foi esse processo de correção.

O teste de 2011 foi composto por 20 cadernos diferentes, cada um contendo quatro blocos de itens, sendo dois de resolução de Matemática e dois de Língua Portuguesa, procedimento adotado desde 2005. Cada aluno do 5<sup>o</sup> ano do ensino fundamental respondeu a um total de 44 itens e os alunos do 9<sup>o</sup> ano do ensino fundamental e da 3<sup>a</sup> série do ensino médio responderam a 52 itens. Para garantir a comparabilidade entre os anos de estudo e os diferentes testes aplicados, são mantidos itens comuns em todos os cadernos de teste, ou seja, existem itens do 5<sup>o</sup> ano iguais, tanto na prova do 5<sup>o</sup> ano como na do 9<sup>o</sup> ano, e outros itens que foram aplicados no teste anterior. Pelo fato de os relatórios técnicos não terem sido publicados, não se têm informações de quantos itens eram comuns. No total, foram utilizados 144 itens, metade para cada área.

Para melhor visualizar as diferenças entre os testes aplicados em 1995 e 2011, é apresentado o Quadro 2.21.

**Quadro 2.21 – Comparação entre o número de itens, blocos, cadernos e itens respondidos e a área de conhecimento do teste respondido por aluno nos ciclos de 1995 e 2011**

Atributos	1995	2011
Número de itens dos 5º e 9º anos do ensino fundamental (4ª e 8ª série) e da 3ª série do ensino médio.	898	144
Número de blocos.	13	4
Número de cadernos por série e disciplina.	208	20
Número de itens respondidos pelos alunos do 5º ano (4ª série) do ensino fundamental.	30	44
Número de itens respondidos pelos alunos do 9º ano (8ª série) do ensino fundamental e da 3ª série do ensino médio.	39	52
Teste respondido por aluno.	Língua Portuguesa <i>ou</i> Matemática	Língua Portuguesa <i>e</i> Matemática

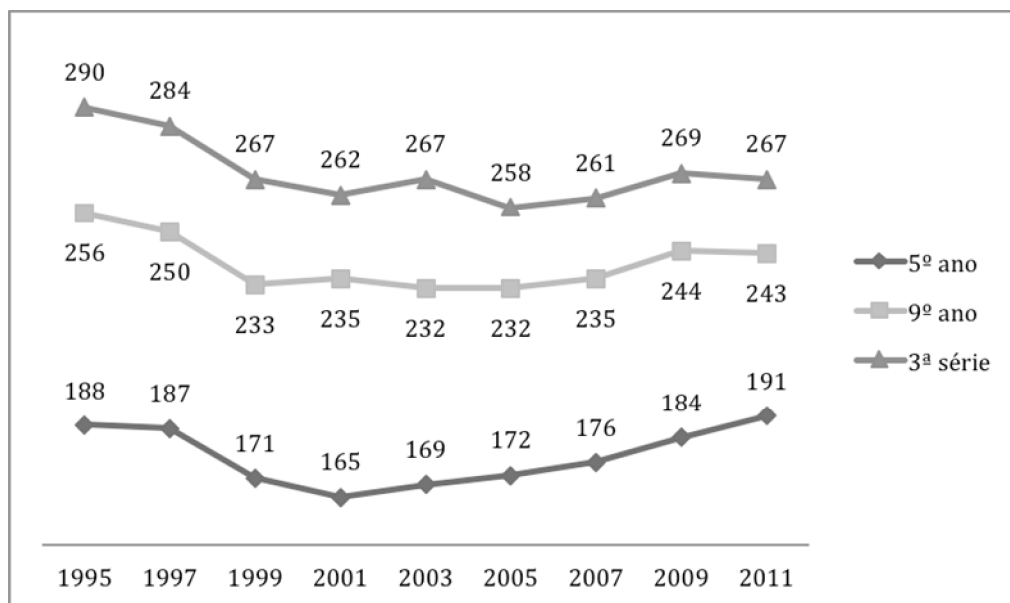
Fonte: INEP, 1998, 2002, elaboração do autor.

Pelos dados apresentados, percebe-se grande variação das características de cada um dos testes. Destaque-se que a quantidade de itens testados caiu, para os anos escolares testados, cinco vezes, entre 1995 e 2011, e aumentou em torno de 40% o número de itens que cada aluno passou a responder. Apesar dessas alterações, nenhum estudo foi apresentado para demonstrar que essas diferenças não causaram problemas para a validade e a comparabilidade do teste.

Essas modificações, de alguma forma, se refletiram nos estudos voltados para a interpretação da escala do Saeb, desenvolvidos com o objetivo de dar um significado pedagógico aos resultados numéricos obtidos. Para detalhar essas alterações, foi construído o Apêndice A desta pesquisa, no qual são discutidos o processo de construção da interpretação da escala do Saeb, os problemas surgidos e as propostas para a normatização do processo.

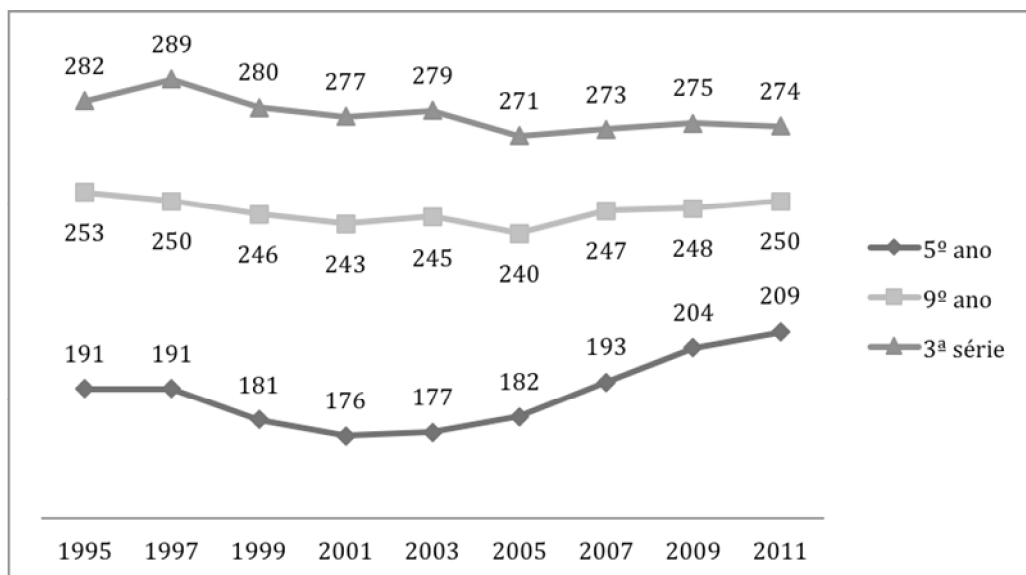
### **A evolução do desempenho do sistema educacional**

Os Gráficos 2.11 e 2.12 mostram como variaram as proficiências médias em Língua Portuguesa e Matemática, no Saeb e na Prova Brasil, entre 1995 e 2011, respectivamente. Nos gráficos são utilizados como base os nove anos de escolaridade, conforme instituído pela Lei 11.274/2006.



**Gráfico 2.11 – Evolução da proficiência média em Língua Portuguesa no Saeb/Prova Brasil – escolas públicas – 1995 a 2011**

Fonte: INEP, com elaboração do autor.



**Gráfico 2.12 – Evolução da proficiência média em Matemática no Saeb/Prova Brasil – escolas públicas – 1995 a 2011**

Fonte: INEP, com elaboração do autor.

Nos dois gráficos, é possível perceber os diversos momentos da variação da proficiência obtida através dos testes. Para começar a análise das proficiências dentro do contexto em que foram colhidas, é preciso ter em mente que em toda a média podem se esconder muitas disparidades, portanto ela não é a medida mais transparente de todas.

Serão discutidos três períodos: o que vai de 1995 a 2001, o ano de 2003 e o que vai 2005 a 2011.

O período entre 1995 e 2001 foi marcado pela queda constante das proficiências em todos os anos escolares e nas duas áreas, com exceção da proficiência média em Matemática, na 3ª série do ensino médio, em 1997. Esse período coincide com os dois governos Fernando Henrique Cardoso e foi marcado por uma grande discussão envolvendo a queda da qualidade da educação pela significativa inclusão das camadas mais pobres da população na escola. Apesar de ser uma hipótese plausível, não se sabe de nenhum estudo que tenha feito essa análise. No entanto, isso poderia explicar a queda da proficiência nos anos iniciais do ensino fundamental, mas não a queda nos anos finais e no ensino médio, a menos que se imagine que o movimento de entrada na escola foi acompanhado por outro de retorno a ela. Por outro lado, esse período foi marcado por muitas alterações nos testes, com a inclusão da Matriz de Referência, em 1997, e de duas alterações posteriores, em 1999 e 2001. Como também não foram divulgados os documentos detalhando o plano amostral para essas aplicações, pode ser que tenha havido problemas também.

Em 2003, já no primeiro governo de Luiz Inácio Lula da Silva, o teste não sofre nenhuma alteração e nos anos iniciais do ensino fundamental há uma melhora, que é prontamente creditada ao novo governo. Mas como explicar ter havido, também, melhoras nos dois outros anos escolares testados?

Em 2005, começa a Prova Brasil, e os alunos de todas as escolas são testados. No caso do ensino médio, a medida continua sendo feita com base na amostra. Também é o início da trajetória ascendente nas proficiências para os anos iniciais e finais do ensino fundamental, apesar de nos anos finais o crescimento ser de magnitude menor. Para o ensino médio, o percurso é errático. Em 2007, no mesmo ano do Ciclo de 2007, entra em cena o Ideb, e o MEC divulga em seu *site*<sup>50</sup> um simulado para ajudar as professoras e as escolas a se prepararem para o teste, e os gestores recebem, meses antes da realização do teste, o valor do Ideb de 2005 das redes e das escolas e as metas que deveriam ser atingidas no Ciclo de 2007. O que esse aumento das proficiências está significando? A escola estaria melhorando a qualidade do ensino oferecido e isso se refletiria na elevação das proficiências medidas? O que pode estar ocorrendo?

Sem pretender aqui ser exaustivo na discussão sobre o que de fato pode estar ocorrendo, mesmo porque esse não é o tema central deste trabalho, algumas hipóteses podem ser levantadas. A primeira é que, encerrado o fluxo acelerado de entrada na escola, tenha

---

<sup>50</sup> Disponível em: <[http://portal.mec.gov.br/index.php?option=com\\_content&view=article&id=16640&Itemid=1109](http://portal.mec.gov.br/index.php?option=com_content&view=article&id=16640&Itemid=1109)>.

havido uma acomodação e com isso tenha sido possível melhorar o atendimento aos alunos. Outra é que o aumento no valor dos programas de transferência de renda, capitaneados pelo Bolsa-Família, e o aumento das famílias beneficiadas tenham contribuído para retirar da miséria muitas crianças e jovens, com a conseqüente melhoria na aprendizagem. O mesmo pode ter acontecido devido ao aumento da renda da população, fruto do período de crescimento econômico vivido nos últimos anos. Até aqui, referimo-nos aos efeitos externos que impactam a aprendizagem, positivos e desejáveis. Outra hipótese é que o aumento da escolaridade para nove anos, começado em 2006, por colocar a criança mais cedo em contato com um ambiente de aprendizagem, possa também estar contribuindo para o aumento da proficiência, mas, nesse caso, por ser uma medida recente, a contribuição ainda deve ser mínima.

Pode ser também que estivesse havendo um problema de gestão nas redes e nas escolas e, com o impacto das pressões por melhores desempenhos, os gestores tenham desenvolvido estratégias para dar mais apoio às escolas, e esse apoio tenha chegado até as professoras, estando eles, assim, mais preparados para melhorar a aprendizagem de seus alunos. Por fim, uma hipótese, que se lê nas entrelinhas das matérias jornalísticas e nas declarações de alguns gestores educacionais, é a de que os gestores escolares e as professoras estivessem fazendo o corpo mole, algo que o Ideb estaria denunciando, e, a partir da divulgação do fato, eles começaram a se esforçar mais, redundando em melhores resultados. Até aqui todas as hipóteses levantadas refletiriam em benefício das crianças e dos jovens.

Mas existem também hipóteses que trariam prejuízos ao ensino. Pode ser que por conta da pressão por melhorias dos indicadores das escolas, esteja havendo uma redução do currículo, que estaria se moldando às Matrizes de Referências dos testes. Além disso, por conta dessa mesma pressão, pode estar levando as escolas a se preparar para o teste, principalmente através da aplicação de simulados, aproveitando os modelos disponibilizados pelo MEC. Esses dois fatores também podem estar na raiz da explicação para o crescimento mais acelerado nos anos iniciais do que nos finais.

Como nos anos iniciais funciona a unicidade, é mais fácil para a professora organizar seu tempo e, dessa forma, preparar melhor seus alunos para o teste, algo mais complicado na etapa seguinte, com várias professoras de diferentes áreas. Essas práticas, além de não colaborarem para a aprendizagem, estariam contribuindo para a homogeneização do ensino. Ao longo deste trabalho, diversas evidências serão trazidas para justificar a grande preocupação com esses fatos.

Destaque-se, também, uma preocupação pouco explorada nos estudos sobre os testes aplicados. Oliveira (2011) comenta, em sua pesquisa junto a equipes escolares de algumas escolas do Distrito Federal, sobre a execução da Prova Brasil, em 2009, que o tempo de aplicação dela é considerado longo pelos alunos, que ficam sujeitos a rotinas diferentes das que estão acostumados no dia a dia escolar. Além disso, Oliveira destaca que a tarefa de responder aos testes, contendo 44 itens para o 5º ano do ensino fundamental e 52 para o 9º ano do ensino fundamental e a 3ª série do ensino médio, em um período que dura duas horas e meia, requer maior disciplina e concentração. Além disso, na maior parte das vezes, os estudantes não são motivados, não compreendem a razão de serem obrigados a participar do teste e respondem aos itens de qualquer forma, fazendo com que a avaliação perca seu significado. Esses fatores devem ter influência sobre o desempenho, ainda que aos dados, resultantes de comportamentos inteiramente descompromissados, sejam aplicados procedimentos estatísticos complexos (VIANNA, 2003, p. 15). Como os fatores apontados aqui tendem a ser mais presentes nos alunos mais velhos, principalmente entre aqueles que estão concluindo o ensino médio, uma parte de seu desempenho deve ser creditado a esse descompromisso. De qualquer forma, essas questões deveriam ser discutidas pelo Inep quando da apresentação dos dados, destacando os limites dos testes avaliativos.

### **Provinha Brasil**

A Provinha Brasil retoma 18 anos depois, uma tradição inaugurada em 1990 pelo Saeb de testar os alunos nos anos iniciais do ensino fundamental para medir o processo de alfabetização. Além disso, com essa iniciativa, o governo federal, acompanha iniciativas semelhantes de diferentes governos estaduais que já vinham aplicando testes para esse segmento do ensino fundamental há algum tempo, como é o caso de São Paulo, desde 2003, com o Saresp; de Minas Gerais, em 1992, 1994 e 1996 com o Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais e desde 2005 com o Proalfa; e do Ceará, desde 2007, com o Spaece-Alfa. Nesse caso, a lógica se inverteu e foi o governo federal que acompanhou a iniciativa dos Estados.

Do desenvolvimento dos instrumentos da Provinha Brasil, participaram, além de pesquisadores do Inep e de consultores convidados, o Centro de Alfabetização Leitura e Escrita (Ceale), da Universidade Federal de Minas Gerais (UFMG); o Centro de Políticas Públicas e Avaliação da Educação (CAEd), da Universidade Federal de Juiz de Fora (UFJF);

o Centro de Estudos em Educação e Linguagem, da Universidade Federal de Pernambuco (UFPE); o Centro de Formação Continuada de Professores, da Universidade de Brasília (UnB); o Centro de Formação Continuada, Desenvolvimento de Tecnologia e Prestação de Serviços para as Redes Públicas de Ensino, da Universidade Federal de Ponta Grossa. Portanto, diferentemente do Saeb, em que apenas uma única instituição coordenou as atividades de seu desenvolvimento, a Provinha Brasil ampliou bastante o número de envolvidos, mostrando o quanto a *expertise* na área de avaliação educacional avançou no Brasil.

A Provinha Brasil testa os alunos no segundo ano de escolaridade, nas áreas de Língua Portuguesa e Matemática, no início e no final de cada ano, utilizando os chamados Testes 1 e 2, respectivamente, em cada um desses momentos. A primeira medida é diagnóstica, e a segunda mede o progresso obtido ao longo do ano. Juntos com os testes são entregues vários materiais informativos destinados aos gestores. Para as professoras, existem materiais que tratam da aplicação do teste e de sua correção e outro com reflexões sobre a prática docente. O Ministério da Educação encaminha os dois testes, conforme o número de matrículas constante no Censo Escolar, e os materiais que os acompanham diretamente para todas as Secretarias de Educação que manifestarem interesse em aplicá-los.

Em sua primeira aplicação, no começo de 2008, o Teste 1 era composto por 27 itens, sendo que 24 deles eram de múltipla escolha, divididos em três tipos: aqueles em que a professora lia totalmente o item e os alunos marcavam a alternativa correta; outros, em que a professora lia uma parte do item, o aluno a outra e, a partir daí, assinalava a resposta correta; por fim, itens que o aluno lia sozinho e marcava a alternativa correta. Para os outros três itens, em dois deles a professora fazia um ditado com uma palavra diferente (õbolaö e õsorveteö) e no último item o ditado era uma frase (õO menino brinca no quintalö). Para os itens de múltipla escolha foram fornecidos os gabaritos e para os três itens relativos ao ditado foi fornecida uma grade de correção com seis possibilidades, em que três delas indicavam possibilidades consideradas corretas. O *Guia de correção* informava que existiam cinco níveis de proficiência, utilizados para classificar os alunos segundo seu estágio de alfabetização, numerando-os de 1 a 5, e a cada um deles estava associado um número de acertos dos itens de múltipla escolha: Nível 1, até 13 acertos; Nível 2, de 14 até 17 acertos; Nível 3, de 18 a 20 acertos; Nível 4, de 21 a 22 acertos; Nível 5 de 23 a 24 acertos. O mesmo instrumento informava que deveria ser dada atenção especial às crianças que estivessem nos Níveis 1 e 2, de modo que, no final do ano, quando seria aplicado o Teste 2, essas crianças pudessem atingir os Níveis 3 e 4, respectivamente. Para cada nível, o guia apresentava uma descrição



das habilidades que os alunos classificados naquele nível deveriam dominar. Por exemplo, para o Nível 1, para as crianças que tiveram até 13 acertos:

Neste nível, encontram-se os alunos que estão em um estágio muito inicial em relação à aprendizagem da escrita. Estão começando a se apropriar das habilidades referentes ao domínio das regras que orientam o uso do sistema alfabético para ler e escrever. Sabem, por exemplo:

- Identificar o valor sonoro das partes iniciais e/ou finais de palavras (algumas letras ou sílabas), para adivinhar ou ler o restante da palavra;
- Reconhecer algumas letras do alfabeto e iniciar a distinção das letras de desenhos e outros sinais gráficos;
- Reconhecer gêneros textuais mais próprios do contexto escolar e mais familiares, com base em características gráficas (INEP, 2008a, p. 5).

O problema de uma classificação como essa é enquadrar um processo tão dinâmico e complexo, como é a alfabetização, em uma determinada categoria, descrevendo-o a partir de itens construídos com base em uma Matriz de Referência que não necessariamente reflete o currículo praticado na escola e aquele definido pelo sistema de ensino em que a escola está inserida. Além disso, o processo, para a professora, parece uma mágica: como associar um número de acertos de itens, independentemente do item que o aluno acertou, com uma etapa do processo de alfabetização que apresenta uma descrição tão fechada? Assim, apesar de na Provinha Brasil a própria professora aplicar a prova, ficar com esta para compreendê-la melhor, corrigi-la e iniciar um processo de análise sobre o significado do desempenho dos seus alunos, tudo isso é feito a partir de um planejamento que não foi realizado por ela, a partir de itens que ela não criou e usando uma interpretação apresentada a ela de forma incompleta. Ou seja, as professoras se encontram tão à margem desse processo quanto daquele utilizado pelo Saeb.

Além de apresentar a descrição do nível, o guia apresenta algumas considerações e sugestões de atividades para serem desenvolvidas em sala de aula, para auxiliar o trabalho da professora no sentido de elevar o desempenho dos alunos de forma a alcançar níveis mais altos, quando da aplicação do Teste 2, no final do ano.

No segundo semestre de 2008, o Inep disponibilizou o Teste 2 e junto com ele um novo *Guia de correção* com informações sobre o novo teste (INEP, 2008b). O novo teste também continha 24 itens de múltipla escolha, mais dois itens envolvendo ditado de palavras (õbananaõ e õtartarugaõ) e outro de uma frase (õA garota perdeu o seu gatoõ). Também foram

disponibilizados o gabarito para a correção dos itens de múltipla escolha e a grade de correção para o ditado, mas dessa vez com quatro possibilidades, sendo que duas delas indicavam possibilidades consideradas corretas. Como os itens utilizados apresentavam características psicométricas diferentes daquelas usados no Teste 1, a quantidade de acertos vinculados a cada um dos cinco níveis utilizados também mudou: Nível 1, 9 acertos; Nível 2, de 10 a 15 acertos; Nível 3, de 16 a 17 acertos; Nível 4, de 18 a 21 acertos; Nível 5, de 22 a 24 acertos. No restante, o guia também trazia uma descrição para cada nível, assim como considerações e sugestão de atividades para as professoras aprimorarem o desempenho de seus alunos. Da mesma forma que aconteceu ocorreu quando da aplicação do Teste 1, novamente as professoras aparecem como coadjuvantes nesse processo, e o fato de ter havido mudança na quantidade de acertos de cada nível pode ter trazido maior incompreensão sobre o processo.

Para 2009, não houve modificação no número de itens, seguindo-se o mesmo esquema utilizado para os Testes 1 e 2, de 2008, mudando-se, é claro, os itens. Com isso, os itens, por terem parâmetros psicométricos diferentes, acabaram fazendo com que o número de acertos que caracteriza cada nível também mudasse para valores diferentes daqueles utilizados em cada um dos testes do ano anterior. Outra alteração que pode ter trazido uma nova dificuldade para as professoras que participaram da aplicação de 2008 é que a descrição dos níveis foi modificada e, além disso, a descrição do mesmo nível do Teste 1 para o Teste 2 modifica-se. Como exemplo ilustrativo, utiliza-se o mesmo Nível 1 do Teste 1, que em 2009 recebe a seguinte descrição:

[...]

[os alunos] Sabem, por exemplo:

- Identificar o valor sonoro das partes iniciais e/ou finais de palavras (algumas letras ou sílabas), para adivinhar o restante da palavra;
  - Identificar relação entre grafemas e fonemas (letra/som) com correspondência sonora única ou com mais de uma correspondência sonora;
  - Reconhecer algumas letras do alfabeto e iniciar a distinção das letras de desenhos e outros sinais gráficos;
  - Ler palavras formadas por sílabas simples (consoante monogâmica + vogal)
- (INEP, 2009a, p. 9).

Comparando a descrição dos dois anos, acrescentaram-se duas novas habilidades em 2009, e uma das de 2008 não aparece nessa nova descrição. Para os especialistas em medidas educacionais, esse fato pode ter uma explicação simples, mas, para a professora que aplicou a

prova no ano anterior e a aplicou no ano seguinte, como explicar que para o mesmo nível existem duas descrições diferentes? Nenhuma informação sobre isso é fornecida pelos documentos consultados e entregues às escolas.

Para o Teste 2 de 2009, também há variação na descrição do Nível 1, não somente em relação ao Teste 2 de 2008, mas também em relação ao Teste 1 de 2009:

[...]

[os alunos] Sabem, por exemplo:

- Identificar o valor sonoro das partes iniciais e/ou finais de palavras (algumas letras ou sílabas), para ãdivinharõ ou õlerõ o restante da palavra;
- Reconhecer algumas letras do alfabeto e iniciar a distinção das letras de desenhos e outros sinais gráficos (INEP, 2009b, p. 8).

Agora a variação da descrição do Nível 1 ocorre no mesmo ano, e novamente não existe uma explicação para isso nos documentos entregues às professoras.

Em 2010, os mesmos problemas relatados aqui se repetem: diferença nas descrições dos níveis de um ano para outro e entre o Teste 1 e o 2. A diferença é que nesse ano os três itens relativos ao ditado foram excluídos do teste, sem que nenhuma explicação tenha sido dada.

Em 2011, sob o argumento de que os alunos estavam aparentando cansaço devido ao tamanho do teste, o número de itens testados cai de 24 para 20. Só que os mesmos problemas persistem. Nesse ano, no segundo semestre, testou-se pela primeira vez, além de Língua Portuguesa, Matemática, usando-se no teste 20 itens de múltipla escolha, com itens de apenas dois formatos: ou eram totalmente lidos pelo aplicador, ou parcialmente lidos por ele. Não havia itens, como em Língua Portuguesa, em que sua leitura ficava totalmente a cargo do aluno, isso provavelmente porque se tinha como objetivo evitar que problemas de compreensão do texto não prejudicassem as medidas relativas à alfabetização matemática.

Pelo exposto, apesar de se deixar a aplicação, a correção e a análise dos dados nas mãos das professoras, as modificações em sua metodologia fazem com que os testes avaliativos continuem com sua aura de mistério. O controle de quase todo o processo continua nas mãos do governo federal, com exceção dos resultados, que ficavam nas mãos de cada rede de ensino.

A forma como a Provinha Brasil é utilizada depende de cada rede. Em algumas, assim que os instrumentos são recebidos, eles são enviados para as escolas, que decidem quando

devem aplicá-los e como utilizam seus resultados. Em outras redes, repete-se o que o governo federal faz com o Saeb: a Secretaria marca a data de aplicação, percorre as escolas aplicando os testes, coleta as folhas de resposta, faz a correção, analisa os resultados e só depois discute com as escolas os resultados, normalmente fixando metas que devem ser atingidas. Assim, um instrumento que deveria ser autonomamente utilizado pelas escolas pode, em alguns casos, repetir o esquema do Saeb e se transformar em mais um teste avaliativo externo, com o objetivo de controlar a escola e suas professoras.

Com o Pacto Nacional de Alfabetização na Idade Certa (Pnaic), que será discutido mais à frente, a Provinha Brasil passa a ser obrigatória em todos os municípios que aderiram ao programa. Além disso, os resultados de cada rede devem ser inseridos em um *software* que será disponibilizado pelo Inep e que permitirá ter um panorama, escola por escola, do desempenho dos alunos.

### ***O Índice de Desenvolvimento da Educação Básica (Ideb)***

O Inep, a partir de proposições de seus dirigentes, apresentou, pelo menos em duas oportunidades, propostas de indicadores educacionais que levassem em consideração o desempenho dos estudantes no Saeb e na Prova Brasil.

O primeiro indicador, o Índice de Qualidade Educacional (IQE) (ARAÚJO; CONDÉ; LUZIO, 2004), foi apresentado em 2004 e referia-se apenas ao ensino fundamental. Pela proposta, o indicador seria formado por três parcelas: o componente atendimento escolar (CAE), relacionado à taxa de atendimento escolar; o componente adequação idade-série (CAIS), relacionado à taxa de distorção idade-série; e o componente desempenho educacional, (CDE), relacionado ao desempenho dos alunos em Matemática e Língua Portuguesa, com base nos resultados do Saeb. Para o cálculo do CDE, estipulou-se um desempenho mínimo considerado satisfatório para cada ano escolar avaliado e para cada área do conhecimento (Matemática e Leitura: 200 pontos para a 4ª série e 300 para a 8ª) calculava-se a diferença entre esse mínimo e o desempenho apurado no teste, sem que se justificasse a razão da escolha da pontuação para representar o conhecimento mínimo satisfatório (se é que possível afirmar que existe algum). Essas diferenças, a seguir, eram ponderadas, também sem apresentar as razões do critério de ponderação adotado. O CDE seria então assim calculado:

$$CDE = \frac{3 \times M4 + 3 \times P4 + 2 \times M8 + 2 \times P8}{24},$$

em que M4 e M8 são as diferenças em Matemática para as 4ª e 8ª séries, respectivamente, e P4 e P8 correspondentes às diferenças em Leitura para as mesmas séries. O cálculo final do IQE também envolvia uma ponderação entre os três componentes e novamente não se explicitam os critérios de ponderação utilizados. O IQE seria assim calculado:

$$IQE = \frac{4 \times CDE + 2 \times CAIS + CAE}{7}$$

O IQE era expresso por um número que variava entre 0 e 100. Esse indicador nunca foi oficialmente utilizado pelo Inep, nem pelo MEC.

O Ideb (FERNANDES, 2007) foi apresentado três anos depois, em 2007. Em suas justificativas, o autor aponta uma série de premissas que balizaram o desenvolvimento do indicador e demonstra matematicamente as hipóteses apresentadas. Segundo sua proposta, o Ideb, expresso em um número que varia entre 0 e 10 seria composto por duas parcelas,<sup>51</sup> deve ser calculado da seguinte forma:

$$IDEB = N \times P,$$

sendo  $\bar{N}$ , a nota média padronizada, construída a partir de uma transformação linear, que toma como limites superior e inferior a média verificada em 1997 nos desempenhos de Matemática e de Língua Portuguesa, acrescidos de três desvios-padrão para cima ou para baixo, e  $\bar{P}$  o indicador de rendimento, construído a partir do inverso da taxa de aprovação da etapa escolar para o ensino fundamental da 1ª à 4ª série e da 5ª à 8ª série e para o ensino médio da 1ª à 3ª série.

Com esse indicador, seria possível, segundo o autor, detectar escolas e/ou redes de ensino cujos alunos apresentem baixa *performance* e monitorar a evolução temporal do desempenho dos alunos dessas escolas e/ou redes de ensino (FERNANDES, 2007, p. 8). Além disso, o autor sugere que os governos poderiam agir, com base no indicador, no sentido de direcionar o financiamento educacional para as redes de ensino em que seus alunos apresentassem piores desempenhos. Ainda segundo ele, assegurado o financiamento, o Ideb

<sup>51</sup> Um exemplo bastante ilustrativo desse cálculo pode ser encontrado ao final do documento "Nota Informativa do Ideb 2011", produzido pelo Inep e disponível em: <[http://download.inep.gov.br/educacao\\_basica/portaI\\_Ideb/o\\_que\\_e\\_o\\_ideb/nota\\_informativa\\_ideb\\_2011.pdf](http://download.inep.gov.br/educacao_basica/portaI_Ideb/o_que_e_o_ideb/nota_informativa_ideb_2011.pdf)>.

poderia ser utilizado também para monitorar os avanços obtidos, definindo condições para a liberação dos recursos. Ideias como essa, que parecem tão lógicas dentro do ambiente empresarial, onde acredita-se muito no incentivo material como base para o avanço, apareceu algumas vezes ao longo desse trabalho, como quando se apresentou o que vem acontecendo nos EUA ou as ideias do senador Darcy Ribeiro quando da apresentação do seu projeto de LDB, e vem se tornando prática corrente nas ações dos governos nacional e estaduais.

Definido o Ideb como um indicador útil e necessário, o governo federal estipula uma meta: que o Brasil alcance o valor 6,0 em 2021, um ano antes do bicentenário da Independência. Para acompanhar seu avanço, foram estabelecidas metas para as escolas de ensino fundamental e para o ensino médio como um todo, calculadas com base no desempenho da Prova Brasil e do Saeb de 2005 e nas taxas de aprovação desse ano.

Comparando as duas propostas, a do IQE e a do Ideb, percebe-se que o Ideb, como indicador, é mais robusto tecnicamente do que a proposta do IQE, oferecendo parâmetros claros para seu acompanhamento e não propondo ponderações que, com o passar do tempo, poderiam sofrer algum tipo de alteração em função dos mais variados interesses. Assim, uma grande vantagem do Ideb, em termos de políticas sociais, é sua transparência e estabilidade. Além disso, permite expressar claramente a grande disparidade existente dentro de cada Estado, entre as redes de ensino e as escolas. Em 2011, por exemplo, no Estado de Minas Gerais, observa-se grande variação entre escolas, redes e municípios, com índices que vão de 0,3 para a Escola Estadual Prof<sup>a</sup> Maria do Socorro Andrade, no município de Belo Horizonte, a 8,6 para a Escola Municipal Carmelia Malagutti, no município de Itau de Minas.

No entanto, o indicador possui algumas limitações, técnicas e políticas que precisam ser esclarecidas para se compreender que sua utilização deve ser feita de forma criteriosa e não indiscriminadamente, como vem acontecendo.

Com relação às limitações técnicas, Soares (2012) afirma que a forma como as proficiências entram no cálculo do Ideb indicam um desempenho superior em Matemática do que em Leitura, algo que testes como o Pisa não confirmam. A segunda limitação tem relação com o próprio universo testado. O autor, baseando-se nos microdados da Prova Brasil e do Saeb de 2009, registra uma ausência de alunos no dia da aplicação do teste, de 15%, 25% e 50%, respectivamente para o 5º e o 9º anos do ensino fundamental e para a 3ª série do ensino médio, algo já comentado por Vianna em ciclos anteriores e discutido neste texto. Com isso, a proficiência medida padece de alguma imprecisão, permitindo que alguma situação esteja sendo mascarada.

Outra limitação técnica também é apontada por Soares (2009), que afirma:

O valor do Ideb aumenta quando um ou os dois indicadores que o compõem aumentam. A melhoria do primeiro termo, que descreve o desempenho, ocorre sempre que há aumento da nota média dos alunos de um sistema ou de uma escola. Uma mesma média, entretanto, pode ser obtida com valores muito diferentes para os percentuais nos diferentes níveis de desempenho. Muitos desses perfis são claramente pouco equitativos, pois, para obter médias de desempenho maiores, deixam para trás alunos de baixo desempenho (p. 37).

O autor está chamando a atenção para o fato de que, como no cálculo do Ideb, utiliza-se a média de desempenho, e esta média pode ser composta por situações muito díspares, com grupos de alunos com desempenhos muito altos e outro grupo com desempenhos muito baixos. Pode inclusive aumentar a média dos alunos de maior desempenho e, ao mesmo tempo, aumentar a porcentagem de alunos com baixo desempenho e, assim mesmo, ocorrer um aumento no Ideb. Portanto, longe de identificar um problema de equidade, o indicador o mascara e acaba contribuindo para que os alunos com mais dificuldade sejam deixados de lado.

Existe outra limitação, esta relacionada com o nível socioeconômico. Duarte (2012) realizou um estudo relacionando o Ideb de 2007 e o de 2009 com a quantidade de alunos por escola, alunos cujas famílias eram beneficiárias do programa Bolsa-Família, os quais representavam 39% da matrícula total do ensino fundamental. Como resultado, Duarte encontrou correlações negativas superiores a  $-0,50$ . A autora destaca que os pobres não estão fadados ao fracasso, mas que, sem políticas sociais que permitam compensar os efeitos negativos da pobreza, os ganhos na aprendizagem tendem sempre a ser baixos. Sendo assim, classificar as escolas pelo Ideb sem considerar os efeitos socioeconômicos sobre a medida acaba sendo injusto com essa população, essas escolas e suas professoras. Mas apenas corrigir a medida, controlando-a pelo nível socioeconômico, como foi feito no relatório do Saeb de 2001, apesar de expressar com mais clareza a situação, não resolve o problema. Sabe-se que existem escolas frequentadas pelas populações pobres que conseguem superar suas limitações, mas, como não poderia deixar de ser, seu número é bastante reduzido. Além disso, um conjunto de situações especiais existe nessas escolas, afetando positivamente o desempenho dos alunos. Por diversas razões, essas situações são difíceis de serem replicadas para a maior parte das escolas públicas, em curto prazo, escolas essas que atendem, em sua maioria, as populações mais pobres.

Esses fatos, longe de mostrarem que o indicador é inadequado para o que se propõe, revelam que uma série de cuidados deve ser tomada quando da utilização do Ideb como o indicador de qualidade da educação brasileira. Além disso, como já foi proposto antes para o caso da divulgação dos desempenhos do Saeb e da Prova Brasil, devem-se deixar bem claras suas limitações no momento de sua divulgação pelo Inep.

Outra limitação, esta de ordem política, tem relação com o fato de o Ideb não contribuir para o fortalecimento de um necessário pacto federativo em matéria educacional, envolvendo os três Entes da Federação.

Para discutir essa limitação, é necessário debater a oportunidade de o governo federal transformar o Saeb, aplicado a uma amostra de escolas, na Prova Brasil, de caráter censitário. Também é necessário analisar o sentido de o governo federal dialogar diretamente com as escolas de ensino fundamental e médio sem a intermediação dos governos estaduais e municipais. Essa análise demanda uma complexa discussão que se procurará sintetizar aqui.

Não é demais lembrar que a nossa Constituição constituiu três níveis de governo independentes entre si e autônomos, delegando a cada um competências próprias e estabelecendo o regime de colaboração entre eles, conforme o Artigo 211 da Constituição Federal, de forma a garantir o desenvolvimento harmônico e sustentável do País. Dessa maneira, no caso educacional, a União é responsável pelo sistema federal de educação, incluída a gestão de sua rede de educação básica e de educação superior; os Estados são responsáveis pelos ensinos fundamental e médio e pela gestão das instituições de educação superior por eles mantidas; os municípios pela educação infantil e pelo ensino fundamental. Sendo assim, a gestão das escolas de educação básica, exceto aquelas que fazem parte da rede federal, é de responsabilidade dos Estados e municípios. São esses entes que mantêm as escolas, contratam as professoras e estabelecem os currículos. Além das atribuições anteriores, a Constituição determina à União ações supletivas, de natureza técnica e financeira, e redistributivas, para minorar as desigualdades regionais. Essas ações deveriam ser dirigidas aos entes federados, pois é mediante o relacionamento horizontal entre as três esferas que se preserva a autonomia e a responsabilidade de cada um.

Apesar de essas premissas parecerem bem claras e respeitarem o espírito dos constituintes em um momento de grande anseio por liberdades democráticas e independência do poder central, não podemos simplesmente esquecer da característica autoritária que prevalece nas relações federativas brasileiras, desde a criação de nossa Federação. Essa característica faz com que muitas vezes a União imponha sua vontade, em detrimento de um envolvimento harmonioso com os outros Entes.



Em relação às políticas educacionais, esse caráter autoritário pode ser observado, por exemplo, em dois programas. Um deles é o do livro didático. O governo federal desde 1938 está envolvido com a questão dos livros didáticos da educação básica, sendo que, nessa época, exigia-se que o livro, antes de ser publicado, recebesse autorização prévia do MEC. Com o Programa Nacional do Livro Didático, o governo federal passa a adquirir os livros, impondo regras às editoras. A partir de 1985, quando as professoras das escolas públicas passam a indicar quais livros devem ser comprados, o MEC continua direcionando a elaboração deles, pois a lista de livros encaminhados às professoras para análise é elaborada depois de uma seleção feita pelo MEC. Ainda com relação ao livro didático, mesmo não havendo um currículo nacional, esses livros acabam indiretamente direcionando qual deve ser o currículo praticado pela escola, e o MEC tem um papel de destaque nisso. Outro exemplo é o Programa Dinheiro Direto na Escola, criado em 1995, que repassa, também diretamente, às escolas recursos para a sua manutenção, em razão do número de alunos matriculados.

Nos dois exemplos, por trás desses meritórios programas, em que pouquíssimos poderiam ser contrários, esconde-se um elevado grau de autoritarismo da União perante os entes da Federação. Em nenhum dos programas, respeitam-se o planejamento e as necessidades dos Estados e dos municípios e de seus órgãos gestores educacionais, pois o diálogo ocorre diretamente com a escola, independentemente da gestão dos sistemas. Programas como esses em nada contribuem para um necessário pacto federativo, uma vez que não favorecem o diálogo construtivo, nem o compartilhamento de responsabilidades. Esses exemplos podem iluminar as causas da imensa dificuldade para o regime de colaboração entre os entes federativos se estabelecer de fato.

Oliveira (2011) traz outro ponto de vista para essas ações que chegam diretamente à escola, sem a mediação dos órgãos intermediários, ao afirmar que estes buscam promover o desempenho escolar, dentro de uma ótica de fortalecimento da regulação. Segundo a autora, existem exemplos em diversos países da América Latina e também em países centrais, sendo que, nesse caso, estão acompanhados por exigência de resultados.

Transformar o teste amostral em censitário tem a mesma gênese, no sentido exposto neste texto. Não contribui para o regime de colaboração o governo federal decidir, unilateralmente, medir o desempenho de todos os alunos da rede pública e divulgar os resultados por escola; desenvolver materiais com orientações de como a escola deve lidar e o que fazer com os resultados obtidos também. Só que, diferentemente dos dois programas discutidos anteriormente, a questão agora é mais complexa, pois o governo federal nada mais faz do que expor uma situação sem que possa desenvolver diretamente qualquer ação para

resolver os problemas apontados pelo teste: as professoras são contratadas pelos dois outros entes, que mantêm suas escolas e órgãos gestores e os currículos obedecem à legislação estadual e se relacionam com os projetos pedagógicos de cada escola. Apenas testar os alunos e divulgar os resultados não asseguram as condições necessárias para melhorar a qualidade da educação. Será que, se por trás dessas metas estipuladas unilateralmente estivesse o pensamento de que, antes dela, todos ó as professoras, as escolas e os gestores ó , estavam sendo relapsos, a divulgação dos resultados seria suficiente para mobilizá-los em prol de uma melhor aprendizagem dos alunos? Que existem problemas, não há dúvidas, pois os próprios resultados dos testes, mesmo tendo um caráter limitado, mostram isso. Mas com certeza os problemas não são apenas de gestão, em seu sentido estrito, e de falta de estímulo para as escolas e as professoras. Se os problemas fossem apenas esses, se a realidade fosse tão simples, tudo se resolveria facilmente.

Outrossim, na área de formação inicial de professoras, a União é a principal responsável, já que, constitucionalmente (art. 211, § 1º), é sua a responsabilidade pelo sistema federal de ensino, que envolve as instituições federais e privadas de educação superior. Sendo assim, tem portanto mandato legal para atuar sobre essa formação, e pouco tem feito para, por exemplo, difundir os resultados que os testes avaliativos nacionais e as pesquisas que tratam dos problemas nos cursos de formação de professoras vêm insistentemente divulgando. Além disso, diferentemente do que ocorre com as escolas, o MEC não disponibiliza para os cursos de formação inicial de professoras materiais que tratem dos resultados dos testes, mesmo sabendo, há muitos anos, que a professora, isoladamente, é considerado o principal fator para impactar diretamente a aprendizagem dos alunos. Identifica-se o problema, mas pouco se faz, além de nomear os responsáveis, para debelar suas causas.

É interessante notar que esse movimento dos testes avaliativos direcionados para os resultados por escola vai contra o espírito federativo com que o Saeb foi instituído. A Portaria MEC nº 1.795/1994, editada pelo então ministro Murilo Hingel, quase no final de sua gestão, determinava a criação de um sistema nacional e como tal previa o estímulo a outras avaliações que se articulassem com a avaliação nacional. Também previa a criação de um conselho, por meio do qual se discutiriam os rumos e os resultados desse sistema que se instituíra. Esse conselho seria composto, entre outros, pelos presidentes do Conselho Nacional de Secretários de Educação (Consed) e da União Nacional dos Dirigentes Municipais de Educação (Undime) e por um representante dos prefeitos das capitais (HORTA NETO, 2007). Essa portaria nunca foi seguida pelos governos seguintes, e o MEC, juntamente com o Inep, passa a tomar as decisões, impondo-as aos Estados e municípios.

O Ideb, como indicador, tem sido muito útil para a realização de diversos estudos. Entre eles destacam-se aqueles que identificam escolas que apesar de atender populações de baixo nível socioeconômico conseguem atingir resultados expressivos<sup>52</sup>, apontando para modelos de sucesso que poderiam ser seguidos. Também tem sido utilizado para ampliar os recursos repassados para as escolas com baixo desempenho no IDEB em valores superiores aos das demais.<sup>53</sup>

O fato é que o Ideb vem ganhando a cada dia mais destaque, e algumas ações são tomadas no sentido de superar algumas de suas limitações. Mas essas ações, ao mesmo tempo, criam uma série de novas dificuldades.

Para superar o problema da ausência de alunos nos dias de aplicação dos testes, como já tratado neste texto, que criaria imprecisões nas medidas de desempenho, os testes vêm sendo impressos com o nome dos alunos, baseado nos dados do *Censo da Educação Básica*, preenchido pelas escolas, algo que pode ter significado um aumento dos custos de impressão dos testes. Caso a ausência de alunos seja superior a 50%, o Ideb da escola ou do município em que isso ocorreu não será divulgado. Essas duas decisões trouxeram como consequência, talvez não prevista, a transformação de um teste de natureza voluntária em teste obrigatório, ainda que não de forma explícita. Esse fato caracteriza bem algo que Barroso (2005) definiu como òregulação das regulaçõesö (p. 734), na medida em que, dada a complexidade de situações, sempre serão necessárias novas formas de regular uma situação não prevista na regulação anterior. Destaque-se que a preocupação da medida é com a precisão da medição, não com a melhoria do processo educacional.

Ainda com o objetivo de aprimorar a medida, surgem na imprensa notícias sobre a possibilidade de o Inep substituir o teste do Saeb, para apurar o Ideb da 3ª série do ensino médio, pelo Exame Nacional do Ensino Médio (Enem). Em artigo opinativo publicado no jornal *Folha de São Paulo*,<sup>54</sup> em setembro de 2012, o presidente do Inep divulga que estudos nesse sentido estão sendo feitos, justificando para isso, além do aprimoramento da medida de desempenho, o fato de que a universalização dos testes para o ensino fundamental teria propiciado maior envolvimento das escolas, da família e da sociedade e o consequente aumento do Ideb. Com relação a isso, pode-se perguntar se o passo seguinte seria tornar o Enem obrigatório e com isso impor um exame de conclusão do ensino médio. Mudanças como a proposta pelo presidente do Inep não podem ser apenas decisões administrativas e

---

<sup>52</sup> Entre elas destacam-se MEC, Inep e Unicef (2007), Unicef, Inep, MEC e Undime (2008), Fundação Lemann; Itaú BBA (2012).

<sup>53</sup> As escolas de baixo desempenho recebem entre R\$ 15 mil e R\$ 75 mil, de acordo com o número de alunos matriculados, enquanto as outras recebem entre R\$ 10 mil e R\$ 37.500,00.

<sup>54</sup> Como o jornal não permite acesso direto dos não assinantes às suas matérias pela internet, esse artigo pode ser lido em: <<http://jorgewertheim.blogspot.com.br/2012/09/luiiz-claudio-costa-metodologia-de.html>>.

devem ser precedidas de um amplo debate sobre sua pertinência, devido a seus possíveis desdobramentos.

O problema das ausências dos alunos nos testes, analisado como um problema ligado à precisão da medida, como ilustrado neste texto, vem sendo discutido na literatura internacional (DARLING-HAMMOND, 2010; RAVITCH, 2010) como uma das várias tentativas de enganar o sistema, com o objetivo de obter escores mais altos nos testes, algo que está se tornando comum em testes *high stakes*. Assim, deixa-se de fora de sala de aula os alunos com dificuldades de aprendizagem, para com isso não prejudicar a medida de desempenho da escola.

Campbell (1976) já afirmava que o quanto mais se usa um indicador social quantitativo para tomar decisões, mais ele estará sujeito à corrupção e mais ele distorcerá o processo social que se pretende monitorar (p. 49). Assim, apesar de o Inep monitorar a quantidade de alunos que participam do teste, com o objetivo de impedir que se engane o sistema, essa não é a única forma de corromper a medida, como já comentado anteriormente.

### ***As ações do governo federal que utilizam os resultados das avaliações***

Neste tópico será discutido o Plano de Desenvolvimento da Educação (PDE), lançado em abril de 2007, no início do segundo governo Luiz Inácio Lula da Silva (2007-2011) e também na segunda gestão de seu ministro da Educação, Fernando Haddad. É importante discutir esse plano, pois ele abrigou em seu interior as diversas ações que foram desenvolvidas pelo MEC, procurando dar um sentido de unidade às ações desenvolvidas. Será discutida também a influência que algumas organizações, em especial o movimento social Todos pela Educação, tiveram nas definições de algumas ações do PDE, em especial ligadas à avaliação educacional.

Cada nova ação entra no PDE para complementar as anteriores que estão em andamento. Esse foi o caso do Pacto Nacional pela Alfabetização na Idade Certa (Pnaic), lançado em 4 de julho de 2012, que instituiu, entre outras coisas, um novo teste censitário, a ser aplicado aos alunos do 3º ano do ensino fundamental. Dada a importância que assume, o Pacto também será analisado, depois do PDE.

## Plano de Desenvolvimento da Educação (PDE)

O PDE constitui-se de um conjunto de ações e programas voltados para a educação básica e superior, desenvolvidos pelo MEC, muitos deles em execução há muitos anos, como o Programa de Merenda Escolar e o Programa Nacional do Livro Didático, e outros recentes, como o desenvolvimento do Ideb e o Plano de Metas Compromisso Todos pela Educação, que será discutido adiante. É importante destacar que a sigla PDE, referindo-se ao Plano de Desenvolvimento da Escola, foi utilizada por outro programa do MEC, desenvolvido no âmbito do Projeto Fundescola,<sup>55</sup> com o objetivo de melhorar a gestão da escola a partir do desenvolvimento de um planejamento estratégico envolvendo as áreas administrativa, financeira e pedagógica da escola, rebatizado depois do Plano, de PDE-Escola.<sup>56</sup>

Antes de detalhar as concepções do PDE, é importante destacar, ainda que muito resumidamente, o momento histórico em que ele foi elaborado, destacando alguns fatos importantes que aconteceram poucos anos antes, no sentido de realçar alguns aspectos que influenciaram sua concepção.

No dia 6 de setembro de 2006, ocorreu nas escadarias do Museu do Ipiranga, em São Paulo, o lançamento do Compromisso Todos pela Educação, reunindo representantes da sociedade civil, da iniciativa privada, das organizações sociais e gestores públicos, com o objetivo de apresentar à sociedade um conjunto de cinco metas para melhorar a qualidade da educação, a serem alcançadas até 2022, ano do bicentenário da Independência do Brasil (TODOS PELA EDUCAÇÃO, 2009). Essa manifestação, que teve grande destaque na mídia escrita e televisiva na época, foi fruto de uma série de discussões começadas em junho de 2005, lideradas pela organização Faça Parte ó Instituto Brasil Voluntário, com sede em São Paulo, organização que tem por objetivo estimular o voluntariado. A partir dessas discussões, foi elaborado o documento *10 causas e 26 compromissos*, de julho de 2005. Os compromissos foram as bases para o lançamento de um movimento em prol da melhoria da educação, que deveria mobilizar diversos setores da sociedade. Entre os 26 compromissos, encontravam-se: combater a repetência e a evasão; ampliar a oferta de educação em tempo integral; abrir as escolas nos fins de semana para atividades sociais, culturais e de lazer; criar fundo de projetos

---

<sup>55</sup> O Projeto Fundo de Desenvolvimento da Escola (Fundescola), desenvolvido em 1998 com recursos do Banco Mundial, tinha por objetivo melhorar os indicadores educacionais nas Regiões Norte, Nordeste e Centro-Oeste, envolvendo recursos da ordem de US\$ 1,5 bilhões. Sucedeu ao Projeto Nordeste, também financiado pelo Banco Mundial, destinado apenas à Região Nordeste, e desenvolveu uma série de ações em 450 municípios dos 19 Estados atendidos em duas etapas. Na primeira etapa, agruparam-se, utilizando a divisão administrativa do IBGE, a capital do Estado e os municípios de sua região e, na segunda etapa, os municípios da região do Estado que apresentasse o mais baixo IDH. Esse grupo de municípios era conhecido como Zonas de Atendimento Prioritárias (ZAP) e para cada uma delas desenvolvia-se um planejamento integrado com ações das três esferas de governo: União, Estado e Município.

<sup>56</sup> Mais informações sobre o PDE-Escola podem ser encontradas no *site* <[www.pdeescola.mec.gov.br](http://www.pdeescola.mec.gov.br)>. Para compreender o programa, recomenda-se FONSECA; TOSCHI; OLIVEIRA, 2004.

escolares com recursos privados; ampliar para nove anos o ensino fundamental (TODOS PELA EDUCAÇÃO, 2009, p. 13). A partir desse documento, começou a ser construído o Pacto Nacional pela Educação, agregando ao movimento instituições-chave na cena educacional, como o Conselho Nacional de Secretários de Educação (Consed) e a União dos Dirigentes Municipais de Educação (Undime), dois fóruns de discussão que congregam os secretários estaduais e municipais de educação. Em seguida, o então ministro da Educação, Fernando Haddad, começou a participar dessas discussões. As discussões avançam, o movimento ganha corpo e os compromissos, em junho de 2006, se transformam em cinco metas: 1) todas as crianças e jovens na escola; 2) todas as crianças e jovens concluindo os ciclos; 3) todas as crianças sabendo ler e escrever; 4) todos os alunos aprendendo; 5) garantia de recursos para a educação. Em seguida, antes do evento nas escadarias do Museu do Ipiranga, essas metas foram mais bem qualificadas e acopladas a indicadores:

Meta 1 ó Toda criança e jovem de 4 a 17 anos na escola. Até 2022, pelo menos 98% das crianças e jovens de 4 a 17 anos deverão estar matriculados e frequentando a escola;

Meta 2 ó Toda criança plenamente alfabetizada até os 8 anos. Até 2010, pelo menos 80%, e até 2022, 100% das crianças deverão apresentar as habilidades básicas de leitura e escrita até o final da 2ª série do ensino fundamental;

Meta 3 ó Todo aluno com aprendizado adequado à sua série. Até 2022, pelo menos 70% dos alunos deverão ter aprendido o que é essencial para a sua série. Pelo menos 70% dos alunos da 4ª série e 8ª série do ensino fundamental e da 3ª série do ensino médio deverão ter desempenho superior a respectivamente 200, 275 e 300 pontos na escala de Português no Saeb e superiores a 225, 300 e 350 pontos na escala de Matemática;

Meta 4 ó Todo aluno com o ensino médio concluído até os 19 anos. Até 2022, pelo menos 95% dos jovens brasileiros de 16 anos deverão ter completado o ensino fundamental e pelo menos 90% dos jovens brasileiros de 19 anos deverão ter completado o ensino médio;

Meta 5 ó Investimento em educação ampliado e bem gerido. Até 2010, o investimento em educação básica deverá alcançar no mínimo 5% do PIB (TODOS PELA EDUCAÇÃO, 2009, p. 18)

As metas destacam as garantias das condições de acesso e conclusão da educação básica, que, uma vez na escola, as crianças sejam alfabetizadas até no máximo aos 8 anos de idade, aprendendo o que seria adequado nos anos escolares posteriores e que sejam assegurados recursos públicos em patamar adequado, para garantir esse processo. São metas

gerais, mas mobilizadoras, colocadas dentro de um prazo razoável e simbólico, o bicentenário da Independência, agregando a cada uma delas indicadores possíveis de serem acompanhados e assim verificar o cumprimento das metas. Duas delas, as de números 2 e 3, foram depois assumidas como objetivos do MEC, como será visto mais à frente.

Outro ponto importante que esclarece a posição do movimento destaca a importância que é dada à avaliação externa:

Toda reforma ou movimento em favor da educação que não chegar à sala de aula e não alterar para melhor o que ali acontece, simplesmente, não merece existir. Em razão disso nossa preocupação básica é com a melhoria do processo aprendizagem-ensino, traduzido em resultados mensuráveis, obtidos por meio de avaliação externa (TODOS PELA EDUCAÇÃO, 2006).

Portanto, o movimento destaca que a avaliação externa deve necessariamente ser o instrumento que medirá a efetividade do processo de ensino-aprendizagem. Essa é uma visão bastante limitada do processo educacional e expõe a importância dada a instrumentos gerenciais para garantir a melhoria da educação. Além disso, destaca a medida e não a avaliação; a ênfase no resultado e não no processo; homogeniza a realidade e não trata cada realidade em sua singularidade. Que a avaliação externa, envolvendo não somente a medida mas a análise, o julgamento de valor e o desenvolvimento de ações para modificar os aspectos não desejáveis daquilo que foi medido, pode trazer impactos importantes sobre o processo educacional, não há dúvida. Mas imaginar que ela seja o único instrumento que permite fazê-lo é superestimar seu papel e subestimar outros que podem ser tão úteis quanto ela em diferentes contextos.

A partir da vitoriosa manifestação nas escadarias do Museu do Ipiranga, o movimento se organiza ainda mais e assume o nome de Todos Pela Educação. São criados um conselho de governança, composto por 19 pessoas, entre elas empresários, líderes sociais e pesquisadores da área educacional; um conselho técnico, composto por 15 pessoas, envolvendo em sua maioria pesquisadores de diferentes áreas, com destaque para a área educacional; além de atuais e antigos dirigentes educacionais, sob a coordenação de uma empresária. As atividades diárias do movimento são executadas por uma equipe de 15 pessoas, dirigidas por uma diretora executiva, e suas ações são financiadas por um grupo de mantenedores, um conjunto de cinco grandes empresas e cinco fundações ligadas a grandes grupos empresariais. O movimento, desde 2008, publica o relatório *De olho nas metas*, um

meticuloso e consistente documento que faz o acompanhamento da evolução das cinco metas estabelecidas, recebendo grande destaque na mídia quando de sua divulgação. Além disso, lança livros, realiza simpósios, realiza ações de capacitação para os jornalistas que cobrem o tema educação, produz *spots* de rádio, que podem ser baixados gratuitamente no *site* do movimento, sobre o tema educação<sup>57</sup> e desenvolve estudos como a Prova ABC, para testar a alfabetização das crianças no fim do 3º ano do ensino fundamental.

Com isso, o movimento ganha visibilidade, mobiliza a sociedade para as causas que defende e também recebe prêmios como o Prêmio Darcy Ribeiro, outorgado em 2012 pela Comissão de Educação da Câmara dos Deputados. Outro aspecto que confirma a importância crescente que o movimento vem alcançando é o fato de dois dos membros de seu conselho de governança fazerem parte da Câmara de Educação Básica do Conselho Nacional de Educação.

Como afirma Gohn, os movimentos sociais são fontes de inovação e matrizes geradoras de saberes (GOHN, 2011, p. 347), saberes esses de natureza político-social. Durante o desenvolvimento desses movimentos, são criadas redes, e elas são essenciais para compreender como as aprendizagens foram geradas e quais valores da cultura política que elas vão construindo pelo caminho. E o Todos Pela Educação vem produzindo aprendizagens e construindo uma cultura política que cada vez mais tem influenciado os gestores educacionais.

É inegável a influência que esse movimento teve sobre as ações do MEC a partir do segundo mandato de Luiz Inácio Lula da Silva, entre 2007 e 2011. Essa influência é destacada em publicação do movimento, em que se afirma que o nome dado ao Decreto nº 6.094/07, Compromisso Todos Pela Educação, que operacionaliza o PDE para a educação básica, foi em reconhecimento à sintonia existente entre as 5 Metas defendidas pelo movimento Todos Pela Educação e os objetivos do Plano de Desenvolvimento da Educação desenvolvido pelo MEC. A publicação também destaca que em consonância com o ideário do Todos Pela Educação, o PDE procura ser um projeto de nação, não apenas de um governo, cujo mandato é finito (TODOS PELA EDUCAÇÃO, 2006, p. 24).

Com a influência das ações desencadeadas pelo movimento Todos Pela Educação, o MEC lança o PDE em abril de 2007. Quatro anos depois, às vésperas da saída de Fernando Haddad do ministério, para se candidatar à Prefeitura de São Paulo, uma publicação do MEC (2011), com o intuito de fazer a apresentação do que seria esse plano, traz algumas luzes

---

<sup>57</sup> As iniciativas de capacitar jornalistas e desenvolver *spots* de rádio foram também utilizadas pelo MEC, durante a vigência do Projeto Fundescola.



sobre as concepções que nortearam a gestão de Fernando Haddad à frente do ministério, que são importantes para a compreensão sobre os sentidos dados ao PDE e às ações relacionadas à centralidade adquirida pela avaliação educacional. Segundo o documento, o PDE possui seis pilares: 1) uma visão sistêmica da educação, em contraposição a uma visão fragmentada que teria prevalecido em governos anteriores, não privilegiando nenhum nível, etapa ou modalidade de ensino; 2) a territorialidade, pois dadas as grandes diferenças regionais seria necessário definir ações, tendo por base o ordenamento territorial; 3) o desenvolvimento, pois a educação teria importância para alcançá-lo; 4) o regime de colaboração, pois nenhum ente isoladamente teria condições de resolver todos os desafios para melhorar a educação; 5) responsabilização, que o texto apresenta como uma tradução de *accountability*, afirmando que se a educação é um direito de todos e dever do Estado e da família, deve-se responsabilizar os que não garantem esse direito, sobretudo a classe política; 6) a mobilização social, fundamental para acompanhar e fiscalizar as ações educacionais, sendo que para isso aconteça deve haver transparência no trato das questões ligadas à educação e nas políticas destinadas a seu desenvolvimento.

Com relação à educação básica, o documento agrupa as seguintes ações: financiamento, formação de professoras, plano de metas e avaliação e responsabilização.

Com relação ao financiamento, o documento refere-se à criação do Fundeb e à Lei nº 11.738/2008, que estabeleceu o piso salarial nacional para as professoras, a única categoria profissional a ter um piso nacional previsto constitucionalmente.

Quanto à formação de professoras, além dos cursos regulares oferecidos pelas Instituições Federais de Educação Superior (Ifes), foi desenvolvido, em 2007, o Plano Nacional de Formação de Professores da Educação Básica (Parfor), para garantir formação superior às quase 600 mil professoras que não o possuíam, na época, ou que atuavam em áreas para as quais não possuíam licenciatura. Pelo plano, previa-se o oferecimento de cursos a distância, oferecidos pela Universidade Aberta do Brasil (UAB) e cursos presenciais. Os cursos presenciais são operacionalizados pela Plataforma Freire, na qual as professoras fazem a inscrição para os cursos oferecidos em polos avançados das Ifes, que são mantidas no interior dos Estados e com recursos dos governos locais. Para os cursos que começarão no segundo semestre de 2013, foram oferecidas 35 mil vagas e autorizadas a frequentar o curso cerca de 17 mil professoras.<sup>58</sup> A UAB atualmente oferece cursos de bacharelado, de tecnólogo, de especializações e de pós-graduação, além dos cursos de licenciatura e aperfeiçoamento. Em dezembro de 2012, estes dois últimos cursos contavam com cerca de

---

<sup>58</sup> Dados obtidos no site <[www.freire.mec.gov.br](http://www.freire.mec.gov.br)>.

161 mil alunos matriculados.<sup>59</sup> Como existem poucas informações sobre esses cursos, não se sabe muito sobre seu currículo, principalmente sobre de que forma os resultados das avaliações externas estão servindo como parâmetros para balizar a melhoria da prática docente. Mais adiante se discutirá também a eficiência desses programas de formação.

Com relação ao plano de metas, o documento refere-se ao Plano de Metas Compromisso Todos pela Educação, objeto do Decreto nº 6.094, de 24 de abril de 2007. Na verdade, o plano, apesar de ser o instrumento que operacionaliza o PDE, nem faz menção a ele. O Plano de Metas nada mais é do que a explicitação de como ocorrerão as transferências voluntárias da União para os Estados e os municípios e em que áreas o MEC prestará assistência técnica e financeira. Sob a justificativa da necessidade de implementar o regime de colaboração entre os entes federados, é estabelecido um compromisso, de que tomam parte a União, os Estados e os municípios, em torno de 28 diretrizes definidas pelo governo federal que garantiriam a melhoria da educação básica. Apesar de isto não estar explícito, sem participar desse compromisso, não existe forma de ter acesso à assistência técnica, nem aos recursos financeiros do MEC.

Entre as 28 diretrizes estabelecidas nos incisos do art. 2º do decreto, 25 delas são imposições aos sistemas de ensino em matéria administrativa, que envolvem, entre outras ações:

[...]

XII ó instituir programa próprio ou em regime de colaboração, para a formação inicial e continuada de profissionais da educação;

XIII ó implementar plano de carreira, cargos e salários, para os profissionais da educação, privilegiando o mérito, a formação e a avaliação de desempenho;

[...]

XV ó dar consequência ao período probatório, tornando o professor efetivo estável após avaliação, de preferência externa ao sistema educacional local;

[...]

XVIII ó divulgar na escola e na comunidade os dados relativos à área de educação, com ênfase no Índice de Desenvolvimento da Educação Básica;

[...]

XXVII ó firmar parcerias externas à comunidade escolar, visando à melhoria da infraestrutura da escola ou à promoção de projetos socioculturais e ações educativas;

XXVIII ó organizar um comitê local do Compromisso, com representantes das associações de empresários, trabalhadores, sociedade civil, Ministério Público,

---

<sup>59</sup> Dados obtidos no *site* <[www.painel.mec.gov.br](http://www.painel.mec.gov.br)>.

Conselho Tutelar e dirigentes do sistema educacional público, encarregado da mobilização da sociedade e do acompanhamento das metas de evolução do Ideb (BRASIL, 2007).

Claramente, o documento impõe aos sistemas de ensino uma série de atribuições que lhes deveriam ser próprias e sobre as quais a União não deveria ter nenhuma ingerência. Além disso, acaba associando o Ideb ao leme que deve guiar as ações desses sistemas, tornando-se, assim, o indicador da qualidade educacional, quando se refere apenas às medidas de desempenho de Língua Portuguesa e Matemática na Prova Brasil, e do fluxo escolar, uma parcela importante, mas não determinante, no conjunto de fatores que impactam uma educação de qualidade.

Com relação a aspectos educacionais propriamente ditos, das 28 diretrizes, no art. 2º do decreto, apenas três se relacionam com o tema:

II ó alfabetizar as crianças até, no máximo, os oito anos de idade, aferindo os resultados por exame periódico específico;

[...]

IV ó combater a repetência, dadas as especificidades de cada rede, pela adoção de práticas como aulas de reforço no contraturno, estudos de recuperação e progressão parcial;

[...]

VIII ó valorizar a formação ética, artística e a educação física (BRASIL, 2007).

Com relação a essas três diretrizes, a primeira estabelece uma meta para que a alfabetização esteja concluída até os oito anos de idade, equivalente ao final do 3º ano do ensino fundamental, algo que Soares (2010) considera possível e que garantiria, em tese, a possibilidade de desenvolvimento de novas habilidades nos anos seguintes a essa etapa de ensino. No entanto, a diretriz impõe a realização de um exame para aferir se a alfabetização ocorreu ou não, uma prática pedagógica condenável, principalmente por envolver crianças de tenra idade. Essa imposição é colocada parcialmente em prática com o lançamento do Plano Nacional de Alfabetização na Idade Certa (Pnaic), a ser comentado adiante. Quanto ao combate à reprovação, existem diversas formas de fazê-lo, e as estratégias apresentadas são algumas das possíveis, mas são apresentadas como se fossem únicas e as mais adequadas. Por fim, a última diretriz tem um caráter geral e universal e espera-se que qualquer escola, além da formação específica, também desenvolva ações na direção apontada. O que se observa é

que também fixam-se atribuições, agora para a escola, que estão longe de ser de responsabilidade da União.

O que impressiona no Decreto é que por trás de um compromisso, um simples acordo envolvendo os Entes da Federação, seja imposta a seus signatários uma série de obrigações fixadas pela União. Poder-se-ia argumentar que como o compromisso é de natureza voluntária, quem estiver em desacordo com seus termos tem a opção de não o assinar. No entanto, o art. 8º do decreto abre a possibilidade para que a União, unilateralmente, se exima de colaborar com o ente que deixar de firmá-lo. O fato é que todos os Estados e municípios, sem exceção, firmaram esse compromisso. Mesmo porque nenhum ente da Federação tem condições de dispensar os recursos federais que podem vir a ser transferidos, nem tampouco deixar de assumir o compromisso público de melhorar a educação sob sua responsabilidade.

O art. 3º do decreto deixa claro o que ficou subtendido nas diretrizes quanto ao Ideb, ao definir que a qualidade da educação básica será aferida, objetivamente, com base no Ideb e ao transformá-lo no indicador objetivo para a verificação do cumprimento de metas fixadas no termo de adesão ao Compromisso. O art. 5º diz que a adesão voluntária ao compromisso significa a concordância com o cumprimento das metas de evolução do Ideb, fixadas pelo Inep para cada escola. Assim, além de o Ideb ser o indicador da qualidade educacional, as metas fixadas pelo Inep para cada escola passam a ser obrigatórias.

O último aspecto a ser destacado do Decreto nº 6.094 tem relação com os arts. 9º e 10, nos quais se esclarece sobre o instrumento utilizado para a União balizar sua assistência técnica e financeira a cada ente, o Plano de Ações Articuladas (PAR). Esse instrumento foi elaborado por equipes do MEC que percorreram os municípios para, segundo o art. 9º, identificar as medidas mais apropriadas para a gestão do sistema, com vistas à melhoria da qualidade da educação básica. Essa equipe foi formada por diversos consultores contratados depois de seleção em chamada pública, os quais deveriam, depois de analisar as condições do município, sugerir quais ações seriam as mais adequadas para a melhoria da qualidade da educação, medida pelo aumento do Ideb. Com base no diagnóstico e na elaboração do PAR, o MEC define de que forma apoiará o município em seu esforço. É importante salientar que essa estratégia é muito semelhante àquela usada pelo Projeto Fundescola para definir como os recursos do projeto seriam distribuídos em cada município da Zona de Atendimento Prioritário (ZAP), como comentado anteriormente.

Voltando à publicação do MEC sobre o PDE (MEC, 2011), o objetivo seria atender imediatamente os 1.242 municípios e as 7.085 escolas com os mais baixos indicadores e, em

curto prazo, todos os que estejam abaixo da média nacional (MEC, 2011, p. 23). Afirma também que os convênios anteriores seriam substituídos por

Planos de Ações Articuladas (PAR), de caráter plurianual, construídos com a participação dos gestores e educadores locais, baseados em diagnóstico de caráter participativo, elaborados a partir da utilização do Instrumento de Avaliação de Campo, que permite a análise compartilhada do sistema educacional em quatro dimensões: gestão educacional, formação de professores e dos profissionais de serviço e apoio escolar, práticas pedagógicas e infraestrutura física e recursos pedagógicos (MEC, 2011, p. 24-25).

Um conjunto de intenções nobres orientadas por um instrumento draconiano. Como é possível um diagnóstico participativo quando uma série de condições são definidas *a priori*? Como fazer uma análise compartilhada do sistema educacional, se uma série de diretrizes deve norteá-la?

Pelo que foi aqui descrito, o PDE e o instrumento que o operacionaliza, o PAR, são instrumentos utilizados para regular verticalmente todo o sistema educacional, utilizando os recursos de assistência financeira da União, para obrigar os demais entes da Federação a participar desse esforço. E os resultados dos testes avaliativos, a partir das metas impostas pelo Ideb, são usados para monitorar se os termos dos acordos estão sendo ou não cumpridos.

### **O Pacto Nacional de Alfabetização na Idade Certa (Pnaic)**

O Pacto, lançado em julho de 2012 e com investimentos previstos de R\$ 3,3 bilhões, é um programa com o objetivo de garantir que todas as crianças estejam alfabetizadas até os oito anos de idade, ao final do 3º ano do ensino fundamental.

Sem pretender esgotar a discussão sobre como o tema da alfabetização aos 8 anos de idade vai ganhando espaço nas discussões nacionais, é importante destacar alguns dos momentos importantes, a maior parte deles tratados nesse texto. Em 2004, no Ceará, depois da criação do Comitê Cearense para a Eliminação do Analfabetismo Escolar, fruto de uma iniciativa da Assembleia Legislativa do Ceará, foi criado o Programa de Alfabetização na Idade Certa (Paic), envolvendo a Associação de Municípios e Prefeitos do Ceará (Aprec) e a Undime do Estado. Apesar de entre seus objetivos não constar a idade de 8 anos, o programa previa a criação de sistemas municipais de avaliação da aprendizagem e do desempenho

docente. Em 2007, a Secretaria de Educação do Ceará assume o programa para si e firma parcerias com a Unicef e a Undime. Um ano antes, em 2006, é lançado o Compromisso Todos pela Educação, o qual, entre suas metas, estava garantir a alfabetização até os 8 anos de idade, na mesma época que, em Minas Gerais, era lançado o Programa de Intervenção Pedagógica (PIP), que tinha o mesmo objetivo. Um ano depois, em 2007, o governo de São Paulo lança o Programa Ler e Escrever, para também garantir a alfabetização até aquela faixa etária.

O Pacto tem dois componentes principais: um curso de formação em serviço para cerca de 360 mil professoras em atuação no chamado ciclo de alfabetização, aquele que compreende do 1º ao 3º ano do ensino fundamental, e a aplicação de testes para verificar o sucesso da alfabetização. Os Entes da Federação que quiseram aderir ao pacto assinaram compromisso envolvendo as duas ações do programa. No total, segundo o MEC, todos os Estados e 5.540 municípios<sup>60</sup> concluíram, até dezembro de 2012, o processo de adesão ao pacto.

O curso de formação em serviço, primeiro componente do Pacto, ocorrerá em diversos momentos de 2013 e 2014, com ênfase em Língua Portuguesa no primeiro ano e em Matemática no segundo ano. O curso é elaborado e coordenado por 34 universidades públicas. Inicialmente os orientadores de estudo, professoras da rede escolhidos pelas Secretarias de Educação, na razão de um para cada grupo de 25 a 34 professoras que atuam no ciclo de alfabetização, participarão de programa de capacitação continuada de 200 horas anuais, dividido em cinco períodos anuais, durante dois anos. Na etapa seguinte, cada orientador de estudo será responsável pelo acompanhamento e pela capacitação, com duração de 120 horas, divididas em diversos períodos do ano, de uma turma de professoras alfabetizadoras. As professoras serão separados em turmas de acordo com o ano escolar em que lecionam: 1º, 2º e 3º anos e classes multisseriadas, tendo sido desenvolvidos materiais instrucionais específicos para cada ano escolar. A gestão do programa será feita por um Comitê Nacional e uma Coordenação Institucional, em cada Estado e no Distrito Federal, com atribuições estratégicas e de mobilização. Além disso, cada Secretaria, Estadual e Municipal, indica um coordenador responsável pela implementação e pelo monitoramento das ações em sua rede de ensino, servindo também de interlocutor junto ao MEC.<sup>61</sup> Está prevista a distribuição de bolsas de estudo para os orientadores de estudo, as professoras orientadoras e os coordenadores,

---

<sup>60</sup> Levantamento de fevereiro de 2013.

<sup>61</sup> Essa estratégia de serem nomeados interlocutores do MEC junto às Secretarias é uma prática que vem desde a época dos projetos financiados com recursos internacionais e se mantém sempre que um grande programa é implantado pelo Ministério. Como existem vários programas, na prática, as Secretarias acabam moldando sua estrutura à estrutura que o MEC adota. Exemplo disso pode ser encontrado em Horta Neto (2006), quando se refere a um membro da Secretaria de Educação do Distrito Federal, que afirma ser reconhecido entre seus colegas como funcionário do MEC, atuando junto à Secretaria, por conta das atividades que realizava.

mediante o Fundo Nacional de Desenvolvimento da Educação (FNDE), autarquia do MEC responsável, entre outras coisas, pelo repasse dos recursos financeiros do ministério e pelo acompanhamento de sua aplicação, e pelas bolsas de pesquisa da Capes para as professoras universitárias.

Apesar de ambicioso, essa não foi a primeira ação do MEC no sentido de desenvolver programas de formação continuada voltados para as professoras alfabetizadoras. Em dezembro de 2000, durante o segundo governo Fernando Henrique Cardoso (1999-2003), o ministério, sob o comando de Paulo Renato Souza, lança o Programa de Formação de Professores Alfabetizadores (Profa), em ato que contou com representantes de todos os Estados. O programa, com duração de 180 horas e com foco em Língua Portuguesa, começou no ano seguinte recebendo a adesão de 18 Estados e do Distrito Federal e de 1.743 redes municipais, envolvendo 4.300 professoras formadoras, as responsáveis pela multiplicação do treinamento, e atingindo 89 mil professoras, apesar de a meta inicial ter sido de 50 mil professoras (BECALLI, 2007).

Em 2004, durante o primeiro governo Luiz Inácio Lula da Silva, entre 2003 e 2007, o MEC, sob o comando de Tarso Genro, lança um programa para envolver as universidades públicas no desenvolvimento de programas de formação continuada para professoras da educação básica, o Sistema Nacional de Formação Continuada de Professores, aproveitando parte da ideia de seu antecessor Cristovam Buarque. A partir dessa iniciativa, cria-se o Pró-Letramento, um programa de formação continuada com foco na alfabetização, que é a base para o curso do Pacto. O programa focava Língua Portuguesa e Matemática, com duração de 120 horas para cada área, sendo que a formação se realizava por revezamento: terminada uma área, as professoras começavam a outra. Previa-se envolver 400 mil professoras com um investimento de R\$ 120 milhões, tendo sido atendidas até 2009 cerca de 254 mil professoras em 12 Estados e 1.055 municípios (LUCIO, 2010).

Em comum, todos os programas envolveram o desenvolvimento de materiais instrucionais, a participação das universidades, o pagamento de bolsas para os tutores ou orientadores de estudo (o valor da bolsa era variado: R\$ 100,00 no Profa, R\$ 400,00 no Pró-Letramento e R\$ 765,00 no Pacto) e o fornecimento de certificados para as professoras que participassem da formação. Nas justificativas para sua criação, em todos os programas aqui descritos foram citados os resultados do Saeb que mostravam as dificuldades das crianças no 5º ano do ensino fundamental em Língua Portuguesa e Matemática.

No entanto, o curso de formação em serviço do Pacto traz diferenças com relação aos demais. Primeiro, oferece bolsa de estudos para as professoras alfabetizadoras no valor de R\$

200,00, talvez para estimular a participação delas no programa. Segundo, as professoras alfabetizadoras são divididas em turmas de acordo com o ano escolar em que trabalham. Terceiro, traz inovações no material instrucional desenvolvido, ao incluir nesse material um caderno que discute a avaliação em sala durante o ciclo de alfabetização (MEC, 2012) e outro que trata das concepções de currículo na alfabetização (MEC, 2012a, 2012b, 2012c). Comparando esses dois tipos de publicação, observa-se que a avaliação da sala de aula aproxima-se muito dos parâmetros da avaliação externa. Além disso, na publicação sobre as concepções de currículo, apresentam-se os Direitos de Aprendizagem, um tema que ainda não foi transformado em diretriz pelo Conselho Nacional da Educação (CNE).

O segundo componente do pacto são dois testes para acompanhar a evolução da aprendizagem. A primeira ocorre no início e no final do 2º ano do ensino fundamental e será produzida pela Provinha Brasil. Até 2012, esses testes eram utilizados apenas por aquelas redes que o quisessem, mas, a partir da adesão ao pacto, todas são obrigadas a realizá-lo, enviando seus resultados ao Inep. Além disso, em 2014, haverá avaliação externa conduzida pelo Inep, para testar os alunos no final do 3º ano do ensino fundamental, último ano do ciclo de alfabetização, nomeada de Avaliação Nacional de Alfabetização, ANA<sup>62</sup>. Essa ação toma como exemplo outra ação, coordenada pelo movimento Todos Pela Educação, que organizou uma pesquisa conhecida como Prova ABC, no segundo semestre de 2011 e de 2012. O movimento dá grande destaque a essa iniciativa pois,

Até 2012, o único instrumento capaz de acompanhar o cenário nacional relativo à alfabetização das crianças foi a Prova ABC. Realizada em 2011 e em 2012, ela deixa de ser aplicada a partir de 2013, quando o MEC deverá dispor de um instrumento próprio de avaliação dos resultados do Pnaic (TODOS PELA EDUCAÇÃO; EDITORA MODERNA, 2013, p. 42).

Assim, o movimento se coloca na vanguarda dos estudos sobre a aprendizagem dos alunos e como indutor de ações de governo.

A iniciativa contou com o apoio do Inep, que forneceu parte de itens já testados para fazer a calibração dos itens da Prova ABC com a escala do Saeb; da Fundação Cesgranrio, que elaborou os itens, montou o teste, fez a análise estatística e coordenou a análise pedagógica dos resultados; e do Instituto Brasileiro de Opinião Pública e Estatística (Ibope), que aplicou o teste em uma amostra de 6 mil alunos das capitais, incluindo o Distrito Federal.

---

<sup>62</sup> Esse teste foi instituído pela Portaria INEP nº 304 de 21/06/2013, publicada em 24/06/2013 no DOU, seção I, páginas 33 e 34



Foram aplicados testes de Leitura, Matemática e Escrita e para cada uma das três áreas estabeleceram-se patamares, considerados pela Fundação Cesgranrio como adequados para aquela etapa do ensino, fixados em 175, 175 e 75 pontos, respectivamente. Os resultados divulgados pelo Todos Pela Educação, em concorrida entrevista coletiva, mostrou que em Leitura, 56% das crianças brasileiras obtiveram desempenho dentro do esperado; em Matemática foram 42%; e em Escrita, foram 53,4%. Esses resultados tiveram divulgação com grande alarde em manchetes como "Com tal base, a casa cai" (*Correio Brasiliense*, Distrito Federal), "Desastre absoluto em 2011" (A educação foi reprovada no Brasil) (*Tribuna da Imprensa*, do Rio de Janeiro), "Menos da metade dos alunos do 3º ano não conseguem realizar contas simples" (*Estado de Minas*, Minas Gerais) ou "Prova ABC, nova avaliação educacional, mostra alunos com sérias dificuldades no primeiro ciclo escolar" (*O Estado de São Paulo*, de São Paulo). Nas matérias, foi dado destaque à opinião do movimento Todos Pela Educação, de que deveria haver um foco maior nas políticas voltadas para os anos iniciais do ensino fundamental, algo que o governo federal tenta fazer um ano depois da divulgação dos resultados da Prova ABC, ao lançar o Pacto. Novamente, observa-se a grande influência desse movimento social nas ações do MEC.

Com o Pacto, os alunos dos anos iniciais do ensino fundamental, de duração de cinco anos, serão testados três vezes pelo governo federal: no 2º ano, por duas vezes, e nos 3º e 5º anos, todos de forma censitária! Um enorme esforço que vai mobilizar toda a comunidade escolar, com o objetivo de classificar as escolas e pressionar suas professoras a apresentar resultados considerados adequados, baseados em metas estipuladas unilateralmente pelo governo federal ou por integrantes de movimentos sociais, sem que a comunidade de professoras participe do processo.

Outra inovação do Pacto é a instituição da premiação por desempenho, estando previstos recursos da ordem de R\$ 500 milhões, para premiar as escolas que apresentarem os maiores progressos, podendo uma parte desse valor ser destinada à premiação da equipe escolar. Assim, o governo federal aproxima-se de outras iniciativas que também premiam a *performance* das escolas, na expectativa de que, com esse bônus, as professoras se esforcem mais para conseguir que seus alunos obtenham desempenhos melhores, como se antes elas não estivessem fazendo isso. O governo parte da mesma lógica empresarial de oferecer bônus financeiro para alavancar desempenhos, na crença de que os incentivos são a mola propulsora do desenvolvimento das organizações.

Alguns aspectos dos três programas de formação continuada apresentados merecem ser discutidos. Primeiro são programas pontuais, que têm por objetivo alinhar os docentes

com as últimas decisões em matéria de política educacional dos governos (LUCIO, 2010, p. 51).

O segundo aspecto, relacionado a uma possível razão para que programas federais de formação continuada sejam pouco efetivos, tem relação com a elevada rotatividade da carreira docente. No Apêndice B, é feita uma discussão sobre a rotatividade das professoras nas escolas das redes públicas estaduais e municipais e na rede privada. Os dados utilizados para o cálculo são baseados nos Censos Escolares do período entre 2007<sup>63</sup> e 2011, mostrando uma rotatividade média, de um ano para outro, de 30% da equipe docente das escolas públicas de ensino fundamental, ou seja, de cada 10 professoras que estão cadastradas em um determinado ano em uma escola, somente sete serão novamente docentes nessa mesma escola no ano seguinte. Em alguns Estados e em algumas redes, essa rotatividade supera 60%. Essa rotatividade chega a ser maior na rede privada, mostrando tratar-se de um problema envolvendo a carreira docente como um todo, e não problemas com a rede pública. Quando, em vez de compararmos ano a ano, comparamos um intervalo de tempo de cinco anos, entre 2007 e 2012, a rotatividade nacional média dos docentes nas escolas atinge patamares alarmantes: 43% para a rede estadual e 66% para a municipal. Com relação à rede privada, a rotatividade das professoras é de 93%.

Quando a rotatividade apresentada nesse estudo é comparada com estudos produzidos para a realidade americana, a rotatividade de professoras nas escolas brasileiras impressiona. As pesquisas nos EUA mostram que 30% das professoras novas abandonam a profissão depois de cinco anos (INGERSOLL, 2001) e que em Estados como o Texas, anualmente, entre 11% e 18% das professoras deixam a profissão e entre 6% a 10% mudam de escola, sendo que a maior porcentagem refere-se a professoras com até dois anos de formadas (HANUSHECK; KAIN; RIVKIN, 1999).

Com uma rotatividade nos patamares em que se encontram, os programas de formação continuada de professoras, complexos e de longa duração, como é o caso dos três aqui discutidos, não podem se limitar a um único ciclo, pois sempre haverá um novo contingente de professoras que não participaram desses programas. Dessa forma, os mesmos programas federais teriam que ser oferecidos por vários anos, ou então, deveriam ser de tal forma estruturados que permitissem às Secretarias continuar essa tarefa por conta própria. Nesse caso, esbarra-se em problemas de falta de pessoal especializado para levar a cabo essa tarefa.

---

<sup>63</sup> A escolha do Censo Escolar de 2007 como ano-base deve-se ao fato de que só a partir desse ano, com a coleta eletrônica de dados, as professoras foram identificadas individualmente pelo CPF. Antes de 2007, o Censo registrava Funções Docentes, já que a mesma professora poderia atuar em várias escolas de diferentes redes de ensino.

Ao analisar o material desenvolvido pelo pacto, ele nada mais é que um curso sobre o processo de alfabetização, com suas técnicas e estratégias, algo que deveria fazer parte de qualquer curso de formação de professoras. Como as professoras em formação não têm a possibilidade de aprofundar esse tema durante a graduação, devido a problemas estruturais de seus cursos, fica a lacuna. Não seria mais produtivo o MEC atuar mais fortemente nos cursos superiores de formação de professoras? Se isso fosse feito, a lacuna seria diminuída, permitindo a discussão com outros entes da Federação, de programas mais simples, mas nem por isso de menor qualidade, voltados a temas específicos que pudessem ser desenvolvidos localmente. Da forma como os programas estão atualmente estruturados, a formação continuada está voltada para a complementação de falhas na formação inicial, não havendo um esforço consistente para melhorá-la.

Além disso, os números elevados de rotatividade de professoras parecem indicar que muitos programas de bonificação por resultados utilizados por diversos governos estaduais, em especial os de Minas Gerais e de São Paulo, parecem não estar apresentando resultados, no geral. Isso porque uma das premissas desses programas é que, para ter direito ao bônus, se estipula para as professoras um prazo mínimo de docência na escola que apresentou melhora no Ideb.

### ***A discussão da qualidade educacional na mídia eletrônica escrita***<sup>64</sup>

Como uma das premissas do *accountability* é que, mediante a exposição pública dos resultados dos testes de desempenho, haja uma mobilização da sociedade em busca da melhoria da escola, uma das dúvidas era como a mídia eletrônica estava tratando os resultados dos testes de desempenho aplicados aos alunos. Para isso, entre 10 de janeiro de 2009 e 11 de outubro de 2012, foram sendo colecionadas matérias disponibilizadas na internet, como artigos de jornais, revistas, *sites* e *blogs*. Isso foi feito com a utilização da ferramenta *Alerta*, do Google, em uma pesquisa com as seguintes expressões-chave: avaliação educacional, avaliação da educação, *educational evaluation*, *evaluación educativa*, *evaluación de la educación*, Prova Brasil, Provinha Brasil, sistema nacional de avaliação da educação básica, Saeb. Além disso, foi utilizado também o *clipping* de notícias do Movimento Todos pela Educação e o do Inep.

---

<sup>64</sup> Esta seção baseou-se, em parte, em um texto de Horta Neto (2013).

No total foram colecionados 5.445 *links* para matérias de internet, sendo que 2.525 destes, depois de coletados, foram descartados por dois motivos principais: não estavam mais disponíveis ou a matéria, apesar de possuir a palavra-chave, tratava de temas distintos do interesse da pesquisa. Os 2.920 *links* restantes foram categorizados em dois grupos. O primeiro grupo com 986 *links*, a maior parte originários de *blogs*, tratavam da realidade americana, sendo que 720 deles discutiam a influência dos resultados na gestão das escolas e dos sistemas de ensino, 126 tratavam dos resultados do Pisa e do Naep e mostravam, na sua grande maioria, as dificuldades que o sistema educacional americano enfrenta para melhorar o desempenho de seus estudantes, quando comparado com outros países, e os outros 100 *links* referiam-se a críticas à avaliação externa. O segundo grupo, de 2.100 *links*, deram origem ao Quadro 2.22, no qual estão classificados os textos nas seguintes categorias: *avaliações, julgamento do resultado obtido, responsável pelo resultado obtido, preparação para o teste e relação entre os resultados e a pobreza.*

**Quadro 2.22 – Tabulação das matérias coletadas no período entre 10 de janeiro de 2009 e 11 de outubro de 2012, utilizando a ferramenta “Alertas do Google”, classificadas em temas gerais e específicos e a frequência com que foram tratados**

Tema geral	Tema específico	Frequência
1. Avaliações	Saeb, Prova Brasil e Ideb	635
	Enem	493
	Avaliações estaduais	165
	Provinha Brasil	174
	Prova ABC	46
	Avaliações municipais	35
	Avaliações privadas	5
	Crítica à avaliação	0
	Subtotal	1.553
2. Julgamento do resultado obtido	Melhora da nota ( <i>desempenho bom</i> )	175
	Queda da Nota ( <i>desempenho ruim</i> )	171
	Subtotal	346
3. Responsável pelo resultado obtido	Professora e equipe escolar	169
	Recursos	86
	Equipe da secretaria	49
	Curso (Secretaria, particulares, <i>on-line</i> )	41
	Subtotal	345
4. Preparação para os testes	Secretarias	132
	Interna à escola	11
	Subtotal	143
5. Relação entre os resultados e a pobreza		36
<b>Total</b>		<b>2.423</b>

Fonte: Elaboração do autor.

Note-se que, como um mesmo texto poderia tratar de diferentes temas, a frequência total foi de 2.423, superior aos 2.100 *links* analisados.

A primeira observação, é que dos 2.100 *links* nenhum deles apresentava críticas à avaliação externa no Brasil. No máximo, o que se encontrou foram sugestões de aprimoramento ao sistema adotado, passando uma imagem de que existiria uma unanimidade com relação a ela. Em dezembro de 2012, portanto, depois do término da coleta dos dados, entrou no ar o *blog Avaliação Educacional*,<sup>65</sup> mantido pelo educador Luiz Carlos Freitas, tratando principalmente dos reflexos das lições aprendidas com as avaliações americanas e

<sup>65</sup> Disponível em: <www.avaliacaoeducacional.com>.

seus reflexos no cenário brasileiro. No caso dos *links* para as reportagens americanas, as críticas apresentadas são contundentes e trazem fortes argumentos contrários ao aumento dos testes e aos usos inadequados que se fazem de seus resultados, seguindo o que já foi discutido quando se tratou das avaliações nos EUA.

Como se observa no Quadro 2.22, o tema mais citado tratou das avaliações em geral, recebendo 1.702 registros e, dentro deles, os mais tratados referiram-se ao Ideb, ao Saeb ou à Prova Brasil, recebendo 635 referências. A maior parte desses registros trata do Ideb, informando o resultado alcançado, como o indicador é calculado, comparando-o com os anteriores e tecendo comentários sobre os desempenhos em Língua Portuguesa e Matemática. Quanto ao Saeb e à Prova Brasil, a maior parte dos textos trata da preparação para a aplicação dos testes e de suas realizações. No geral, a principal característica desses textos é ser de natureza informativa. Aqueles que fazem algum julgamento de valor estão agrupados no item 2 do Quadro 2.22.

Com 493 referências, o segundo tema mais citado envolve o Enem, com grande parte dos textos discutindo os problemas que ocorreram em suas aplicações, mas também identificando as melhores escolas, segundo o *ranking* divulgado pelo Inep, e apresentando as estratégias que foram utilizadas para a obtenção de tal resultado. Nesse grupo, encontra-se uma reportagem muito interessante, publicada pela revista *Trip* nº 203, datada de 21 de setembro de 2011, com texto da jornalista Karla Monteiro, com o título *“A pior escola do Brasil?”*. Nele, a jornalista relata sua visita à escola estadual D. Pedro I, classificada em último lugar no Exame, localizada no Estado do Amazonas, na região do Alto Solimões, e frequentada por 600 alunos do povo ticuna, uma das nações indígenas da região. A jornalista comenta os costumes daquele povo e os relatos que ouviu, falados em português não fluente, sobre os sonhos de alguns dos jovens entrevistados. Sobre a escola, afirma que as aulas são dadas por professoras, em sua maioria, ticunas, que se comunicam com os alunos na sua língua natal, utilizando livros escritos em uma língua que não é a deles, o português. A reportagem também comenta sobre um dos sonhos daqueles jovens: frequentar um curso superior, para poder ajudar seu povo a se desenvolver, sem abandonar seus costumes e sua cultura. Alguns deles conseguem atingir esse objetivo, graças ao sistema de cotas para índios da Universidade Estadual do Amazonas (Ufam), e por isso fazem o Enem. O diretor da escola, um ticuna que, segundo a reportagem, possui pós-graduação em Pedagogia, afirma que acha justo o Enem, mas ressalta que o exame não reflete nem respeita a cultura amazônica.

Apesar de esse trabalho não ter como objetivo analisar o ensino médio, nem o exame de seleção para vagas em cursos universitários, papel assumido pelo Enem, é importante o exemplo trazido por essa reportagem, pois demonstra concretamente quão inconsistente é julgar e classificar escolas utilizando para isso os resultados dos testes, desconsiderando o contexto em que a escola está localizada, além de outros fatores que são importantes para uma educação de qualidade. Além disso, é importante destacar a solução encontrada por países como a Nova Zelândia, um país com duas línguas oficiais: o inglês e o maori, a língua falada pelos índios. O sistema educacional desse país convive com dois sistemas que funcionam em paralelo e interligados, desde a educação infantil até a educação superior, um em que o inglês é a língua utilizada e o outro que utiliza o maori.

O fato de o Enem aparecer com o segundo tema mais citado reflete o destaque dado à discussão sobre o acesso à universidade, em detrimento de outras discussões sobre o ensino oferecido pelas escolas de educação básica.

Ainda com relação aos textos que tratam das avaliações, também tiveram destaque a Provinha Brasil, as avaliações próprias produzidas por Estados ou municípios e a Prova ABC, sobre a qual já se comentou antes.

Sobre a Provinha Brasil, apesar de ser um teste, à época da pesquisa, que permitia a cada escola aplicá-lo e corrigi-lo e a partir daí discutir internamente as melhores estratégias para superar os problemas encontrados, as matérias produzidas pelas diversas mídias mostram que, gradativamente, as Secretarias de Educação assumem o controle da aplicação, da apuração e da divulgação dos resultados, utilizando para isso julgamentos de valor e o ranqueamento de escolas, tendo por base os níveis definidos pelo Inep para classificar os desempenhos dos alunos.

No caso das avaliações próprias, além dos testes elaborados pelos Estados, as matérias apontaram o crescente número de municípios que utilizam essa estratégia a partir de 2011. Juntamente com os testes, os municípios desenvolvem também indicadores, tomando como base o Ideb. Foram notícia os municípios de: Rio Branco, no Acre; Tauá, Juazeiro e Pacatuba, no Ceará; Congonhas e Arcos, em Minas Gerais; Campo Grande, em Mato Grosso do Sul, talvez o mais antigo sistema municipal, que desde 1999 testa alunos dos 3º, 4º e 7º anos do ensino fundamental; Caruaru e Serra Talhada, em Pernambuco; Castelo do Piauí, no Piauí; Toledo, no Paraná; Anchieta e Rio das Ostras, no Rio de Janeiro; Rondonópolis, em Rondônia; Caxias, no Rio Grande do Sul; Valentim Gentil e São Paulo, em São Paulo. No caso do município de São Paulo, depois da posse do prefeito Fernando Haddad, foi anunciado que os testes aplicados antes não o seriam mais, e que o município utilizaria apenas os testes

do governo federal. Além dos testes produzidos pelas Secretarias Municipais, empresas privadas também estão produzindo testes e vendendo esse produto para as prefeituras ou para as escolas privadas, todos com características semelhantes aos testes da Prova Brasil. Entre essas empresas, encontram-se a Avalia, de Minas Gerais, que também é responsável por uma das avaliações do Simave, e o Instituto de Avaliação e Desenvolvimento Educacional (Inade), ligado ao grupo paranaense Pitágoras. Além deles, as redes de escolas adventistas, maristas e Dom Bosco, todas de natureza confessional, também desenvolveram instrumentos próprios para uso em suas redes.

Apesar de as notícias tratarem de um número pequeno do total de 5.568 municípios brasileiros, o que se percebe é um crescente desenvolvimento e a aquisição de testes para medir o desempenho dos alunos. O tempo gasto nas escolas, para aplicar os testes do município, do governo estadual e do federal, e na preparação dos alunos, para aumentar seu desempenho nesses testes, principalmente na forma de simulados, vem aumentando significativamente, e deve estar tomando uma parcela importante do tempo que deveria ser dedicado à aprendizagem.

O segundo tema, com 346 referências, refere-se aos julgamentos feitos sobre o valor alcançado no Ideb, conferindo, aos sistemas estaduais, municipais ou às escolas, ora referências elogiosas ao bom desempenho ou críticas e explicações para o desempenho ruim. Com base nesses textos, selecionaram-se 19 matérias que retratam, na média, os tipos de textos produzidos sobre o Ideb, buscando-se selecionar diferentes regiões, tipos de mídia e ano das publicações, dando origem ao Quadro 2.23, no qual estão transcritos trechos dos textos selecionados.



**Quadro 2.23 – Trechos de matérias publicadas em diversas mídias, em que o tema central é o Ideb, por localidade da publicação**

Localidade/Estado a que se refere o texto	Trecho selecionado do texto
1. Diversos Estados	Conheça os governadores reprovados no Ideb. Treze redes estaduais de ensino não bateram metas de crescimento no Índice de Desenvolvimento da Educação Básica, estipuladas pelo Ministério da Educação. Nos anos iniciais do ensino fundamental, Alagoas, Paraná e Rio de Janeiro estão abaixo da meta. Doze Estados não conseguiram alcançar a meta nos anos finais desta etapa: Alagoas, Amapá, Bahia, Espírito Santo, Pará, Paraíba, Rio de Janeiro, Rio Grande do Norte, Rio Grande do Sul, Rondônia, Roraima e Sergipe. No ensino médio, Alagoas, Distrito Federal, Espírito Santo, Pará, Rio Grande do Norte, Rio Grande do Sul e Sergipe não atingiram as metas. <sup>66</sup>
2. Rio Branco, AC	Ideb da capital acreana saiu de 4,1 pontos para 5 pontos (numa escala que vai até 6), fazendo com que Rio Branco passasse para a 6ª colocação entre as capitais brasileiras com melhor desempenho. <sup>67</sup>
3. Eirunepé, AM	Em 2005, a nota do Ideb da escola estadual Dom Bosco, em Eirunepé, sudeste do Amazonas, era de 2,7, numa escala que vai de zero a dez. Quatro anos mais tarde, em 2009, a nota foi de 8,7. O aumento, de 322 pontos percentuais, é o maior registrado em todo o País. A escola atende 340 estudantes dos anos iniciais do ensino fundamental e lida com crianças em situação de risco social. Mais de 70% dos estudantes vêm de famílias que recebem Bolsa-Escola. A diretora, Maria de Fátima Libânio da Silva, foi uma das dez educadoras condecoradas com a medalha da Ordem Nacional do Mérito pela presidente Dilma Rousseff, na segunda-feira, dia 21 de março de 2010, em cerimônia no Palácio do Planalto, em Brasília. <sup>68</sup>
4. Apuarema, BA	Município baiano. Apuarema tem pior nota do Ideb. O pior Ideb dos 5.404 municípios analisados em 2009. Apuarema ficou com a nota 0,5. <sup>69</sup>
5. Retribua, CE	A Educação de Reriutaba está melhorando a cada ano. Como prova disso, no Ideb de 2011, Reriutaba alcançou 5,4 nos anos iniciais, superando a meta do Ceará (4,9) e até mesmo a do Brasil (5,0). A meta seria de 4,1 para 2011, mas Reriutaba foi mais longe e bateu a meta de 2011. Nos anos finais, alcançou 4,4 e atingiu a meta de 2015 (Ceará 3,9 e Brasil 4,1). <sup>70</sup>
6. Distrito Federal, DF	Uma dose de disciplina. Uma xícara de vontade de aprender. Duas colheres de força de vontade. Meia dúzia de união. Três gotas de estudo. Quatro litros de concentração. Uma pitada de compromisso. Essa receita não tem erro. Misture todos os ingredientes e o resultado não pode ser diferente: boas notas e bom aprendizado. Porém, cada Escola tem uma fórmula secreta na hora de colocar a mão na massa. O Centro de Ensino Fundamental (CEF) Santos Dumont, de Santa Maria, garantiu o 2º lugar na disputa, e tem se destacado no Ideb desde 2005. Em 2007, ficou entre os cinco melhores. O pozinho mágico usado pela equipe do colégio, segundo o diretor Alberto Vieira do Nascimento, é a rigorosidade com a disciplina em cima dos 1,3 mil alunos do 1º ao 9º ano. Se o aluno deixar de fazer o dever de casa três vezes, chamamos a família e ele vai para a coordenação. Ou quando alguém começa a faltar demais, nós vamos atrás para saber o que está acontecendo, conta. Alberto explica que o trabalho da Escola não se baseia na inovação. Segundo ele, o bê-á-bá é o feijão com arroz que tem dado certo. O projeto pedagógico ganha ainda uma pitada de comprometimento do corpo docente. Trabalhamos fora do horário, usamos nossos celulares para funções de trabalho. <sup>71</sup>
7. Vila Velha, ES	Os cerca de 35 mil alunos do ensino fundamental da rede municipal de Vila Velha terão mais aulas de Português e Matemática. O reforço virá com o aumento da carga horária dessas disciplinas, a partir de hoje, quando começa o ano letivo. Segundo a secretária de Educação de Vila Velha, Wanessa Sechim, essa mudança foi baseada no desempenho dos alunos nas avaliações da rede e no Ideb. Alguns alunos, entre o 1º e o 5º ano, não sabem ler bem, escrever ou fazer as quatro operações matemáticas. Esse é um problema global. Por isso, decidimos ampliar a carga horária para fortalecer a aprendizagem. Para não aumentar a carga horária anual, a Secretaria de Educação reduziu o número de aulas de Artes, que poderão ser compensadas com atividades extracurriculares. <sup>72</sup>

(continua)

<sup>66</sup> Disponível em: <<http://ultimosegundo.ig.com.br/educacao/2012-08-15/conheca-os-governadores-reprovados-no-Ideb.html>>. Acesso em: 3 maio 2013.

<sup>67</sup> Disponível em: <<http://educacaonacional.iela.ufsc.br/noticia/53623/>>. Acesso em: 3 maio 2013.

<sup>68</sup> Disponível em: <[http://portal.mec.gov.br/index.php?option=com\\_content&view=article&id=16461:escola-carente-esta-entre-as-de-melhor-desempenho-do-pais&catid=372:agenda](http://portal.mec.gov.br/index.php?option=com_content&view=article&id=16461:escola-carente-esta-entre-as-de-melhor-desempenho-do-pais&catid=372:agenda)>. Publicado em 25/03/2011. Acesso em: 3 maio 2013.

<sup>69</sup> Disponível em: <<http://jornalistanery.blogspot.com.br/2010/07/municipio-baiano-apuarema-tem-pior-nota.html>>. Publicado em 10/07/2010. Acesso em: 3 maio 2013.

<sup>70</sup> Disponível em: <<http://www.panicomania.com/portal/noticia.php?pg=noticia&id=4471>>. Acesso em: 3 maio 2013.

<sup>71</sup> Disponível em: <<http://www.todospelaeducacao.org.br/comunicacao-e-midia/educacao-na-midia/10416/formulas-de-sucesso>>. Publicado em 12/09/10. Acesso em: 3 maio 2013.

<sup>72</sup> Disponível em: <[http://gazetaonline.globo.com/\\_conteudo/2012/02/noticias/a\\_gazeta/dia\\_a\\_dia/1107574-estudantes-de-vila-velha-terao-mais-aulas-de-portugues-e-matematica.html](http://gazetaonline.globo.com/_conteudo/2012/02/noticias/a_gazeta/dia_a_dia/1107574-estudantes-de-vila-velha-terao-mais-aulas-de-portugues-e-matematica.html)>. Acesso em: 3 maio 2013.

(conclusão)

Localidade/Estado a que se refere o texto	Trecho selecionado do texto
8. Itaú de Minas, MG	Escola ãnota 106 de Minas aposta em livros como chave do sucesso; instituição atribui resultado no Ideb ao incentivo cultural e às professoras; valores, incentivo à leitura, integração com a família e capacitação do profissional: estes são os ingredientes da receita de sucesso que levou a escola a alcançar a pontuação de 8,7. <sup>73</sup>
9. Maracanã, PA	O município de Maracanã tem a 3ª melhor educação do Pará. <sup>74</sup>
10. Recife, PE	Ideb: concorrência de 30 por vaga para estudar no colégio com melhor Ideb da 8ª série: Colégio de Aplicação da Universidade Federal de Pernambuco, no Recife, tem uma média de 1.800 inscritos para 60 vagas. Ideb: 8,0. <sup>75</sup>
11. Cacoal dos Alves, PI	Escolas campeãs no Piauí têm prova toda semana e alta reprovação. Ideb de Cocal dos Alves é bom, mas seria bem maior se não fosse a alta taxa de repetência. A escola Teotônio Ferreira , com 390 alunos tem Ideb de 4,5, acima da média geral do Estado, de 3,8, e do País, de 4 pontos. Pela nota da Prova Brasil, no entanto, o conceito poderia ser de até 6,6 ó acima das particulares que chegam a 5,9 ó , se não fosse a taxa de fluxo escolar (que considera repetências e evasão) ser de apenas 67%. Escola é hexaformadora de medalhista das Olimpíadas Brasileiras de Matemática das Escolas Públicas (OBMEP). <sup>76</sup>
12. Paraná	89% dos colégios têm falhas estruturais. Não é por acaso que os problemas de infraestrutura nas escolas estaduais foram apontados como uma das razões para a queda da nota do Paraná (de 4,2, em 2009, para 4,0, no ano passado) no Ideb, divulgado semana passada. <sup>77</sup>
13. Rio de Janeiro, RJ	A secretária Municipal do Rio de Janeiro comemorou o avanço do desempenho da cidade no Ideb, que passou a ocupar o quarto lugar no <i>ranking</i> das capitais brasileiras com a melhor avaliação nos anos iniciais, ou seja, entre o 1º e o 5º anos. A cidade também evoluiu na lista das capitais brasileiras com as maiores notas na avaliação dos anos finais, entre o 6º e o 9º anos, ficando em quinto lugar. Cláudia Costin afirmou que o avanço se deve a fatores como: a definição de um currículo claro e único entre as escolas da rede; a provas bimestrais, que permitem que a recuperação do estudante seja feita ao longo do ano; a um processo de aceleração do aprendizado de jovens para a redução da distorção idade/série; e a um processo colaborativo de universidades públicas do Rio na elaboração de conteúdo de apoio digital. <sup>78</sup>
14. Parelhas, RN	Resultado do Ideb em Parelhas apresenta avanço na educação. <sup>79</sup>
15. Jaru, RO	Ideb: escolas de Jaru superam metas. <sup>80</sup>
16. Rio Grande do Sul	Educação gaúcha supera média nacional no Índice de Desenvolvimento da Educação Básica. <sup>81</sup>
17. Santa Catarina	Confira o melhor e o pior desempenho do Índice de Desenvolvimento da Educação Básico (Ideb) em SC. <sup>82</sup>
18. Sergipe	Alunos sergipanos do 5º ano têm uma das piores médias em Português e Matemática. <sup>83</sup>
19. Eldorado, SP	Nota no Ideb de Eldorado sobe 93% em relação a 2007. Eldorado eleva sua nota no Ideb de 2,9 (2007) para 5,6 em 2009. <sup>84</sup>

Fonte: elaboração do autor.

<sup>73</sup> Disponível em: <<http://noticias.r7.com/minas-gerais/noticias/escola-nota-10-de-minas-aposta-em-livros-como-chave-do-sucesso-20120819.html>> Acesso em: 3 maio 2013.

<sup>74</sup> Disponível em: <<http://blogdoisaiasfilho.blogspot.com.br/2010/07/municipio-de-maracana-tem-3-melhor.html>>. Publicado em 13/07/2010>. Acesso em: 3 maio 2013.

<sup>75</sup> Disponível em: <<http://g1.globo.com/vestibular-e-educacao/noticia/2010/07/colégio-com-melhor-ideb-da-8-serie-tem-concorrença-de-30-por-vaga.html>>. Publicado em 11/07/2010. Acesso em: 3 maio 2013.

<sup>76</sup> Disponível em: <<http://ultimosegundo.ig.com.br/educacao/escolas+campeas+no+piaui+tem+prova+toda+semana+e+alta+reprovacao/n1597057061603.html>>. Publicado em 01/07/2010>. Acesso em: 3 maio 2013.

<sup>77</sup> Disponível em: <<http://www.gazetadopovo.com.br/vidacidadania/conteudo.phtml?id=1289112&tit=89-dos-colegios-tem-falhas-estruturais>>. Acesso em: 3 maio 2013.

<sup>78</sup> Disponível em: <<http://noticias.r7.com/educacao/noticias/rio-sobe-11-posicoes-e-tem-o-2-maior-crescimento-em-indice-de-desenvolvimento-da-educacao-20120815.html>>. Acesso em: 3 maio 2013.

<sup>79</sup> Disponível em: <<http://franciscodopt.blogspot.com.br/2010/07/resultado-do-ideb-2009-apresenta.html>>. Publicado em 07/10/10>. Acesso em: 3 maio 2013.

<sup>80</sup> Disponível em: <<http://www.ariquemesonline.com.br/textos.asp?codigo=15320>>. Acesso em: 3 maio 2013.

<sup>81</sup> Disponível em: <<http://www.estado.rs.gov.br/noticias/1/84679/Educacao-gaucha-supera-media-nacional-no-Indice-de-Desenvolvimento-da-Educacao-Basica>>. Acesso em: 3 maio 2013.

<sup>82</sup> Disponível em: <<http://www.clicrbs.com.br/especial/sc/horasantacatarina/19,0,3438691,Confira-o-melhor-e-o-pior-desempenho-do-Indice-de-Desenvolvimento-da-Educacao-Basico-Ideb-em-SC.html>>. Publicado em 08/08/2011>. Acesso em: 3 maio 2013.

<sup>83</sup> Disponível em: <<http://www.atalaiaagora.com.br/conteudo.php?c=25196&sb=1&t=ALUNOS+SERGIPANOS+DO+5o+ANO+TEM+UMA+DAS+PIORES+MEDIAS+EM+PORTUGUES+E+MATEMATICA>>. Acesso em: 3 maio 2013.

<sup>84</sup> Disponível em: <<http://diariodeiguape.com/2010/07/14/nota-no-ideb-de-eldorado-sobe-93-em-relacao-a-2007>>. Publicado em 14/07/2010>. Acesso em: 3 maio 2013.

Como é possível perceber pelos trechos das matérias selecionadas, todos associam a qualidade do ensino oferecido ao valor alcançado pelo Ideb, mostrando o sucesso obtido pelo MEC quando lançou o indicador. Além disso, a maior parte dos textos preocupa-se em divulgar em qual posição do *ranking* se encontra o município ou a escola, procurando oferecer a receita para o sucesso ou a explicação para o fracasso. Outrossim, alguns deles chamam a atenção pela forma como abordam o tema.

O texto 1, publicado pelo portal de notícias IG, refere-se ao desempenho de 13 Estados, responsabilizando seus dirigentes pelo não alcance das metas fixadas pelo Ideb. O teor da notícia mostra um alinhamento com o Projeto de Lei nº 7.420/2006, de autoria da deputada Federal Raquel Teixeira, que dispõe sobre a qualidade da educação básica e a responsabilidade dos gestores públicos pela sua promoção. O PL tem 14 outras proposições apensadas a ele, uma delas de 2010, de autoria do governo federal. Pelo projeto, em tramitação na Comissão Especial da Câmara de Deputados, criada para discutir a matéria, define-se a qualidade da educação a partir de nove fatores obrigatórios, envolvendo, entre outros, oferta de ensino regular de pelo menos sete horas diárias para o ensino fundamental e de pelo menos cinco horas para o ensino médio; a oferta regular de programas de formação continuada de professoras; padrões definidos de infraestrutura e funcionamento das escolas, de acordo com custo-aluno-padrão-qualidade fixado para cada etapa e modalidade de ensino. Contraditoriamente, informa que essa qualidade será aferida por processo nacional de avaliação de desempenho dos alunos. Também determina que o desempenho dos estudantes e as taxas de repetência e evasão apuradas em um período devem ser melhores que as apuradas em período anterior, e que nenhum estudante deve ter um desempenho inferior a um mínimo aceitável, não definido no projeto. Pelo projeto, o descumprimento do disposto na Lei caracteriza-se como crime de responsabilidade, ato de improbidade, e levará à suspensão das transferências voluntárias da União ao Estado ou ao município. Pelo prazo que a proposta tramita, percebe-se a dificuldade para caracterizar o que seja qualidade em educação e, a partir daí, apontar os responsáveis por não alcançá-la e as sanções decorrentes por não a ter alcançado.

O texto 3 trata do resultado obtido pelo município de Eirunepé e foi publicado no portal do MEC, talvez com o objetivo de mostrar que, mesmo nos municípios mais distantes e com situações sociais adversas, seria possível melhorias no Ideb. Para isso, destaca um aumento do indicador em quatro anos, portanto depois de dois ciclos da Prova Brasil, de 322 pontos percentuais. Graças a esse feito, a diretora da Escola estadual Dom Bosco foi

condecorada em 2010. Pelos dados do Inep, os valores para o Ideb foram de 2,7 em 2005, 3,6 em 2007, 8,7 em 2009 e 7,1 em 2011. Apesar da queda de 1,6 pontos no Ideb entre 2009 e 2011, é impressionante a evolução no indicador para essa escola. Impressionante e pouco crível. Como o desempenho dos alunos depende de uma série de fatores, em que o nível socioeconômico e a formação inicial e continuada das professoras têm papel de destaque, é como se a realidade da escola e de seu entorno houvesse se modificado por completo em curto espaço de tempo. Mesmo que toda a equipe da escola, aquela que teria sido responsável pelo baixo Ideb, tenha sido modificada nesse período por outra muito bem qualificada e com baixíssima rotatividade, imaginar que ela tenha conseguido superar as dificuldades sociais daquela população é como acreditar que o processo educacional ocorre como em um passe de mágica, que existe uma bala de prata capaz de eliminar todos os problemas que estariam causando a não aprendizagem. O que impressiona é que essa imagem não é difundida por algum órgão da imprensa, mas pelo próprio MEC. Algumas hipóteses podem ser levantadas para esse aparente sucesso do município. A mais provável é que esteja havendo um forte estreitamento do currículo, e que os alunos estejam sendo fortemente preparados para o teste, sem deixar de lado a possibilidade da ocorrência de fraudes no processo. Esses dois aspectos já foram discutidos quando da análise das avaliações nos EUA, mostrando que são uma realidade. Para confirmá-las, seria necessário pesquisar os processos desenvolvidos pela escola, e não apenas obter respostas a algumas poucas perguntas, como foi feito pela reportagem.

O texto 7, sobre o município de Vila Velha, mostra o aumento do número de aulas de Língua Portuguesa e Matemática, mudança baseada no desempenho dos alunos, realizado pela rede de ensino municipal e pelo Ideb. Como não se vai aumentar a carga horária da escola, a solução encontrada foi diminuir o número de aulas de Artes, que, segundo a secretária de educação, podem ser compensadas com atividades extracurriculares, o que é contraditório, já que não há aumento na carga horária. De qualquer forma, observa-se o quanto a preocupação com os números toma conta da discussão, indicando novamente soluções simples para problemas complexos. Sem um projeto pedagógico adequado, professoras bem formadas e participantes constantes de programas de formação, entre outros fatores, não será com a ampliação do número de aulas das duas áreas que fazem parte dos testes da Prova Brasil que o desempenho dos alunos irá ser modificado.

Voltando a analisar o Quadro 2.22, o segundo tema mais citado nas matérias colecionadas na pesquisa, com 345 referências, refere-se àquelas matérias que mostram quais seriam os responsáveis pelos resultados, fossem estes bons ou ruins. Atribui-se, em sua

maioria, às professoras e à equipe escolar a responsabilidade pelos bons resultados, associando a função docente, normalmente, a uma imagem messiânica, relacionada a um dom para educar, e à escola como o templo sagrado do saber, não como um local de disputas e conflitos. As matérias ressaltam a dedicação e o empenho das professoras e o espírito de equipe das profissionais da escola em torno do objetivo de garantir a aprendizagem dos estudantes como responsáveis pelo sucesso da escola. Esses fatores superariam todo tipo de obstáculos, principalmente a falta de recursos, os baixos salários e as condições sociais adversas das famílias dos alunos. Também para justificar resultados positivos, realça-se o aumento dos recursos e, em outros referenciais, o trabalho da Secretaria e de seu principal gestor como os responsáveis por esses resultados. Com apenas 41 citações, aparecem as referências a cursos, presenciais ou a distância, oferecidos às professoras, seja pela equipe pedagógica da Secretaria, seja por empresas contratadas pelo poder público. Com relação a esse ponto, destaca-se o fato de não haver nenhuma referência aos programas de formação inicial ou continuada, conduzidos pelo governo federal, o que é surpreendente, dada a abrangência que esses programas têm nos últimos anos. Além disso, a pouca referência a cursos como impactantes nos resultados reforça a visão messiânica da professora, em que basta apenas vontade e dedicação para superar a adversidade, como se o conhecimento nessa profissão fosse algo menos importante que o dom de ensinar e o compromisso com a causa da educação.

Quando os resultados são ruins, via de regra, nas matérias locais, culpa-se a falta de recursos e, logo a seguir, afirma-se que, apesar dessa dificuldade, haverá um esforço enorme para modificar o panorama descrito. Em matérias nacionais, que procuram fazer uma abordagem envolvendo um panorama mais geral, apontam-se principalmente problemas de gestão nos governos estaduais, em menor número, e em maior número nos governos municipais. Algumas dessas matérias chegam a comentar que apesar dos recursos destinados à educação terem aumentado, como percentual do PIB, esse aumento não se traduziu em melhorias na mesma proporção no Ideb, concluindo que isso caracterizaria um problema de gestão.

O tema *preparação para os testes* envolveu matérias que tratavam explicitamente de estratégias utilizadas pelas Secretarias ou pelas escolas, para melhorar o desempenho dos alunos na Prova Brasil, que estava prestes a ser aplicada, e com isso aumentar o Ideb. Para ter mais clareza sobre como as matérias descreveram as estratégias utilizadas pelas Secretarias estaduais ou municipais, responsáveis por 132 referências coletadas, construíram-se os dois quadros, um destacando as matérias relativas à rede estadual e outro à rede municipal.

O Quadro 2.24 destaca, para a rede estadual, um conjunto de matérias que retratam de que forma os Estados listados estão preparando seus alunos para a Prova Brasil, mostrando, para cada Estado, o número de escolas e de alunos em cada etapa do ensino fundamental e as variações observadas no Ideb de 2009 a 2011. Com isso, pode-se ter um panorama básico do contexto em que as matérias foram geradas.

**Quadro 2.24 – Tabulação das matérias coletadas no período entre 10 de janeiro de 2009 e 11 de outubro de 2012, utilizando a ferramenta “Alertas do Google”, que trata das redes estaduais e de suas ações para a melhoria do Ideb, mostrando Estado, tamanho da rede e variação no Ideb do ensino fundamental em 2009 e 2011**

Estado	Tamanho da rede**	Variação Ideb**			Ação
		ai/af*	2009	2011	
AL	269 escolas e 103 mil alunos	ai	3,3	3,4	Agosto de 2011: criação do Sistema de Avaliação de Alagoas (Saeval), com aplicação três meses antes da Prova Brasil, para preparar também os alunos para esse teste. <sup>85</sup>
		af	2,7	2,5	
AM	540 escolas e 280 mil alunos	ai	4,5	4,8	Desde 2010, a Secretaria vem preparando os alunos com a aplicação de simulados. A gestora destaca que, em 2011, ôcerca de 5 milhões de simulados foram distribuídos. Além da familiarização do aluno com questões semelhantes às que compõem o exame, todas as professoras, segundo a gerente, receberam manuais de operações pedagógicas para que as dificuldades dos alunos fossem detectadas e corrigidas. <sup>86</sup>
		af	3,6	3,9	
MA	641 escolas e 122 mil alunos	ai	4,0	4,0	Encontro, em agosto de 2011, para discutir a avaliação educacional em Larga Escala e estudar as matrizes da Prova Brasil, com o objetivo de melhorar o desempenho dos alunos. <sup>87</sup>
		af	3,6	3,6	
PA	682 escolas e 256 mil alunos	ai	3,7	4,0	Em 2011, foram oferecidos cursos em Língua Portuguesa e Matemática, voltado para as professoras, com foco no conteúdo da Prova Brasil. Eles participaram também de oficinas para ajudá-los a formular provas parecidas, para que os alunos tenham melhor desempenho nos testes. Em 2011, foram realizados dois simulados, um em junho e outro em agosto. <sup>88</sup>
		af	3,1	3,1	
PI	459 escolas e 97 mil alunos	ai	3,8	4,1	Agosto de 2011: pretende-se transformar os coordenadores pedagógicos em multiplicadores dos treinamentos que estão acontecendo em todo o Estado, para que o desempenho na Prova Brasil melhore. <sup>89</sup>
		af	3,4	3,6	
RO	379 escolas e 136 mil alunos	ai	4,4	4,7	Agosto de 2011: governador afirma que ôtemos um grande desafio: preparar bem os alunos nesses próximos três meses para a Prova Brasil, e precisamos melhorar nossa nota no Ideb, porque isso melhora também a imagem do Estado. Tenho escrito todos os dias uma palavrinha para as professoras e diretores no meu <i>blog</i> e espero que em dezembro possamos melhorar nossa nota. <sup>90</sup>
		af	3,4	3,7	
<b>Total</b>		<b>6 redes estaduais, envolvendo 2.970 escolas e 994 mil alunos matriculados</b>			

Fonte: Elaboração do autor.

Legendas: \* ai ó anos iniciais do ensino fundamental; af ó anos finais do ensino fundamental.

\*\* Dados extraídos do Data Escola Brasil (<http://www.dataescolabrasil.inep.gov.br>) e do portal do Ideb (<http://Ideb.inep.gov.br>).

- O tom cinza indica que a meta do Ideb fixada pelo Inep foi atendida para o ciclo da Prova Brasil do ano indicado.

<sup>85</sup> Disponível em: <<http://www.educacao.al.gov.br/comunicacao/sala-de-imprensa/noticias/janeiro/agosto/escolas-da-rede-publica-se-preparam-para-aplicacao-do-saveal>>. Acessado em: 4 maio 2013.

<sup>86</sup> Disponível em: <<http://www.d24am.com/noticias/concursos/-prova-brasil-envolve-mais-de-160-mil-alunos-no-amazonas/40981>>. Acesso em: 4 maio 2013.

<sup>87</sup> Disponível em: <<http://governo-ma.jusbrasil.com.br/politica/7551388/professores-de-sao-luis-participam-de-formacao-do-prova-brasil>>. Publicado em 24/08/2011. Acesso em: 4 maio 2013.

<sup>88</sup> Disponível em: <<http://diariodopara.diarioonline.com.br/N-138891-DEPOIS+DO+DESCANSO++A+VOLTA+AOS+LIVROS.html>>. Acesso em: 4 maio 2013.

<sup>89</sup> Disponível em: <[www.riachaonet.com.br/seduc-apresenta-plano-de-intervencao-pedagogica-em-picos.html](http://www.riachaonet.com.br/seduc-apresenta-plano-de-intervencao-pedagogica-em-picos.html)>. Acesso em: 4 maio 2013.

<sup>90</sup> Disponível em: <[www.riachaonet.com.br/seduc-apresenta-plano-de-intervencao-pedagogica-em-picos.html](http://www.riachaonet.com.br/seduc-apresenta-plano-de-intervencao-pedagogica-em-picos.html)>.

Pelo Quadro 2.24, observa-se que as redes citadas possuem, no ensino fundamental regular, quase três mil escolas com quase um milhão de alunos matriculados e que estão localizadas nas regiões geográficas mais pobres do Brasil. Pelos trechos destacados, percebe-se a preocupação em preparar os alunos para melhorar o valor do Ideb, utilizando principalmente simulados baseados no modelo da Prova Brasil. Na maioria dos Estados listados, nota-se que as ações começam três meses antes da aplicação do teste, sendo que um deles, Alagoas, aproveita a criação de sua avaliação própria para utilizá-la de forma a preparar os alunos para o teste nacional. Assim, parece que a preocupação não é com o aprendizado das crianças e dos jovens e, sim, elevar o valor do Ideb para, dessa forma, demonstrar que o que vem sendo feito em matéria educacional pelo governo estadual está dando resultado. Ou seja, a preocupação é com a imagem política do governo. Essa preocupação com o Ideb demonstra que, de certa forma, o governo federal está conseguindo mobilizar, pelo menos, a classe política, para melhorar os indicadores educacionais. No entanto, da forma como isso parece estar sendo feito, a melhora no indicador não estaria significando uma melhoria na aprendizagem, muito menos garantindo que ela esteja sendo significativa, no sentido de garantir a autonomia e o desenvolvimento dos alunos como seres sociais ativos na sociedade. Parece que a preocupação é adestrá-los para o teste.

É preciso deixar claro que é possível, em algumas situações específicas, utilizar simulados como uma das possíveis estratégias que a professora tem a seu dispor, que, como o nome bem diz, simulam determinada situação, para trabalhar alguns aspectos de determinada disciplina. Mas essa estratégia, para ter algum sentido e fazer diferença em aprimorar determinados aspectos do que se quer desenvolver, tem de ter relação com a realidade de sala de aula: deve ser preparada pela professora, dentro de determinado contexto, e tendo por base os conhecimentos que ela tem sobre o conjunto de seus alunos. Caso contrário, o teste acaba tendo o mesmo caráter dos simulados utilizados para preparar os alunos do ensino médio para o Enem, em seu papel de vestibular nacional. Nesse caso, os alunos buscam individualmente estratégias para obter uma classificação mais alta que seus concorrentes, garantindo, assim, uma vaga no curso e na universidade escolhidos. Passada essa situação, e tendo sucesso em seu intento, muito pouca coisa fica do que foi estudado, e o jovem, em muito pouco tempo, esquece grande parte daquilo que utilizou para se preparar para o exame. Nesse caso, não houve aprendizagem significativa, pois tudo o que foi estudado só fazia sentido para garantir a vaga. No caso de testes como a Prova Brasil, o que seria de se esperar não é a busca de uma melhor classificação, mas sim recolher informações que ajudassem a desenhar um quadro

representativo da realidade educacional brasileira, para assim poder agir no sentido de melhorá-la. Da forma como parece que os simulados estão ocorrendo, o quadro pode até parecer bonito, mas não representa, fidedignamente, a realidade que se quer conhecer.

As células do Quadro 2.24, coloridas de cinza, indicam que a meta fixada pelo Inep foi atingida, e as em branco, o contrário. Assim, para o ciclo de 2011, com exceção de Alagoas, em o todo o ensino fundamental, e do Pará, nos anos finais do ensino fundamental, todos os outros Estados atingiram suas metas. Dessa forma, parece que essa estratégia de preparar para o teste pode estar dando o resultado esperado, e, com isso, ganhando cada vez mais simpatia dos gestores. Por outro lado, não se pode deixar de considerar que também podem estar sendo gestadas ações comprometidas com a aprendizagem dos alunos. O problema é que, a cada ciclo, maiores são os valores do Ideb exigidos, pois novas complexidades vão sendo agregadas, exigindo proficiências mais elevadas nos testes. Com isso, fica cada vez mais difícil cumprir as metas fixadas, e essa estratégia de aplicação de simulados pode não ser suficiente para aumentar o Ideb.

O Quadro 2.25 apresenta alguns excertos de matérias que mostram como os municípios estão se preparando para os testes da Prova Brasil.

**Quadro 2.25 – Tabulação das matérias coletadas no período entre 10 de janeiro de 2009 e 11 de outubro de 2012, utilizando a ferramenta “Alertas do Google”, que trata das redes municipais e das ações para a melhoria do Ideb, mostrando Estado, município, tamanho da rede e variação no Ideb, no ensino fundamental, em 2009 e 2011**

Estado/Cidade	Tamanho da rede**	Variação do Ideb**			Ação
		ai/af*	2009	2011	
AM/Manaus	366 escolas e 172 mil alunos	ai	3,5	4,1	Setembro 2011: Caravana Pró-Ideb para mobilizar as escolas a elevarem o Ideb do município. Desenvolvimento de três simulados <i>on-line</i> , disponíveis para os alunos e as escolas. Para as professoras, foram disponibilizados CDs com banco de questões. Para as escolas sem internet, a equipe se desloca com 80 <i>notebooks</i> para sua realização. <sup>91</sup>
		af	2,9	3,1	
BA/Barreiras	80 escolas e 16 mil alunos	ai	3,8	4,1	Simulado em setembro de 2011 para as professoras identificarem as dificuldades dos alunos e com isso melhorarem o desempenho dos alunos na Prova Brasil. <sup>92</sup>
		af	3,7	3,5	
BA/Ipororó	26 escolas e 3 mil alunos	ai	2,9	3,7	Três oficinas, em setembro, outubro e novembro de 2011, para preparar professoras do 5º e do 9º ano, com foco nas Matrizes. O foco é a melhoria da aprendizagem, visando melhor o aproveitamento e elevar os resultados da Prova Brasil. <sup>93</sup>
		af	2,4	2,8	

(continua)

<sup>91</sup> Disponível em: <<http://semed.manaus.am.gov.br/semed-realiza-caravana-pro-Iddeb>>. Acesso em: 4 maio 2013.

<sup>92</sup> Disponível em: <<http://jornalnovafronteira.com.br/index2.php?p=MConteudo&i=4456>>. Acesso em: 4 maio 2013.

<sup>93</sup> Disponível em: <<http://educacaoitororo.blogspot.com.br/2011/10/secretaria-de-educacao-de-itororo.html>>. Acesso em: 4 maio 2013.



(continuação)

Estado/Cidade	Tamanho da rede**	Variação do Ideb**			Ação
		ai/af*	2009	2011	
ES/Vila Velha	60 escolas e 34 mil alunos	ai	5,0	4,9	Capacitação em junho de 2011: "Momento Prova Brasil Ó Oficina de Atividade", para as professoras. O objetivo é promover a reflexão e estimular o desenvolvimento de atividades que preparem melhor os alunos para as avaliações externas, como a Prova Brasil, que será aplicada em novembro. <sup>94</sup>
		af	4,0	4,0	
MA/Imperatriz	118 escolas e 31 mil alunos	ai	4,2	4,3	Secretaria de Educação capacita professoras para melhor desempenho dos alunos na Prova Brasil. Em 2011 foram realizados encontros de estudos e formação continuada com professoras de quinto e nono anos do ensino fundamental de toda a rede municipal de ensino. <sup>95</sup>
		af	3,9	4,0	
MS/Sidrolândia	12 escolas e 6 mil alunos	ai	4,1	4,6	Agosto 2011: para se prepararem para a Prova Brasil, os alunos voltam à escola no contraturno, para atividades culturais e esportivas. <sup>96</sup>
		af	3,8	4,2	
MT/Cuiabá	85 escolas e 27 mil alunos	ai	4,5	4,8	Em setembro de 2011, é lançado o programa Força Tarefa, para motivar o engajamento da comunidade escolar e superar as metas do Ideb. O programa realizou dois simulados, em setembro e outubro. <sup>97</sup>
		af	4,1	4,2	
MT Nova Mutum	7 escolas e 3 mil alunos	ai	5,3	5,4	Junho e outubro de 2011: dois simulados. O resultado permite redimensionar a prática das professoras em sala de aula, com vistas ao resultado da Prova Brasil. <sup>98</sup>
		af	--	5,3	
PA/Belém	67 escolas e 44 mil alunos	ai	3,9	4,4	Outubro de 2009: simulado organizado pelo Núcleo de Informática Educativa (Nied) da Secretaria. O objetivo do teste é verificar o desempenho dos estudantes, depois da realização de oficinas integradas junto às professoras de sala de aula e de informática da Rede, que exploraram as disciplinas especificadas pela Prova Brasil. Segundo a coordenadora do núcleo, "Percebemos a evolução dos alunos desde o primeiro simulado. Hoje eles têm tranquilidade e estão mais seguros, adquirindo responsabilidades que antes não tinham e, com certeza, estão mais preparados para a Prova Brasil". <sup>99</sup>
		af	3,5	3,7	
PB/Patos	46 escolas e 7 mil alunos	ai	3,9	4,2	Setembro de 2011: Secretaria ofereceu às professoras do 5º e do 9º ano uma formação sobre a Prova Brasil de Língua Portuguesa, durante três dias, com o objetivo de melhorar os indicadores. <sup>100</sup> Realização da 1ª Bizurada [simulado] da Prova Brasil. <sup>101</sup>
		af	3,0	3,0	
PI/São Miguel do Tapuio	44 escolas e 3 mil alunos	ai	3,2	3,9	Agosto 2011: Projeto Gestão Ideb 2011, que contém um banco de questões elaboradas e distribuídas pela Secretaria, para as escolas se prepararem para a Prova Brasil. <sup>102</sup>
		af	3,4	4,0	

(continua)

<sup>94</sup> Disponível em: <<http://www.vilavelha.es.gov.br/noticias/professores-participam-de-capacitacao-para-a-prova-brasil-843>>. Acesso em: 4 maio 2013.

<sup>95</sup> Disponível em: <<http://www.oprogreso-ma.com.br/noticiario/14200/cidade/2011/9/11/secretaria-de-educacao-capacita-professores-para-melhor-desempenho-dos-alunos-na-prova-brasil>>. Acesso em: 4 maio 2013.

<sup>96</sup> Disponível em: <[www.sidrolandianews.com.br/0,0,0,8243-39030-SIDROLANDIA+COM+REFORCO+ESCOLAR+PARA+GARANTIR+NOTA+Ideb.htm](http://www.sidrolandianews.com.br/0,0,0,8243-39030-SIDROLANDIA+COM+REFORCO+ESCOLAR+PARA+GARANTIR+NOTA+Ideb.htm)>. Acesso em: 4 maio 2013.

<sup>97</sup> Disponível em: <[www.cuiaba.mt.gov.br/noticias?id=3763](http://www.cuiaba.mt.gov.br/noticias?id=3763)>. Acesso em: 4 maio 2013.

<sup>98</sup> Disponível em: <[www.sonoticias.com.br/noticias/3/129888/mais-de-400-participam-em-nova-mutum-da-prova-brasil](http://www.sonoticias.com.br/noticias/3/129888/mais-de-400-participam-em-nova-mutum-da-prova-brasil)>. Acesso em: 4 maio 2013.

<sup>99</sup> Disponível em: <<http://www.diariodopara.com.br/N-65157.html>>. Publicado em 19/10/2011. Acessado em 03/05/2013.

<sup>100</sup> Disponível em: <<http://www.patosonline.com/interna.php?modulo=publicacao&codigo=22162>>. Publicado em 30/09/2011. Acesso em: 4 maio 2013.

<sup>101</sup> Disponível em: <[www.patos.pb.gov.br/noticias/prova-brasil-em-patos-a2440.html](http://www.patos.pb.gov.br/noticias/prova-brasil-em-patos-a2440.html)>. Acesso em: 4 maio 2013.

<sup>102</sup> Disponível em: <[www.saomi.gueldotapuio.pi.gov.br/noticias.php?id=877](http://www.saomi.gueldotapuio.pi.gov.br/noticias.php?id=877)>. Acesso em: 4 maio 2013.

(conclusão)

Estado/Cidade	Tamanho da rede**	Variação do Ideb**			Ação
		ai/af*	2009	2011	
PI/Teresina	150 escolas e 60 mil alunos	ai	5,2	5,2	Foram realizados vários simulados, o primeiro em julho, com o objetivo de diagnosticar o processo de ensino-aprendizagem. A escola e a família foram orientadas para incentivar os alunos a estudar não só para o teste, mas para superar seus próprios desafios. <sup>103</sup>
		af	4,7	4,4	
RJ/Barra Mansa	53 escolas e 15 mil alunos	ai	4,9	5,1	A prefeitura homenageou os quatro estudantes com melhores resultados do I Simulado da Prova Brasil, realizado pela Secretaria em outubro. Entre 900 alunos, os quatro estudantes de 14 anos do nono ano, da rede pública local, foram os melhores colocados e por isso foram recebidos pelo prefeito em seu gabinete. <sup>104</sup>
		af	4,3	4,1	
RJ/Itatiaia	11 escolas e 3 mil alunos	ai	4,4	4,9	Setembro de 2011: simulados e encontros com as professoras para debater os resultados e, assim, preparar melhor os alunos para a Prova Brasil. <sup>105</sup>
		af	3,4	4,1	
RJ/Itaboraí	67 escolas e 9 mil alunos	ai	4,1	4,6	A preparação vem ocorrendo desde 2010, quando fizemos simulado com os alunos do 4º e do 8º ano e, nesse ano, com os mesmos alunos que estão agora no 5º e no 9º ano. <sup>106</sup>
		af	3,8	3,9	
RS/Santa Maria	52 escolas e 13 mil alunos	ai	4,4	4,9	Agosto de 2011: três simulados, com o objetivo de identificar as deficiências e apontar onde cada escola, cada classe e cada aluno tinha dificuldade para corrigi-las. Vamos entrar na Prova Brasil bem mais preparados. <sup>107</sup>
		af	4,0	4,3	
RS/Victor Graeff	2 escolas e 200 alunos	ai	5,6	5,4	Setembro 2011: realização de simulados e informação para os pais sobre o resultado obtido pelos alunos. <sup>108</sup>
		af	--	5,5	
SE/Itabaiana	50 escolas e 1 mil alunos	ai	3,4	3,4	Julho de 2011: o primeiro de muitos encontros com os diretores para discutir a Prova Brasil. A Secretaria preparou várias sugestões de atividades para que as professoras possam realizar com os alunos que farão a prova. <sup>109</sup>
		af	3,4	3,4	
SE/Riachuelo	8 escolas e 400 alunos	ai	3,2	3,4	Aplicamos o simuladinho, uma forma de fazer com que os estudantes se acostumem com o estilo e o ritmo da prova. Até cartão-resposta colocamos. Esse simulado funciona como reforço escolar. <sup>110</sup>
SP/Jacareí	26 escolas e 9 mil alunos	ai	5,6	5,7	Agosto de 2011: simulado sobre a Prova Brasil, elaborado por equipe de profissionais da Gerência Técnico-Pedagógica. <sup>111</sup>
<b>Total</b>		<b>20 redes municipais, envolvendo 1.330 escolas e 456.600 alunos matriculados</b>			

Fonte: Elaboração do autor.

Legendas: \* ai óanos iniciais do ensino fundamental; af ó anos finais do ensino fundamental.

\*\* Dados extraídos do Data Escola Brasil (<<http://www.dataescolabrasil.inep.gov.br>>) e do Portal do Ideb (<<http://Ideb.inep.gov.br>>).

• O tom cinza indica que a meta do Ideb fixada pelo Inep foi atendida para o ciclo da Prova Brasil do ano indicado.

<sup>103</sup> Disponível em: <[www.teresina-pi.gov.br/noticias/semec/simulado-prova-brasil-aponta-bom-desempenho-de-alunos-de-escolas-municipais-2010.html](http://www.teresina-pi.gov.br/noticias/semec/simulado-prova-brasil-aponta-bom-desempenho-de-alunos-de-escolas-municipais-2010.html)>. Acesso em: 4 maio 2013.

<sup>104</sup> Disponível em: <<http://diariodovale.uol.com.br/noticias/4,50589,Estudantes-sao-premiados-por-bom-desempenho-em-simulado.html#ixzz23tQykpHC>>. Acesso em: 4 maio 2013.

<sup>105</sup> Disponível em: <[http://www.avozdacidade.com/site/page/noticias\\_interna.asp?categoria=3&cod=5649](http://www.avozdacidade.com/site/page/noticias_interna.asp?categoria=3&cod=5649)>. Acesso em: 4 maio 2013.

<sup>106</sup> Publicado em <[www.osaogoncalo.com.br/site/geral/2012/2/9/30240/alunos+serão+preparados+para+prova+brasil+do+mec](http://www.osaogoncalo.com.br/site/geral/2012/2/9/30240/alunos+serão+preparados+para+prova+brasil+do+mec)>. Publicado em 20/08/2011. Acesso em: 4 maio 2013.

<sup>107</sup> Disponível em: <[www.santamaria.rs.gov.br/index.php/\\_secretarias/pdf/www.ucsrn.com/\\_eventos/docs/simbolos/turismo?secao=noticias&id=2500](http://www.santamaria.rs.gov.br/index.php/_secretarias/pdf/www.ucsrn.com/_eventos/docs/simbolos/turismo?secao=noticias&id=2500)>. Acesso em: 4 maio 2013.

<sup>108</sup> Disponível em: <[www.osaogoncalo.com.br/site/geral/2012/2/9/30240/alunos+serão+preparados+para+prova+brasil+do+mec](http://www.osaogoncalo.com.br/site/geral/2012/2/9/30240/alunos+serão+preparados+para+prova+brasil+do+mec)>. Acesso em: 4 maio 2013.

<sup>109</sup> Disponível em: <<http://www.itabaiana.se.gov.br/ler.asp?id=623&titulo=Noticias>>. Acesso em: 4 maio 2013.

<sup>110</sup> Disponível em: <[www.infonet.com.br/educacao/ler.asp?id=120432](http://www.infonet.com.br/educacao/ler.asp?id=120432)>. Acesso em: 4 maio 2013.

<sup>111</sup> Disponível em: <<http://www.nossajacarei.com.br/noticias/1-cotidiano/1005-jacarei-realiza-simulado-da-prova-brasil.html>>. Acesso em: 4 maio 2013.

No Quadro 2.25 foram tabulados trechos de matérias obtidas de 20 municípios que possuem 1.330 escolas e mais de 456 mil alunos. Da mesma forma que aconteceu com as Secretarias estaduais, nas municipais, tanto para grandes redes, como a de Manaus, com 172 mil alunos, como a da pequena cidade de Victor Graeff, com 200 alunos, o retorno às aulas no segundo semestre de 2011 marcou a realização de simulados para preparar os alunos para a Prova Brasil, sendo que em alguns deles foram aplicadas até três simulados. Em um deles, Itaboraí, a preparação começou em 2010, acompanhada de simulados com turmas de 4º e 8º anos do ensino fundamental, que em 2011 fariam o teste nacional. Algumas das matérias relatam programas de formação para as professoras, sempre direcionados àquelas que lecionam Língua Portuguesa e Matemática nos 5º e 9º anos do ensino fundamental, também concentrados no segundo semestre de 2011 e direcionados a melhorar o desempenho de seus alunos nos temas das Matrizes de Referência da Prova Brasil. No geral, os municípios conseguiram cumprir as metas impostas pelo Inep.

Resultado semelhante aparece no Quadro 2.26, que destaca matérias referentes a oito escolas localizadas em diferentes Estados, as quais desenvolvem, por iniciativa própria, simulados ou então aulas de reforço em Matemática e Língua Portuguesa, nos meses que antecederam a aplicação da Prova Brasil em 2011.

**Quadro 2.26 – Tabulação das matérias coletadas no período entre 10 de janeiro de 2009 e 11 de outubro de 2012, utilizando a ferramenta “Alertas do Google”, que trata das escolas e de suas ações para a melhoria do Ideb, mostrando município, tamanho da escola e efeito no Ideb, nos anos iniciais e finais do ensino fundamental**

Estado/Cidade	Escola	Matrículas (*)	Variação do Ideb (*)			Ação
			ai/af (**)	2009	2011	
BA/Itapetinga	Isaura C. Costa	345 alunos	ai	3,4	4,7	Outubro de 2011: anuncia a realização do 2º Simulado Escola Isaura Costa Prova Brasil (Seic), com o objetivo de preparar os alunos para o teste. <sup>112</sup>
BA/Santaluz	Sisalândia	133 alunos	af	4,1	4,0	Realização do simulado em junho, no qual os alunos foram mal em Matemática, inclusive õchutandoõ algumas questões, justificando o fato por terem realizado primeiro a prova de Língua Portuguesa. õConversamos sobre esse problema e esperamos que no segundo simulado o quadro se normalize e eles tenham um bom desempenho na Prova Brasilõ. <sup>113</sup>
PA/Ananindeua	Cândida S. de Souza	916 alunos	af	4,2	4,4	Informa que está realizando aos sábados aulas de Português e Matemática, para o aprimoramento escolar dos alunos de 8ª série na Prova Brasil 2011. <sup>114</sup>
PR/Toledo	Luiz Augusto M. Rego	544 alunos	af	<b>4,0</b>	<b>4,0</b>	Outubro de 2011: de olho em melhorar o desempenho na próxima Prova Brasil, o Colégio realizou simulado. <sup>115</sup>
	Washington Luiz	100 alunos	ai	7,2	7,9	Prova Brasil muda o dia a dia nas escolas. Durante o ano letivo, a professora trabalhou com os alunos simulados e atividades. Em 2009, isso não foi feito, mas agora a cobrança é maior. õComo alcançamos o melhor índice no município, muitos acham que tivemos sorte, por isso nos esforçamos para mantê-lo e provar que o ensino é de qualidadeõ. <sup>116</sup>
RN/Felipe Guerra	José P. Barra	359 alunos	af	3,5	3,4	Em setembro, a escola informava estar realizando seu terceiro simulado do ano preparatório para a Prova Brasil. <sup>117</sup>
RR/Boa Vista	Mario D. Andreazza	400 alunos	ai	5,1	5,3	Recebe recursos do Mais Educação para, entre outras coisas, promover o acompanhamento pedagógico nas áreas de leitura, escrita e matemática e, segundo a gestora, o foco é "a melhoria do Ideb e também da Prova Brasil. Esse é o diferencial das nossas atividadesõ. <sup>118</sup>
			af	3,9	4,1	
SC/Capivari de Baixo	Escola Pequeno Polegar	137 alunos	ai	4,9	sem média (***)	Os estudantes preparam-se bastante. A cada 15 dias, os alunos fazem um simulado. As provas são fornecidas pela Secretaria e as professoras fazem uma adaptação. <sup>119</sup>

Fonte: Elaboração do autor.

Legendas: \* Extraído de Data Escola Brasil (<<http://www.dataescolabrasil.inep.gov.br>>) e Portal do Ideb (<<http://Ideb.inep.gov.br>>).

\*\* ai ó anos iniciais do ensino fundamental; af ó anos finais do ensino fundamental.

\*\*\* - menos de 50% dos alunos matriculados participaram do teste e portanto a média não foi divulgada pelo Inep

<sup>112</sup> Disponível em: <<http://escolaisauracosta.blogspot.com.br/2011/10/2-simulado-escola-isaura-costa-seic.html>>. Acesso em: 4 maio 2013.

<sup>113</sup> Disponível em: <<http://escolade1grausalandia.blogspot.com.br/2011/07/simulado-da-prova-brasil.html>>. Acesso em: 4 maio 2013.

<sup>114</sup> Disponível em: <<http://escolacandida21.blogspot.com.br/2011/08/escola-candida-na-prova-brasil.html>>. Acesso em: 4 maio 2013.

<sup>115</sup> Disponível em: <<http://jornaldoeste.orangologa.com.br/cidade/moraes-rego-realiza-simulado-para-a-prova-brasil-10504>>. Acesso em: 4 maio 2013.

<sup>116</sup> Disponível em: <<http://www.jornaldoeste.com.br/cidade/prova-brasil-muda-o-dia-a-dia-nas-escolas-11896>>. Acesso em: 4 maio 2013.

<sup>117</sup> Disponível em: <<http://brejocity.blogspot.com.br/2011/09/realizacao-do-3-simulado-prova-brasil.html>>. Acesso em: 4 maio 2013.

<sup>118</sup> Disponível em: <[www.bvnews.com.br/esporte1019.html](http://www.bvnews.com.br/esporte1019.html)>. Acesso em: 4 maio 2013.

<sup>119</sup> Disponível em: <[http://www.notisul.com.br/n/geral/alunos\\_estudam\\_para\\_prova\\_brasil-31301](http://www.notisul.com.br/n/geral/alunos_estudam_para_prova_brasil-31301)>. Acesso em: 4 maio 2013.

- O tom cinza indica que a meta do Ideb fixada pelo Inep foi atendida para o ciclo da Prova Brasil do ano indicado.

Assim como no caso das Secretarias de Educação, algumas escolas também desenvolvem, por iniciativa própria, simulados para preparar seus alunos para o teste nacional. No caso das escolas, as matérias são mais escassas e foram colhidas na maior parte das vezes de *blogs* mantidos por elas ou por alguma professora.

Dois casos merecem destaque, quando se analisa o Quadro 2.26. O primeiro é o da Escola Washington Luiz, em Toledo/PR. Essa escola participou da Prova Brasil pela primeira vez em 2009 e já alcançou um patamar elevado que se manteve em 2011. Pode ser que nessa escola o simulado seja utilizado da forma como foi discutido anteriormente: dentro de um contexto determinado, pela professora que conhece sua turma e que utiliza os resultados para rever as práticas de sala de aula e para auxiliar os alunos com mais dificuldades. Mas também pode ser que esteja havendo um estreitamento do currículo, em que as atividades se concentrem nas Matrizes de Referência. De qualquer forma, qualquer julgamento, sem conhecer em detalhes a realidade dessa escola, seria precipitado. Por essa e outras razões, não é possível fazer inferências sobre a qualidade dessa ou de qualquer outra escola apenas tomando como referência o valor do Ideb, ainda mais sem definir claramente a qual qualidade se está referindo.

O segundo caso é o da Escola Pequeno Polegar, em Capivari de Baixo/SC, onde a matéria afirma que se realizaram simulados a cada 15 dias. Nesse caso, em especial, existe grande possibilidade de o ensino estar todo centrado na preparação para o teste. No entanto, um dado traz informações sobre uma possível tentativa de fraude, pois, em 2011, menos de 50% dos alunos matriculados fizeram o teste; por isso, no Quadro 2.26, aparece a informação *sem média*. Algo comum nesses casos é deixar de fora os alunos com baixo desempenho, para não prejudicar a média da escola<sup>120</sup>.

Tomando por base os Quadros 2.23 e 2.24, essa pesquisa trouxe informações de que os simulados estão sendo utilizados como uma das estratégias para conseguir cumprir as metas do Ideb, principalmente nos três meses anteriores à realização da Prova Brasil. São dados agregados de redes estaduais e municipais, que congregam 4.300 escolas de ensino fundamental onde estão matriculados quase 1,5 milhões de alunos, representando 6% do total de alunos matriculados no ensino fundamental regular. Menos do que atentar para a pequena porcentagem, o mais importante é notar que essa forma de obter resultados, na maioria da

---

<sup>120</sup> Do total de 71.224 escolas que participaram do teste em 2011, em 1.113 delas, cerca de 1,5% do total, a taxa de participação foi inferior ao mínimo estabelecido, e as escolas não tiveram seu Ideb divulgado (INEP, 2012a).

vezes, pode estar mascarando os resultados que estão sendo coletados, reforçando o alerta para o uso inadequado que se está fazendo dos resultados do Ideb.

Mesmo que se considere que seja necessário relativizar as informações obtidas a partir de matérias publicadas na internet, já que poderiam não apresentar dados fidedignos, a quantidade de alunos envolvidos no processo e a quantidade de depoimentos em favor da preparação para o teste não podem esconder que essa realidade é mais presente do que se poderia imaginar. Esse fato, além de representar uma corrida atrás dos números, pode ser também uma tentativa de burlar o processo que se está querendo medir.

O último tema classificado e apontado no Quadro 2.22 destaca a relação entre os resultados no Ideb e a pobreza, um tema da maior importância, que recebeu apenas 36 referências. A maior parte das matérias ligadas a esse tema ressaltava os baixos desempenhos das escolas situadas em regiões de baixo nível socioeconômico, sendo que algumas delas descreviam algumas pesquisas, acadêmicas ou não. Uma delas apresentava o resultado de um estudo que concluía que os estudantes de baixa renda recebiam menos conteúdo escolar, quando comparados aos estudantes de renda elevada, com base no cruzamento dos desempenhos da Prova Brasil com as respostas aos questionários socioeconômicos, em que as professoras informavam se conseguiam trabalhar em sala tudo o que havia sido planejado. Outra comentava que o programa Bolsa-Família havia contribuído para a diminuição do abandono escolar. Outras poucas matérias, mas que receberam bastante destaque, enfatizavam os resultados positivos de escolas situadas em regiões pobres, mas que conseguiram superar essa situação com o apoio e a dedicação das professoras, novamente dentro daquela visão messiânica da educação que supera todas as barreiras.

Esse tema envolvendo a relação entre a pobreza e o desempenho educacional vem sendo muito estudado nos EUA e está no centro das discussões sobre o *gap* existente entre os desempenhos maiores e menores. Lá, essa discussão ganhou impulso para realçar o empenho que o país deveria fazer para igualar a diferença existente entre negros e brancos, durante os anos 1960, e evoluiu para a diferença entre os chamados *Aprendizes de Língua Inglesa*, aqueles cujas famílias emigraram para o país e aqueles cujos pais não são nascidos nos EUA. O que se discute agora é que essas duas situações são faces da mesma moeda, pois se referem a grupos que estão entre os mais pobres daquele país, tendo, portanto, que receber suporte adicional para superar as dificuldades sociais (RAVITCH, 2010).

No caso brasileiro, a pesquisadora Silvia Yannoulas, líder do grupo de pesquisa Trabalho, Educação e Discriminação (TEDis), da Universidade de Brasília (UnB), afirmou em entrevista, em que comenta os trabalhos desenvolvidos pelo grupo, fazer parte do senso

comum, estando especialmente arraigada entre as professoras da escola pública, a crença de que os alunos mais pobres sempre terão os desempenhos mais baixos (UFMG, 2013). Dessa forma, os alunos pobres estariam fadados ao fracasso e nada poderia ser feito para mudar essa situação predeterminada. Isso poderia explicar, pelo menos em parte, o fato de a relação entre a pobreza e os baixos desempenhos ter recebido pouco destaque nas matérias publicadas. Se faz parte do senso comum que o fracasso está predeterminado, o tema não é novidade e não merece a atenção da mídia eletrônica escrita.

Pelo menos duas pesquisas de campo mostram o contrário, revelando que, apesar das condições socioeconômicas desfavoráveis, algumas escolas conseguem obter desempenhos superiores, quando medidos pela Prova Brasil ou pelo Ideb, buscando demonstrar que a pobreza não implica necessariamente baixos desempenhos. Destacam que a solução para esse problema começa com uma atenção diferenciada que essa população em situação de vulnerabilidade social deve ter, não porque possuem dificuldades para aprender, mas porque enfrentam situações sociais desfavoráveis para a aprendizagem. A primeira pesquisa identificou escolas com altos desempenhos na Prova Brasil, em que pelo menos 30 alunos tenham participado do teste, frequentadas por alunos de baixo nível socioeconômico de cidades também pobres (MEC, INEP, UNICEF, 2007). A outra, mais recente e com critérios de seleção de casos mais rigorosos, cruzou dados do Ideb de escolas em que pelo menos 70% dos alunos matriculados participaram da Prova Brasil com o nível socioeconômico desses alunos (FUNDAÇÃO LEMANN; ITAÚ BBA, 2012). Nos dois casos, o envolvimento da escola com a comunidade que a cerca, o esforço de incorporar as famílias no cotidiano da vida escolar de seus filhos e o sentimento da equipe escolar que todos alunos podem aprender são fatores, entre outros, que ajudaram a superar as dificuldades de aprendizagem dos alunos. Apesar de não terem sido destacados nesses estudos, provavelmente a baixa rotatividade das professoras, sua experiência profissional e a qualidade de sua formação inicial devem ter sido fatores que muito contribuíram para os resultados positivos apontados.

Outras matérias não foram classificadas, pois tratavam de assuntos diversos. Apesar disso, um tipo delas que mereceu pouco destaque, mas que tem muita importância na discussão sobre os usos dos indicadores educacionais, tem a ver com os sistemas estruturados de ensino desenvolvidos por grandes grupos educacionais e comercializados nas pequenas prefeituras de todas as regiões.

Uma delas é uma matéria publicada no jornal *Gazeta do Paraná*, intitulada "Escolas públicas contratam métodos privados de ensino",<sup>121</sup> que traz algumas declarações de dirigentes e professoras sobre a eficácia dessas iniciativas. Uma secretária de um município paranaense afirma que a compra de um desses sistemas foi importante porque auxilia o trabalho da professora e vem acompanhado de capacitação pedagógica. Uma diretora de escola, também do Paraná, afirma que, depois que seu município adquiriu o produto, o Ideb da escola passou de 1,5 em 2005 para 5,5 em 2009.

A possibilidade de que esses sistemas estruturados possam contribuir para elevar o Ideb dos municípios vem fazendo com que um crescente número de empresas os estejam comercializando e um também crescente número de municípios os estejam adquirindo. Pesquisa coordenada pela pesquisadora Paula Louzano a partir de questionários enviados aos secretários municipais de educação do Estado de São Paulo confirma essa tendência (FUNDAÇÃO LEMANN; FGV, 2008). Segundo a pesquisa, em 1998, nenhum dos 645 municípios paulistas utilizavam esses sistemas estruturados e dez anos depois 29% deles o faziam, o equivalente a 187 municípios. No total, foram identificadas 17 empresas que ofereciam esse tipo de sistema, sendo que a liderança é a do Curso Oswaldo Cruz (COC), utilizado por 80 municípios paulistas. Em continuidade a essa pesquisa, constatou-se, como era de se esperar, que não se poderia atribuir somente a utilização desses métodos às melhoras de desempenho verificadas (FGV; FUNDAÇÃO LEMANN, 2009).

Como pode ser verificado pelos dados apresentados, as iniciativas do governo federal em relação aos indicadores de desempenho da educação básica receberam bastante destaque. Uma pesquisa de opinião comandada pelo Ibope, que tinha por objetivo identificar, a partir de uma amostra nacional, quais seriam os problemas que deveriam ser enfrentados pelos candidatos à presidência da República, na eleição de 2011 (IBOPE; FUNDAÇÃO SM; TODOS PELA EDUCAÇÃO, 2010) traz algumas informações interessantes a respeito do que pensa o brasileiro sobre a qualidade da educação e do que deve ser feito para melhorá-la. Essa pesquisa repetiu as mesmas perguntas de outra sobre o mesmo tema, realizada em 2006, também antecedendo a eleição presidencial. Entre as diversas perguntas, destaca-se a de número 13: "Desta lista de órgãos e grupos relacionados direta ou indiretamente com a escola pública, gostaria que o(a) Sr.(a.) me dissesse qual deles, em sua opinião, pode contribuir mais concretamente para assegurar a qualidade da educação básica no Brasil? E qual o segundo? E

---

<sup>121</sup> Disponível em: <<http://www.gazetadopovo.com.br/ensino/conteudo.phtml?id=1163731&tit=Escolas-publicas-contratam-metodos-privados-de-ensino>>. Acesso em: 4 maio 2013.



qual o terceiro?ö (IBOPE, FUNDAÇÃO SM; TODOS PELA EDUCAÇÃO, 2010, p. 39). O resultado das respostas é apresentado no Quadro 2.27.

**Quadro 2.27 – Resposta à pergunta estimulada, com três opções de resposta, sobre a percepção da população, referente a quem pode contribuir mais para garantir a qualidade da educação básica – 2006-2010**

Órgão e grupo	2006 (%)	2010 (%)
Governo federal	27	55
Prefeituras	23	42
Ministério da Educação	29	39
Governos estaduais	16	33
Secretarias Municipais de Educação	18	25
Secretarias Estaduais de Educação	12	18
Pais de alunos	13	18
Diretores de escolas	10	12
Professoras das escolas públicas	27	10
Alunos das escolas públicas	8	9
Deputados e senadores	2	8
Associação de Pais e Mestres (APM)	12	8
Empresas privadas	6	5
Organizações Sociais (ONGs, fundações, institutos)	17	5
Igrejas	-	4
Nenhum destes	5	0
NS/NR	10	1

Fonte: IBOPE; FUNDAÇÃO SM; TODOS PELA EDUCAÇÃO, 2010, p. 39.

Observa-se no Quadro 2.27 que, isoladamente, o grupo que mais cresceu entre as duas pesquisas realizadas foi o governo federal, com 28 pontos percentuais, fruto talvez do grande destaque que o Ideb, provavelmente identificado como uma ação federal, alcançou no noticiário nacional e local. Por outro lado, apesar de toda a ênfase que as políticas têm dado ao desempenho das professoras, principalmente aquelas relacionadas à premiação por desempenho, a população diminuiu o peso relativo dessas profissionais, quando se trata de apontar quem mais pode contribuir para a qualidade educacional.

Para melhor discutir as variações ocorridas, entre os dois períodos, montou-se o Quadro 2.28, em que foram reunidos os grupos por nível de governo e o grupo da escola.

**Quadro 2.28 – Agregação das respostas relacionadas aos governos e às escolas, fornecidas pela pesquisa do Ibope**

Nível	2006 (%)	2010 (%)
Federal	56	94
Municipal	41	67
Estadual	28	51
Escola	37	22

Fonte: IBOPE; FUNDAÇÃO SM; TODOS PELA EDUCAÇÃO, 2010, com elaboração do autor.

O Quadro 2.28 traz dados instigantes. Primeiramente, destaca-se a importância que a população atribuiu ao nível federal para a melhoria da qualidade da educação, citado por 94% dos entrevistados. Se, por um lado, foi a partir de uma iniciativa federal que os resultados dos testes foram divulgados por escola, fazendo com que ela seja ranqueada e seu desempenho exposto publicamente, recai sobre o nível federal a maior expectativa quanto à contribuição para garantir a qualidade do ensino. Outro dado interessante é que, desde 2006, a população atribui muito mais aos governos do que à escola a parcela de contribuição para a melhoria da qualidade educacional. E essa percepção se agudiza ainda mais em 2010, quando a população atribui aos governos o maior peso para garantir a qualidade educacional, um aumento médio de 29 pontos percentuais, diminuindo o peso da escola, um decréscimo de 15 pontos percentuais. No caso das escolas, no período da pesquisa, diminuiu em 17 pontos percentuais a percepção de que seriam as professoras aquelas que poderiam contribuir mais para essa melhoria, uma diminuição de 27% para 10%. Assim, parece que, pelo menos aos olhos da população, a pressão que se queria exercer sobre a escola e suas professoras para melhorar o desempenho de seus alunos parece ter se voltado para os governos e mais fortemente para o governo federal.

Finalizando esse capítulo, é possível perceber a trajetória do processo avaliativo instituído pelo governo federal que passa de uma preocupação com a aprendizagem dos alunos e com a formação das professoras para uma preocupação com os resultados, desconectando-os do processo educacional. Com isso, induz-se a preparação para o teste e o estreitamento do currículo, pois o importante é atingir metas, mesmo que elas sejam definidas unilateralmente e estejam distantes da realidade vivenciada pela escola.

No capítulo seguinte, serão discutidos os processos avaliativos desenvolvidos pelos Estados de Minas Gerais e São Paulo.

## CAPÍTULO 3

### OS TESTES AVALIATIVOS NOS ESTADOS DE MINAS GERAIS E SÃO PAULO

Neste capítulo são discutidos os caminhos traçados pelos Estados de Minas Gerais e São Paulo na direção da construção de sistemas avaliativos de suas redes estaduais de educação e a utilização de seus resultados para a formulação de políticas educacionais. As redes estaduais de Minas Gerais e São Paulo correspondem, juntas, a 40% do total de professoras em efetivo exercício e a 40% das matrículas do total das redes estaduais de ensino fundamental brasileiras. Portanto, analisar o movimento dos testes avaliativos nesses Estados é significativo no contexto brasileiro, dado o impacto para o conjunto do País.

Para melhor caracterizar cada um dos Estados, construiu-se o Quadro 3.1, que mostra a população, a quantidade de municípios e, no âmbito das redes estaduais de ensino fundamental, o número de escolas, de alunos matriculados e de professoras em exercício, comparando-os com os dados do Brasil.

**Quadro 3.1 – Caracterização das redes de ensino fundamental regular estaduais, a partir da população total, do número de municípios, escolas, alunos matriculados e professoras em exercício – MG, SP e Brasil**

Estado	População <sup>(1)</sup>	Municípios <sup>(1)</sup>	Escolas <sup>(2)</sup>	Matrículas <sup>(2)</sup>	Professoras <sup>(3)</sup>
MG	19.597.330	853	3.305	1.270.198	75.730
SP	41.262.199	645	4.888	2.406.085	120.546
Brasil	190.755.799	5.565	25.768	9.083.704	495.390

Fonte: (1) IBGE. *Censo Demográfico 2010*.

(2) INEP. *Censo da Educação Básica 2012* ó Matrícula inicial.

(3) INEP. *Censo da Educação Básica 2011* ó Professoras em efetivo exercício.

Para a pesquisa, foram utilizados documentos produzidos pelos órgãos de governos estaduais, em especial pelas Secretarias de Educação, relativos às concepções dos sistemas avaliativos implantados, aos resultados obtidos e à documentação referente aos principais programas desenvolvidos, buscando compreender o percurso desenvolvido pelas políticas traçadas. As principais fontes de informação foram os *sites* das Secretarias de Educação e das Assembleias Legislativas dos dois Estados.

A coleta dessas informações foi uma tarefa árdua, pois os dados estão esparsos e, muitas vezes, os mecanismos de busca existentes nos *sites* não facilitam o trabalho daqueles que buscam informações. Foi necessário montar diversos quebra-cabeças com o objetivo de

ilustrar o movimento nos dois Estados. Em algumas situações, foi necessário recorrer a pessoas que vivem os processos de perto, para esclarecer alguns pontos fundamentais, de modo a fechar o círculo das informações.

### ***O Estado de Minas Gerais***

Os testes avaliativos em Minas Gerais são bastante abordados na literatura educacional, destacando-se importantes estudos: SILVA, 2011; SOUZA, 2007; SOUZA; OLIVEIRA, 2007 e BROOKE; CUNHA; FALEIROS, 2011.

São abordados neste capítulo os esforços dos diferentes governos mineiros para implantar a avaliação educacional no Estado, englobando o período entre 1988 e 2012. Os primeiros estudos começam em 1988, na mesma época em que também ocorrem os primeiros estudos do governo federal, e deram origem, quatro anos depois, ao Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais. Esse programa foi deixado de lado e em seu lugar criou-se, em 2000, o Sistema Mineiro de Avaliação da Educação Pública (Simave). Também serão discutidos algumas ações e programas desenvolvidos pela Secretaria de Educação, que levam em consideração os resultados da avaliação educacional.

### **As primeiras pesquisas para avaliar o ensino, a consolidação e o abandono do Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais – 1988 a 1996**

A implantação de mecanismos de avaliação educacional em Minas Gerais está prevista na Constituição do Estado de 1989. O seu inciso X do artigo 196 determina que deve existir um padrão de qualidade do ensino garantido a partir de avaliação cooperativa periódica promovida pela Secretaria de Educação, envolvendo as professoras e os responsáveis pelos alunos. O fato de ter sido prevista uma avaliação cooperativa foi registrado como o único caso em que uma Constituição estadual destacou o real significado do termo *avaliação* e sua contribuição para o aprimoramento do ensino (VIANNA; ANTUNES; SOUZA, 1993). No entanto, essa determinação foi sendo deixada de lado com o passar dos anos e atualmente não é mais cumprida.

Em 1988, durante o governo Newton Cardoso (1987/1991), foi realizada a primeira avaliação educacional, envolvendo o Ciclo Básico de Alfabetização (CBA), uma estratégia implantada em Minas Gerais, em 1985, que recebera muitas críticas, principalmente das professoras que discordavam do fato de não haver reprovação na passagem da 1ª para a 2ª série. A Secretaria de Educação tinha a intenção de, com essa avaliação, obter dados para rebater a crítica frequente nos setores educacionais de que as crianças sabem cada vez menos (GOULART, 1992, p. 5). Como essa avaliação foi restrita, decidiu-se pela realização de uma nova, mais abrangente, que ampliasse seu escopo, desenvolvendo-se a pesquisa *Avaliação do Ciclo Básico de Alfabetização (AVA-CBA)*. Essa pesquisa, coordenada por Heraldo Vianna, pesquisador da Fundação Carlos Chagas (FCC), envolvido com a elaboração do Saeb, previa, como um de seus instrumentos, a aplicação de testes de Língua Portuguesa, Matemática e Ciências aos alunos. O planejamento para a aplicação dos testes foi executado durante o último ano do governo Newton Cardoso, mas os testes só foram aplicados em março de 1991, durante o governo de Hélio Garcia (1991/1995).

Ainda durante 1991, ocorreram os primeiros estudos para ampliar as pesquisas, no sentido de implantar uma avaliação educacional. Esses estudos, coordenados também por Heraldo Vianna, deram origem ao Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais.

A avaliação educacional já fazia parte dos planos do secretário de educação daquele governo, Walfrido dos Mares Guia, como demonstra um texto de sua autoria, em que delineia as principais ações desenvolvidas pela pasta:

[...] promoção da autonomia da escola, em todos os seus aspectos: pedagógico, administrativo e financeiro; fortalecimento da direção da escola, através da liderança do diretor e da participação de um colegiado que representa a comunidade; capacitação e aperfeiçoamento dos profissionais da educação; avaliação da escola; integração com os municípios (GUIA NETO, 1994, p. 12).

Percebe-se que a avaliação educacional, juntamente com outras ações ligadas à autonomia da escola, assumiu destaque dentro das ações da Secretaria.

Segundo o secretário, a avaliação seria peça importante dentro do programa de Qualidade Total, implementado com o objetivo de melhorar a gestão do sistema. Assim, a partir dos resultados da avaliação, pretendia-se atuar sobre a gestão de todo o sistema de ensino, não somente sobre a escola. No mesmo texto, o secretário destaca terem sido criadas

as condições para que, havendo interesse, os municípios pudessem participar do Programa de Avaliação. Não foi possível obter informação sobre se, de fato, isso aconteceu e quantos municípios participaram.

O Programa de Avaliação de Minas Gerais nasceu mais abrangente do que o Saeb, Tanto em relação ao número de alunos, pois o teste era censitário para a população-alvo, quanto às áreas de conhecimento testadas, incluindo, além de Língua Portuguesa e Matemática, as áreas de Ciências, História, Geografia, Física, Química, Biologia e a inclusão de redação. Para garantir a execução do Programa, a Secretaria de educação contava com recursos da ordem de US\$ 302,0 milhões do Projeto Qualidade na Educação Básica de Minas Gerais, conhecido como Pró-Qualidade, recursos esses divididos em partes iguais entre o Banco Mundial e o governo do Estado (SANFELICE; PEDROSA, 2005). Como um dos objetivos do Programa era melhorar o fluxo escolar do ensino fundamental, prejudicado principalmente pela repetência, o monitoramento do sistema, por meio da avaliação educacional, surge como importante elemento na direção desse objetivo.

O Programa de Avaliação foi normatizado pela Resolução SEE nº 6.908, editada em 17 de janeiro de 1992, o mesmo ano em que ocorreu seu primeiro ciclo e dois anos depois da primeira aplicação do Saeb. Vianna, Antunes e Souza (1993), ao descreverem as características do programa, destacam que a avaliação da qualidade da educação deve levar em consideração variáveis relacionadas

[...] ao *contexto* (atuação na escola), a *fatores não relacionados diretamente à escola* (*status* socioeconômico da família, nível de educação dos pais, recursos educacionais do lar, participação dos pais no processo educacional, atividades educacionais fora da escola, atividades de lazer e sociais e atitudes dos estudantes) e a *informações ligadas à escola*, por intermédio da coleta de elementos de *entrada* (características do corpo docente, programas escolares e participação dos pais), do *processo* (currículos e práticas instrucionais) e, finalmente, de dados relativos ao *produto*, representado pelo desempenho escolar e a formação de atitudes (VIANNA, ANTUNES; SOUZA, 1993, p. 6).

Percebe-se que existia uma preocupação em não caracterizar a qualidade educacional apenas por meio do desempenho dos alunos no teste, mas, sim, enxergando-a como algo muito mais amplo, alcançando, inclusive, fatores externos à escola e relacionados às atitudes dos alunos frente a alguns temas. Note-se que, enquanto a proposta do secretário limitava-se apenas aos resultados do teste, o Programa era mais abrangente. Em um ponto, tanto o

secretário como os formuladores do Programa concordavam: consideravam o Programa de Avaliação um instrumento para a gestão do sistema, não se percebendo nos textos consultados nenhuma intenção em classificar a escola, nem responsabilizá-la individualmente.

Foram quatro os objetivos traçados pelo Programa de Avaliação:

I ó conhecer o desempenho dos alunos nas escolas estaduais;

II ó levantar dados e informações para servirem de subsídios à tomada de decisões sobre o desenvolvimento do processo de ensino;

III ó identificar, nos conteúdos curriculares, pontos críticos que necessitam maior atenção e requerem intervenção imediata e prioritária para a melhora da aprendizagem;

IV ó fornecer, a partir dos dados levantados, subsídios para que os professores possam atuar com maior eficiência na condução do processo de ensino e a Secretaria da Educação possa colaborar para maior eficiência do sistema (VIANNA; ANTUNES; SOUZA, 1993, p. 9).

Pela descrição dos objetivos, fica claro o caráter do Programa de Avaliação como auxiliar do processo de aprendizagem, ao fornecer informações que pudessem aprimorá-lo, e contribuir para que a Secretaria se organizasse no sentido de garantir a eficiência do sistema.

Os autores destacam, ainda, que se buscava com o Programa criar uma cultura de avaliação em todo o sistema de ensino, pois, segundo eles, a prática da avaliação era muito rara na escola, além de influenciar, a partir dos modelos de testes aplicados, na melhoria da qualidade das provas aplicadas pelas professoras. Portanto, nesse caso, buscava-se, mediante um modelo externo, induzir modificações no interior da escola. Outro objetivo previsto era o de contribuir para que as professoras desenvolvessem uma atividade cooperativa dentro da escola, a partir de atividades como a correção dos testes e elaboração de relatórios, que deveriam contar, inclusive, com o envolvimento dos membros do colegiado. Além desses objetivos, havia mais um, ligado à autonomia da escola, pois se esperava que as escolas atingissem a autonomia pedagógica a partir dos resultados do Programa.

É possível perceber que os objetivos traçados eram bastante ambiciosos, não se limitando a testar os alunos e a divulgar os resultados. Esperava-se um grande movimento que envolvesse toda a comunidade escolar, dentro de um espírito cooperativo, no sentido de melhorar as práticas pedagógicas e o aprendizado dos alunos.

O alvo do Programa eram as 3<sup>a</sup>, 5<sup>a</sup> e 8<sup>a</sup> séries do ensino fundamental e a 2<sup>a</sup> série do ensino médio e, além deles, as 3<sup>a</sup> e 4<sup>a</sup> séries do curso de Habilitação para o Magistério. A justificativa para essa escolha era que:

A avaliação de alunos do ensino fundamental, no início da 3<sup>a</sup> série, tem permitido verificar em que medida o Ciclo Básico de Alfabetização (CBA) estaria conseguindo atingir com êxito seus objetivos. A 5<sup>a</sup> série representa uma mudança radical na vida do aluno, que deixa a primeira fase do ensino fundamental, ingressando em uma sala de aula onde o ensino é diversificado em seu conteúdo, nas metodologias de ensino e na pluralidade dos professores. A 8<sup>a</sup> série, por ser terminal de um nível de ensino que não tem continuidade para uma parcela expressiva do alunado que não prossegue seus estudos em nível de ensino médio. A 2<sup>a</sup> série do ensino médio, que ainda não se acha inteiramente contaminada pela influência dos cursos preparatórios para o acesso ao ensino de 3<sup>o</sup> grau, foi selecionada para integrar o projeto, fornecendo subsídios sobre o sistema em algumas importantes áreas de conteúdo nesse nível: Português, Matemática, Física, Química e Biologia, além de informações sobre o funcionamento da escola. A inclusão da 3<sup>a</sup>/4<sup>a</sup> série da Habilitação do Magistério resultou de razões óbvias: conhecer aspectos da formação dos futuros professores que irão atuar nas quatro primeiras séries do ensino fundamental (VIANNA; ANTUNES; SOUZA, 1993, p. 18).

Reforça-se, nesse trecho, uma grande preocupação com o processo educacional, pois a escolha das séries estava ligada aos períodos escolares considerados mais críticos para os alunos. Percebe-se, também, uma preocupação com os aspectos que pudessem auxiliar os alunos em seu desenvolvimento e não com o fato de eles estarem ou não aprendendo. Três dados chamam a atenção: testar os alunos no início da 3<sup>a</sup> série, para obter informações sobre o Ciclo Básico de Alfabetização (CBA); obter informações sobre o ensino médio, a partir dos alunos da 2<sup>a</sup> série, evitando com isso contaminar os dados com possíveis preparações para o vestibular; testar os alunos dos cursos de Habilitação para o Magistério, a única iniciativa que se tem notícia, dentro da avaliação educacional brasileira, para obter informações sobre a formação de professoras.

O Programa de Avaliação previa, além da aplicação de testes por áreas do conhecimento, a aplicação de questionários, tanto para os alunos como para a administração da escola. Para cada série, o teste seria composto por diferentes números de itens.

No início da 3<sup>a</sup> série do ensino fundamental, em um único dia de aplicação, os alunos responderiam a 30 itens de diferentes áreas do conhecimento, de tal forma que 1/3 deles seria



submetido ao teste de Língua Portuguesa e redação, outro 1/3 ao de Matemática e o 1/3 restante ao de Ciências. Para operacionalizar esse arranjo, durante a aplicação dos testes, os alunos da escola eram enturmados por área do conhecimento. É importante destacar que, apesar de estar previsto testar a área de Ciências Sociais, isso não foi feito porque

[...] por sua própria natureza e pela grande diversidade cultural, social, histórica e geográfica do Estado, apresentam [as Ciências Sociais] um alto grau de heterogeneidade em sua abordagem nas turmas nos primeiros anos de escolaridade de 1º grau (ANTUNES; XAVIER; FREITAS, 1992, p. 34).

Portanto, percebe-se que o objetivo do Programa de Avaliação esteve desde seu início ligado à integralidade do processo de aprendizagem do aluno e a seu processo de formação como cidadão. Quando uma área do conhecimento deixou de ser testada, a razão foi de natureza técnica, ligada às dificuldades na definição de parâmetros válidos para todo o Estado, justificadas pela heterogeneidade, natural em qualquer processo educacional. Esse fato demonstra a grande preocupação com a realidade vivenciada pelas escolas.

As áreas de conhecimento avaliadas estão apresentadas no Quadro 3.2, que também indica a quantidade de itens utilizados no teste e as séries testadas.

**Quadro 3.2 – Previsão de instrumentos e anos escolares a serem testados dentro do Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais**

Testes/Questionários		Número de itens por série				
		1º Grau			2º Grau	Habilitação Magistério
		3ª	5ª	8ª	2ª	3ª/4ª
Língua Portuguesa + Redação		30	30	30	30	ó
Matemática		30	30	30	ó	ó
Ciências		30	ó	ó	ó	ó
Conhecimentos Gerais	História	ó	15	15	ó	ó
	Geografia	ó	15	15	-	ó
	Ciências	ó	20	20	-	ó
Estudos Sociais	História	ó	ó	ó	20	ó
	Geografia	ó	ó	ó	20	ó
Matemática/Física		ó	ó	ó	20 + 20	ó
Química/Biologia		ó	ó	ó	20 + 20	ó
Conteúdos básicos para Magistério		ó	ó	ó	-	50
Questionário para aluno		10	30	50	-	50
Questionário de atitude		ó	ó	ó	34	ó
Questionário para escola		50	50	50	50	50

Fonte: VIANNA; ANTUNES; SOUZA, 1993, com elaboração do autor.

Quanto aos itens, o formato escolhido foi o de múltipla escolha, por possibilitar uma correção mais ágil, com exceção para a 3ª série do ensino fundamental, em que o formato escolhido foi o de resposta construída, dadas às características da população.

O Programa de Avaliação não trabalhava com uma Matriz de Referência, baseada no currículo da Secretaria, mas com uma lista que continha dez tópicos, por área de conhecimento, envolvendo os conteúdos básicos a serem testados. Essa lista foi elaborada em conjunto pelos técnicos da Secretaria, por professoras que atuavam no CBA e por professoras das diferentes disciplinas.

Quanto aos questionários respondidos pelos alunos, o objetivo era levantar fatores socioeconômicos e hábitos de estudo, procurando relacioná-los aos desempenhos obtidos. Apenas para a 2ª série do ensino médio, estava previsto um Questionário de Atitudes para levantar informações sobre algum aspecto de interesse, sendo que, para o primeiro ciclo, o tema escolhido foram as atitudes quanto à Ciência. Esse questionário, com 34 questões, tratava da relação da Ciência com a experimentação, com a sociedade e com a escola; como profissão e como processo (SOUZA, 2007). Além dos alunos, a escola também respondia a um questionário, com o objetivo de obter informações sobre sua realidade.

Os resultados eram expressos em uma escala de zero a dez, que indicava o percentual de acertos, tomando por base os dez tópicos testados e a quantidade de itens do teste.

As séries escolhidas voltariam a ser testadas a cada dois anos, mesma periodicidade adotada para o Sistema de Avaliação da Educação Básica (Saeb), período necessário para que houvesse tempo de implementar medidas de intervenção, com vistas a resolver as deficiências de ensino identificadas, permitindo que as melhorias implementadas pudessem se manifestar nos resultados. Nos anos pares, seriam testadas as 3ª e 5ª séries do ensino fundamental e, nos ímpares, os outros anos escolares, reservando-se o ano sem aplicação de testes para fazer a pré-testagem dos itens.

Do Programa de Avaliação, três aspectos relacionados à participação da comunidade escolar merecem destaque (VIANNA; ANTUNES; SOUZA, 1993). O primeiro foi a implantação da figura do controlador de qualidade da aplicação dos testes, função exercida pelos familiares dos alunos, membros do colegiado da escola e indicados por esse órgão para exercer a função. O segundo refere-se à forma como os relatórios eram produzidos para divulgar os resultados, pois envolvia a participação do diretor, do supervisor, do coordenador, das professoras e dos representantes dos pais (VIANNA, 1992). A elaboração do relatório da Secretaria partia da síntese dos relatórios das escolas para a construção dos relatórios por município, a partir dos quais se elaboravam os relatórios por Superintendência Regional de

Ensino, culminando no relatório final. Por fim, o terceiro aspecto diz respeito à participação da professora em todo o processo, comportando-se como um avaliador: na elaboração dos itens, na aplicação dos testes, na elaboração dos relatórios e no uso dos resultados para aprimorar o processo de aprendizagem (SOUZA, 2007).

Com relação à utilização dos resultados do Programa de Avaliação, Vianna, Antunes e Souza (1993) destacam que foram fundamentais para identificar municípios com altos riscos educacionais, embasar a construção de uma proposta de conteúdos básicos de 1ª a 4ª série do 1º grau e subsidiar a programação de uma capacitação docente.

O Programa de Avaliação teve três ciclos: 1992-1993, 1994-1995 e 1996-1997 (SOUZA, 1999). Os dados relativos a esses ciclos estão indicados no Quadro 3.3.

**Quadro 3.3 – Quantitativo dos três primeiros ciclos do Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais – 1992 a 1995**

	Ano	Mês	Série	Alunos	Escolas	Recursos Humanos			
						Regionais	Unidade escolar		
						Técnicos	Direção	Professoras	Pais
1º ciclo	1992	março	3ª 1º grau	325.000	5.570	1.795	11.754	19.698	7.059
	1992	novembro	8ª 1º grau	124.000	2.173	1.441	7.131	6.958	3.555
	1993	novembro	5ª 1º grau	407.443	5.353	1.407	9.052	19.556	4.982
	1993	novembro	2ª 2º grau	69.163	813	987	3.209	5.539	1.367
	1993	novembro	Magistério	6.760	434	874	1.771	1.300	668
2º ciclo	1994	março	3ª 1º grau	356.530	4.764	1.723	13.328	20.627	6.746
	1994	novembro	5ª 1º grau	265.645	5.570	1.145	8.541	20.381	5.632
	1995	novembro	8ª 1º grau	162.000	2.604	1.500	7.500	7.000	4.000
	1995	novembro	2ª 2º grau	102.030	1.044	1.000	3.300	5.600	1.500
	1995	novembro	Magistério	17.782	548	900	1.800	1.500	700
3º ciclo	1996	março	3ª 1º grau	43.941	2.643	s/d	s/d	s/d	s/d
	1996	novembro	5ª 1º grau	431.826	2.261	s/d	s/d	s/d	s/d
<b>Totais</b>			<b>12</b>	<b>2.312.120</b>	<b>-</b>	<b>12.773</b>	<b>67.206</b>	<b>108.159</b>	<b>36.209</b>

Fonte: SOUZA, 1995, p. 30; SOUZA, 2007, p. 58, com elaboração do autor.

Observa-se no Quadro 4.3 que, durante o 1º e 2º ciclos, deve ter havido um grande esforço da Secretaria para a realização do Programa de Avaliação: foram dez testes aplicados, cinco por ano, envolvendo grande quantidade de profissionais da Secretaria, técnicos das Superintendências Regionais de Ensino, direção e professoras das escolas, e mobilizando

grande quantidade de alunos. Merece destaque a quantidade de pais que participaram do processo, cerca de 20% do total de recursos humanos envolvidos no programa.

Os dois primeiros ciclos ocorreram em governos diferentes: Hélio Garcia (1991-1995) e no primeiro ano do governo Eduardo Azeredo (1995-1999). No terceiro ciclo, coincidindo com o segundo ano do governo Eduardo Azeredo, houve perda de ritmo, apesar de os secretários que sucederam Walfrido dos Mares Guia, no comando da Secretaria de Educação, serem ligados a Walfrido: Ana Lucia Machado Pinheiro, subsecretária na gestão anterior, e João Batista dos Mares Guia, irmão do antigo secretário.

Para garantir uniformidade nas ações, uma série de manuais contendo instruções e procedimentos foram elaborados, tratando: da coordenação geral, da coordenação e supervisão local, de aplicação de provas, da apuração de resultados, da correção de redação, do controle da qualidade e da análise e produção de relatórios (SOUZA, 2007).

Apesar do expressivo esforço da Secretaria durante os dois primeiros ciclos do Programa de Avaliação, não se identificou qualquer estudo produzido com sua chancela, e pouquíssimos dados foram disponibilizados. Dos estudos que a pesquisa localizou, todos foram publicados na revista *Estudos de Avaliação Educacional*, editada pela Fundação Carlos Chagas (FCC), a principal parceira do governo mineiro no programa. Destaca-se um estudo muito interessante, baseado no questionário respondido pela escola, quando da aplicação do teste aos alunos da 3ª série do 1º grau, que mostra, entre outras coisas, as resistências das professoras à implantação do Ciclo Básico de Alfabetização, as práticas da escola e as estratégias das professoras para o ensino de Língua Portuguesa, Matemática e Ciências (VIANNA; SQUÁRCIO; VILHENA, 1992). Outros estudos que merecem destaque abordaram os textos produzidos pelos alunos do CBA (MENDES, 1992; SIMÕES, 1996), os resultados dos questionários da escola que oferecia a 8ª série do ensino fundamental (ANTUNES, 1994), o desempenho dos alunos dessa mesma série em História (ALVES, 1994), Geografia (CARDOSO, 1994), Língua Portuguesa (SANTOS, 1994) e Matemática (SILVÉRIO; XAVIER; PORTINHO, 1994). O comum em todos os estudos citados é que figuravam técnicos da Secretaria de Educação entre seus autores, reforçando o caráter de pesquisa do Programa de Avaliação e sua importância para a Secretaria como elemento de estudo da realidade educacional mineira. Além disso, a FCC não estava envolvida apenas com a concepção do Programa de Avaliação, mas também realizava esforços para disseminar os resultados obtidos, reforçando o papel de vanguarda da FCC na pesquisa educacional brasileira.

## **A modificação do Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais e a implantação do Simave**

Em 1998, durante o governo Eduardo Azeredo (1995-1998), o Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais sofre uma grande modificação que praticamente o descaracteriza, quando comparado a seus objetivos iniciais. Com a introdução, nessa época, do Regime de Progressão Continuada, as séries testadas passaram a ser as últimas de cada etapa do ensino fundamental, as 4ª e 8ª séries. As áreas de conhecimento testadas passam a ser Língua Portuguesa e Matemática e mais Redação. Com isso, o programa fica semelhante ao Saeb, do governo federal, sem que tenha sido identificado os motivos da alteração. Além dos questionários do aluno e da escola, introduziram-se o da professora e o do pai ou responsável pelo aluno. Os testes das outras séries e dos cursos de Habilitação do Magistério, que não mais constaram dos testes aplicados em 1996, foram definitivamente eliminados dos testes seguintes, assim como o foram os das outras áreas do conhecimento. A partir daí, no material disponível sobre o Programa, não se encontra mais referência a suas premissas iniciais. O compromisso inicial com a participação da comunidade e, principalmente, das professoras foi deixado de lado. Não foi possível obter informações sobre o que ocasionou essas alterações no Programa de Avaliação, mas, provavelmente, as causas estivessem ligadas à diminuir o esforço de mobilização de recursos humanos e financeiros exigidos da Secretaria.

Nova alteração na avaliação educacional mineira ocorre durante o governo seguinte, de Itamar Franco (1999-2003), sem que, novamente, existam razões claras sobre o que a motivou. A única referência feita é a de que, apesar das diversas avaliações feitas,

[...] o que se constata é que as unidades escolares, em sua maior parte, não tomaram conhecimento dos seus objetivos e não conseguiram perceber com clareza nem os objetivos e nem os desdobramentos resultantes dessas avaliações (SOUZA; OLIVEIRA, 2007, p. 110, *apud* MINAS GERAIS, 2000).

Essa citação parece se referir aos acontecimentos ocorridos no governo Eduardo Azeredo, durante a realização do 3º ciclo do Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais, em 1996. Isso porque, pelos documentos analisados, nos ciclos anteriores ocorreu justamente o oposto, havendo grande mobilização nas escolas.

O novo secretário de educação, Murfílio Hingel, ex-ministro da Educação durante o governo de Itamar Franco como presidente, implanta um projeto denominado Escola Sagarana, que possuía como ideal construir coletivamente a escola, tendo por base sua gestão democrática e os seguintes objetivos: promover a qualidade do ensino e da aprendizagem; melhorar a qualidade do gerenciamento escolar, melhorando a gestão dos recursos; fornecer autonomia gerencial e pedagógica para a escola (MACHADO, 2004). Nesse processo, a avaliação merecia destaque.

Segundo documento sobre a Escola Sagarana (MINAS GERAIS, 2001), o Simave, nova denominação da avaliação educacional, surge nas discussões de um evento chamado Seminário Travessia para o Futuro, realizado em 1999, com o objetivo de discutir estratégias, parcerias e cooperação entre a sociedade, o governo e as instituições de ensino de Minas Gerais. O documento destaca, também, que na formulação do Simave foi feito um acordo com o governo da França para aproveitar a experiência daquele país na área de avaliação educacional, apesar de estar claro, pela estrutura e pelo desenvolvimento dos instrumentos utilizados, que a maior inspiração para o Simave veio do Saeb.

O documento mostra que o Simave foi desenvolvido para testar os alunos bienalmente, com a participação das 27 instituições regionais de ensino superior como parceiras na implementação de políticas públicas e não como simples prestadoras de serviço, pois não se trata apenas de aplicar testes, mas de formular democraticamente, em conjunto com a comunidade educacional, as políticas do setor (MINAS GERAIS, 2001, p. 60). Caberia, então, ao Simave fornecer as diretrizes para a avaliação e para o teste, desenvolvendo e discutindo seus resultados de forma consorciada. O documento da Escola Sagarana mostrava, também, que, com o Simave, as escolas teriam uma participação ativa no processo, e que as diferenças regionais, mesmo em uma avaliação estandardizada, seriam respeitadas de alguma forma.

Quanto ao Saeb, o documento da Escola Sagarana diz que o Simave significaria um avanço, quando comparado ao sistema do governo federal, pelo fato de ser universal, coletando informações de todas as escolas, e de estar adequado às especificidades do Estado, com vistas à formulação de políticas educacionais próprias. Apesar disso, o sistema deveria permitir uma comparação com o sistema nacional, mas não indicava como isso seria realizado. Importa ressaltar que, apesar de os resultados da avaliação serem públicos, o objetivo declarado não era expor indevidamente a escola.

Comparado à proposta do Programa de Avaliação, percebe-se que o Simave tinha como preocupação envolver as instituições regionais de ensino no processo, talvez para se

contrapor à destacada participação dos técnicos de outros Estados no desenvolvimento e no acompanhamento do Programa de Avaliação. De toda a forma, procurava-se envolver as instituições formadoras de professoras no processo avaliativo, esperando com isso influir para aprimorar a formação das futuras professoras. Não se faz referência, nos documentos de apresentação do Simave, ao papel dos técnicos da Secretaria e à necessidade de apropriação das metodologias utilizadas; ao contrário, a coordenação do processo ficou a cargo da Universidade Federal de Minas Gerais (UFMG), por meio do Centro de Políticas e Avaliação da Educação (CAEd), vinculado à Faculdade de Educação, uma das instituições regionais de ensino chamadas a participar do Simave.

O Simave é regulamentado pela Secretaria por meio da Resolução nº 14, de 3 de fevereiro de 2000, reeditada com alterações pela Resolução nº 104, de 14 de julho de 2000. Dos cinco objetivos traçados para o Simave, destacam-se:

[...]

- Criar novos instrumentos de participação da sociedade e dos profissionais da Educação na gestão da Escola Pública;
- Democratizar o acesso à informação sobre a Educação Pública;
- Desenvolver procedimentos de gestão baseados na avaliação continuada das políticas públicas educacionais e em critérios de equidade [...] (MINAS GERAIS, 2000).

Atrelada à ideia de democratizar o acesso à informação, estava a participação tanto da sociedade como dos profissionais da educação, apesar da inexistência dos devidos esclarecimentos sobre como isso aconteceria. Com relação ao acesso às informações da escola, a resolução apenas reafirma o que o documento da Escola Sagarana já havia expressado, mas sem garantir que não haveria uma exposição indevida das escolas. Por fim, aponta a avaliação continuada das políticas públicas como um instrumento de gestão e de busca de equidade. Mesmo com objetivos muito diminutos com relação à proposta do Programa de Avaliação de 1988, os objetivos ainda eram mais amplos do que apenas medir o desempenho das escolas.

São relacionados na resolução oito princípios gerais que deveriam fundamentar a avaliação contínua da rede pública. Entre eles, alguns merecem relevo, quais sejam: a participação dos profissionais da educação básica na implementação do Simave; a centralidade da escola, percebendo-se a educação básica como um processo coletivo e

institucionalizado; a geração de insumos para as políticas de formação inicial e continuada da professora; e estar a serviço da sociedade civil e da democracia participativa.

Como instrumento do Sistema, é instituído o Programa de Avaliação da Rede Pública de Educação Básica (Proeb). Determinava-se que a coordenação do Proeb deveria ficar a cargo de uma das universidades públicas sediadas em Minas, que seria escolhida por licitação pública. A essa instituição, estariam ligadas as atividades de coordenação das atividades, a discussão e a divulgação dos resultados e o apoio aos trabalhos realizados pelas outras instituições de educação superior que atuavam em parceria com as Superintendências Regionais de Ensino. A Resolução estabelecia que deveriam ser testados, a cada ciclo de dois anos, alunos das 4<sup>a</sup> e 8<sup>a</sup> séries do ensino fundamental e da 3<sup>a</sup> série do ensino médio, testando, alternadamente, em um ciclo, as áreas de Língua Portuguesa e Matemática e, no seguinte, as de Ciências Humanas e Ciências da Natureza. Estabelece, ainda, que os testes deveriam ser construídos com a ajuda das professoras das escolas públicas. A resolução, em seu artigo 17, determina também que depois da aplicação dos testes, deveriam ser implementados, nas Regionais, subprogramas de avaliação continuada, para promover o debate sobre os resultados, elaborar e implementar estratégias de enfrentamento dos problemas identificados e refletir sobre as práticas correntes nas unidades escolares (MINAS GERAIS, 2000).

Pelo exposto, percebe-se a Secretaria construindo o Simave como um instrumento não somente centrado nas testagens dos alunos, mas como uma ferramenta a serviço da avaliação de políticas públicas, calcado na participação das diferentes instituições educacionais; das professoras, inclusive no desenvolvimento dos testes; e da sociedade. Dessa forma, mantinha-se a proposta original do Programa de Avaliação e os princípios de uma avaliação de qualidade. Por meio da participação das instituições de educação superior em todo o processo do Simave, percebe-se claramente uma preocupação com os resultados chegarem até os responsáveis pela formação das professoras, de modo a influenciar diretamente a qualidade dos novos profissionais formados.

Um fato histórico importante foi apurado por Souza e Oliveira (2007). Segundo os autores, um Projeto de Lei de 2001 propunha a criação de uma Agência Mineira de Avaliação Educacional como responsável pelo Simave, prevendo a aplicação de seus instrumentos. No entanto, a proposta não foi implementada.

Analisando-se o período discutido, em pouco mais de dez anos, sem que fossem feitos estudos para embasar as mudanças implementadas, a avaliação educacional em Minas mudou três vezes ao sabor dos novos governos eleitos. Passa-se de uma proposta muito próxima de uma avaliação cooperativa, como previsto na Constituição estadual, um projeto inovador e



centrado no processo educacional e com preocupações com a escola e seus integrantes, para outra, em que ainda se previa alguma participação, mas que estava centrada nos resultados.

Com o passar de alguns poucos anos, a única preocupação seria com os resultados, e os integrantes da escola seriam meros coadjuvantes do processo, discutindo seus resultados a partir de um roteiro programado pela Secretaria, com o objetivo de produzir relatórios para serem encaminhados aos órgãos de gestão. Em uma pesquisa sobre o uso dos resultados do Simave por diretores escolares de Minas Gerais, Silva (2011) apresenta o final desse processo, ao concluir, com base nas entrevistas coletadas, que os gestores se sentem obrigados a intervir na realidade escolar a fim de alterar, não importa como, os índices de proficiência da escola (p. 198). Essa conclusão exemplifica claramente o que Barroso (2005) chamou de *regulação autônoma*, na qual os próprios atores adaptam a regulação institucional às suas conveniências, em um processo de adaptação às regras do jogo.

## **Os instrumentos do Simave**

O Simave possui três instrumentos: o Proeb, que o acompanha desde seu nascimento, conforme comentado anteriormente; o Programa de Avaliação da Alfabetização, (Proalfa); e o Programa de Avaliação da Aprendizagem Escolar (PAAE). Apesar de este último estar sendo utilizado no ensino médio atualmente, existe o objetivo de que ele seja expandido para os anos finais do ensino fundamental.

Para compreender melhor a lógica do Simave, é importante discutir sobre esses três instrumentos, como surgiram, como são utilizados e que resultados produzem.

### ***Programa de Avaliação da Rede Pública de Educação Básica (Proeb)***

O primeiro dos instrumentos do Simave foi o Proeb, que nasceu com a regulamentação do Sistema, com o objetivo de medir o desempenho dos alunos em Língua Portuguesa e Matemática.

O documento sobre a proposta da Escola Sagarana registra que os primeiros resultados do Proeb começaram a ser discutidos na rede de ensino a partir da instalação do Sistema de Ação Pedagógica (Siape) (MINAS GERAIS, 2001). Esse sistema envolvia uma série de materiais escritos e em vídeo relativos ao Simave e seus resultados, cuja discussão deveria ser

realizada quinzenalmente em reuniões nas escolas. Essas discussões seriam sintetizadas em um documento encaminhado às Regionais de Ensino, que identificariam as necessidades das escolas com o objetivo de desenvolver ações de apoio a elas.

A estrutura do teste segue uma lógica parecida com aquela utilizada pelo Saeb até o ano 2003: são utilizados, para cada ano escolar testado, 169 itens de Língua Portuguesa e a mesma quantidade de Matemática. Os itens de cada área são agrupados em 13 diferentes blocos, contendo cada um 13 itens. A partir desses blocos, são construídos 26 cadernos de prova, cada um dos quais contendo três blocos e 39 itens. Essa organização pode ser mais bem compreendida mediante o Quadro 3.4.

**Quadro 3.4 – Organização dos cadernos de teste, a partir da divisão dos 169 itens em blocos e cadernos.**



Fonte: MINAS GERAIS, 2010a, p. 34.

Utilizando-se um modelo estatístico baseado na TRI, é determinado o desempenho médio dos alunos da escola. Os desempenhos do Proeb são expressos, como no Saeb, em uma escala que varia de zero a 500 pontos. A aplicação dos testes ocorre no final do mês de novembro.

Os itens são elaborados pela equipe do CAEd e são pré-testados antes de serem incluídos no teste. Dessa forma, o item da Resolução SEE nº 14/2000, de instituição do Simave, que determina a construção dos testes com a ajuda das professoras, foi desconsiderado, pois quase todas as etapas do processo estão fora do controle das professoras. A única etapa em que elas são chamadas a participar é na aplicação dos testes.

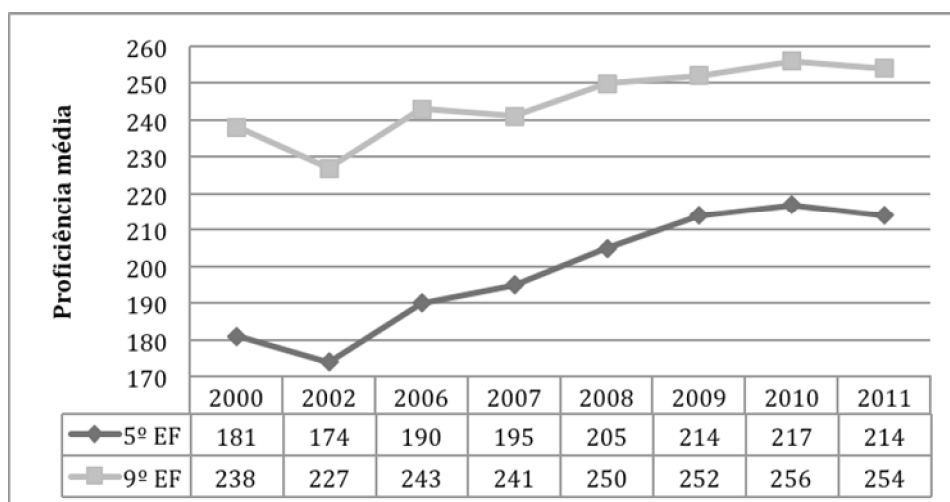
A divulgação dos resultados é feita por meio de quatro instrumentos: a *Revista do Sistema de Avaliação*, que apresenta as Matrizes de Referência, a composição do teste e sua metodologia de análise; a *Revista do Gestor*, que apresenta os resultados por Superintendência Regional de Ensino e pelos municípios e escolas sob sua jurisdição; o *Boletim Pedagógico da Escola*, destinado à escola e contendo a interpretação pedagógica dos resultados, os padrões de desempenho e a análise de alguns itens; e a *Revista Fatores Associados*, que relaciona os fatores intra e extraescolares que interferem no desempenho dos alunos, obtidos por meio da correlação entre o desempenho demonstrado e esses fatores, identificados nos questionários respondidos pelos alunos, professoras e diretores de escolas, e que apresenta, também, o Índice de Eficácia da Escola, o qual demonstra o desempenho da escola, levando em consideração o nível socioeconômico dos alunos.

Nos dois primeiros anos de criação do Simave, a Resolução 14 foi cumprida na íntegra, pelo menos no que tange à aplicação dos testes. Assim, em 2000, o Proeb testou as áreas de Língua Portuguesa e Matemática e, em 2001, as áreas de Ciências Humanas e Ciências da Natureza. Em 2002, quando se deveria repetir o ciclo do ano 2000, testou-se apenas Língua Portuguesa, sem que tenha sido encontrado, nos documentos disponíveis, qualquer razão para a mudança.

Em 2003, durante o primeiro ano do governo de Aécio Neves (2003-2007), em que deveria estar se repetindo o ciclo de 2001, conforme estabelecia Resolução 14, apenas a área de Matemática foi testada. Em 2004 e 2005, o Proeb não foi aplicado e desde 2006 vem sendo aplicado regularmente todos os anos, testando apenas Língua Portuguesa e Matemática. Novamente, nenhuma razão para essas alterações se encontra nos documentos disponíveis.

## A expressão dos resultados do Proeb

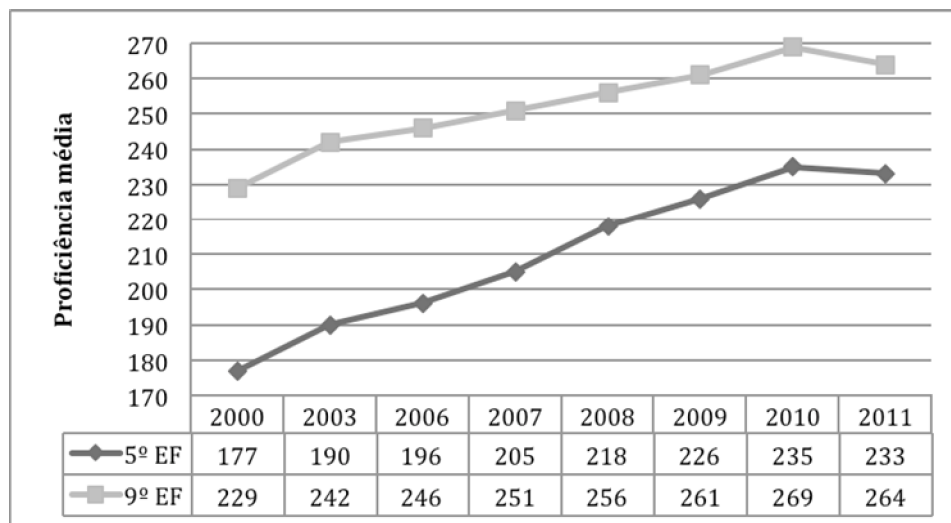
Os resultados do Proeb vêm sendo divulgados em diversas publicações da Secretaria de Educação, com destaque para a proficiência média dos alunos. Os Gráficos 3.1 e 3.2 apresentam essa proficiência média entre 2000 e 2011, para as áreas de Língua Portuguesa e Matemática, respectivamente.



**Gráfico 3.1 – Proficiência média de Língua Portuguesa no Proeb para os 5º e 9º anos do ensino fundamental<sup>122</sup> da rede estadual – 2000 a 2011**

Fonte: MINAS GERAIS, 2010; MINAS GERAIS, 2011; ATLAS DA EDUCAÇÃO (disponível em: <<http://crv.educacao.mg.gov.br/atlas>>), com elaboração do autor.

<sup>122</sup> Minas Gerais iniciou a implantação do ensino fundamental de nove anos em 2004. Para garantir coerência com os documentos oficiais e permitir comparações com os resultados das medidas efetuadas antes desse ano, neste trabalho foi feita a correspondência entre as séries do ensino fundamental de oito anos com os anos do ensino fundamental de nove anos.



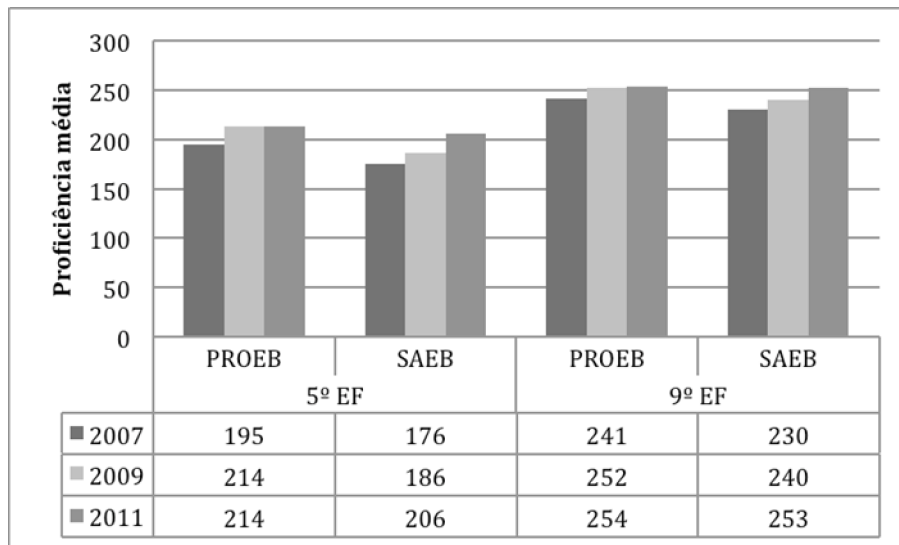
**Gráfico 3.2 – Proficiência média de Matemática no Proeb para os 5º e 9º anos do ensino fundamental da rede estadual – 2000 a 2011**

Fonte: MINAS GERAIS, 2010; MINAS GERAIS, 2011; e ATLAS DA EDUCAÇÃO (disponível em: <<http://crv.educacao.mg.gov.br/atlas>>), com elaboração do autor.

Pela análise dos gráficos, é possível perceber um crescimento constante da proficiência entre 2006 e 2010 e uma queda em 2011. Essa queda não foi comentada nos documentos oficiais do Proeb, tanto que no *Boletim Pedagógico* da Escola, de 2011, em que tradicionalmente se discutem e se apresentam os resultados obtidos, nada se comenta sobre o desempenho daquele ano. A queda da média de proficiência em 2011 foi comentada somente em matéria jornalística divulgada pela assessoria de imprensa do governo do Estado, destacando que apesar da queda, a tendência de crescimento é constante (MINAS GERAIS, 2011). Claramente um paradoxo! Repete-se aqui a omissão no enfrentamento da situação. Quando existe uma queda de desempenho, em vez de buscar explicações consistentes sobre o problema, que poderia até ser associado a algum erro de medida, esconde-se o fato ou, então, justifica-se a queda tomando por base fatos positivos anteriores. Com a importância dada ao número que expressa a média de proficiência, qualquer queda, por menor que seja, torna-se um fator difícil de ser tratado, tendo-se que recorrer, como nesse caso, a argumentos esdrúxulos.

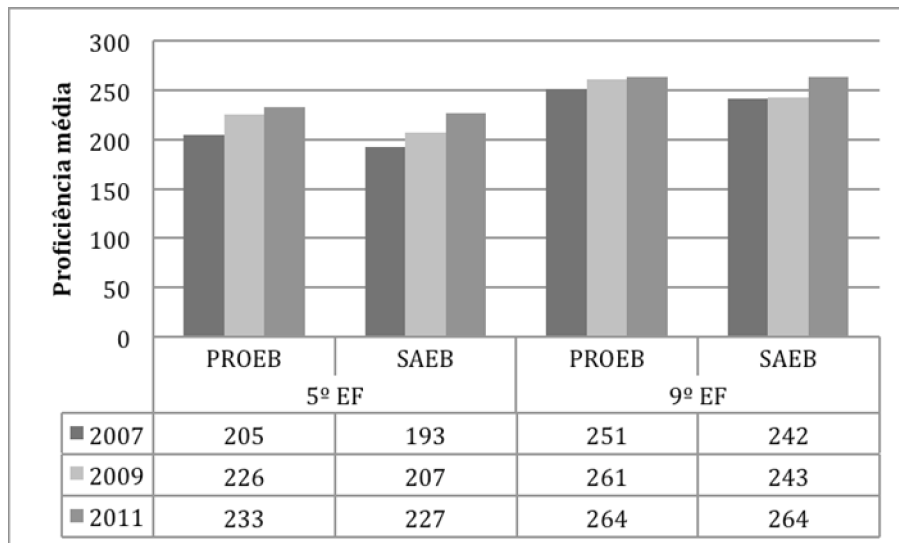
Como a escala de proficiência do Proeb passa por um processo de equalização com a escala do Saeb, as duas podem ser comparadas e, assim, o comportamento de uma escala deve acompanhar o da outra. Para comparar os resultados obtidos pelo Proeb e observar se o padrão de desempenho se mantém quando se utilizam os dados do Saeb, e o que pode ter

ocorrido em 2011, foram construídos os Gráficos 3.3 e 3.4, para os anos em que existem dados para os dois testes, ou seja, 2007, 2009 e 2011.



**Gráfico 3.3 – Proficiência em Língua Portuguesa para os 5º e 9º anos da rede estadual de ensino fundamental no Saeb e no Proeb entre 2007 e 2011**

Fonte: INEP, s/d (<disponível em <http://sistemasprovabrazil2.inep.gov.br/resultados>>); Minas Gerais, 2012, com elaboração do autor.



**Gráfico 3.4 – Proficiência média em Matemática para os 5º e 9º anos da rede estadual de ensino fundamental no Saeb e no Proeb entre 2007 e 2011**

Fonte: INEP, s/d (disponível em: < <http://sistemasprovabrazil2.inep.gov.br/resultados>>); Minas Gerais, 2012, com elaboração do autor.

Primeiramente, pelos Gráficos 3.3 e 3.4, percebe-se que as proficiências do Saeb e do Proeb seguem a mesma tendência, o que era esperado. No entanto, percebe-se também que existe uma variação que pode chegar até 19 pontos, quando se comparam as duas medidas, para Língua Portuguesa, em 2007, e para Matemática, em 2009. Não foi encontrada qualquer explicação nos documentos disponíveis sobre as razões dessa diferença. Apesar disso, se as duas medidas seguem a mesma tendência, o que pode ter acontecido é que a queda de 2011, quando comparada com 2010 tenha sido causada por problemas na calibração dos itens do Proeb para a escala do Saeb em 2010. Ou seja, os valores apurados em 2010 foram maiores do que deveriam ser. O fato de não se tratar o evento com transparência só contribui para gerar mais desconfianças quanto aos testes avaliativos.

Além de divulgar as médias de proficiência, os resultados do Proeb também são divulgados a partir de alguns padrões de desempenho. São eles:

- Recomendado: [...] são capazes de realizar tarefas que exigem habilidades mais sofisticadas. Eles desenvolveram habilidades esperadas para o período de escolaridade em que se encontram.
- Intermediário: [...] demonstram ter ampliado o leque de habilidades tanto no que diz respeito à quantidade quanto no que se refere à complexidade dessas habilidades, as quais exigem um maior refinamento dos processos cognitivos nelas envolvidos.
- Baixo: [...] revelam ter desenvolvido competências e habilidades muito aquém do que seria esperado para o período de escolarização em que se encontram. Por isso, este grupo de alunos necessita de uma intervenção focada, de modo a progredirem com sucesso em seu processo de escolarização. Para esse grupo de estudantes, é importante o investimento de esforços, para que possam desenvolver habilidades mais elaboradas (MINAS GERAIS, 2011a, p. 49).

Como se observa, esses padrões classificam o desempenho das escolas segundo o que seus alunos seriam capazes de executar, com base nos testes aplicados. Apesar de não estarem identificados os critérios utilizados pela Secretaria para estabelecer esses padrões, as informações expressas em cada um deles são muito genéricas e dificilmente teriam utilidade como referência para qualquer ação, seja da Secretaria, das escolas ou das professoras.

Parece que se buscou dar uma interpretação sintética à média de proficiência. Mas o resultado prático é que se adjetiva a escola com base em medidas muito limitadas que não traduzem a realidade da escola, seja porque são testadas apenas duas áreas de conhecimento, seja porque, mesmo para essas áreas, nem tudo o que é trabalhado em sala é passível de ser

medido e mesmo que o fosse, por conta do tamanho limite que um teste deve ter para não cansar os alunos, nem tudo poderia ser transformado em item de teste. Portanto, a divulgação desse tipo de informação serve apenas como mecanismo de pressão sobre a escola e seus integrantes e pouco auxiliam no desenvolvimento do processo educacional. Esse processo é típico do que Maroy (2011) chamou de regulação por resultados, baseado na lógica de regulação das escolas e de seus agentes pelos resultados.

Nos relatórios de cada área testada e para cada ano escolar, procura-se agregar algumas informações adicionais, talvez com o sentido de melhorar a compreensão dos padrões estabelecidos. Um exemplo disso é a expressão do padrão recomendado para Língua Portuguesa do 5º ano do ensino fundamental:

Os alunos que apresentam esse padrão de desempenho conseguem realizar tarefas que exigem habilidades de leitura mais sofisticadas, pois, além de conseguirem localizar informações que se encontram explícitas em textos mais extensos e com vocabulário mais sofisticado, realizam inferências que requerem o recurso ao sentido global do texto, articulam partes do texto no processo de produção de sentidos e, ainda, começam a estabelecer relações entre textos. Esses alunos *desenvolveram habilidades que superam aquelas esperadas para o período de escolaridade em que se encontram* (MINAS GERAIS, 2010c, p. 29, grifo do autor).

O trecho destacado assinala que os alunos desenvolveram habilidades que superam aquelas esperadas, enquanto que a descrição genérica dos padrões de desempenho afirma que os alunos desenvolveram habilidades esperadas, duas descrições contraditórias. A descrição mais detalhada para o padrão recomendado mostra as habilidades de Língua Portuguesa que os alunos do 5º ano do ensino fundamental demonstraram dominar a partir do resultado do teste. O problema é que a informação ainda continua genérica, pois, por exemplo, o que seriam textos mais extensos e com vocabulário mais sofisticado? Essa informação está referenciada nos textos utilizados para compor os testes. Dessa forma, a possível mensagem que chega às professoras é a de que devem ser utilizados textos e vocabulários similares aos apresentados no teste, contribuindo, assim, para o estreitamento do currículo e a diminuição da autonomia da professora, de adaptar o currículo às necessidades do processo educacional e de seus atores. Além disso, a informação continua genérica, tanto para o órgão gestor como para as escolas e suas professoras, com uma capacidade muito diminuta de influenciar positivamente a aprendizagem dos alunos. Apesar de não ser detalhado aqui, o padrão recomendado para Matemática tem análise semelhante. Essa ação da Secretaria revela uma



preocupação em regular todo o sistema educacional, como discutido em Maroy (2011) e Barroso (2005).

A cada um desses padrões e para cada ano escolar e área do conhecimento, a Secretaria fez a correspondência com uma pontuação na escala, conforme pode ser observado no Quadro 3.5.

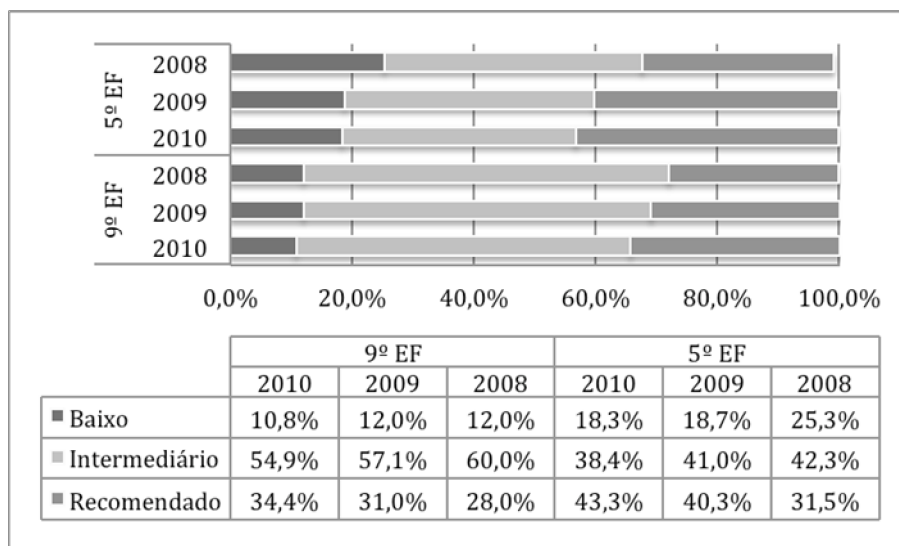
**Quadro 3.5 – Padrões de desempenho no Proeb para os 5º e 9º anos do ensino fundamental – Língua Portuguesa e Matemática**

Etapa de escolaridade	Padrões de desempenho		
	Baixo	Intermediário	Recomendado
<i>Língua Portuguesa</i>			
5º ano EF	até 175	175 a 200	acima de 225
9º ano EF	até 200	200 a 275	acima de 275
<i>Matemática</i>			
5º ano EF	até 175	175 a 225	acima de 225
9º ano EF	até 225	225 a 300	acima de 300

Fonte: MINAS GERAIS, 2010b, com elaboração do autor.

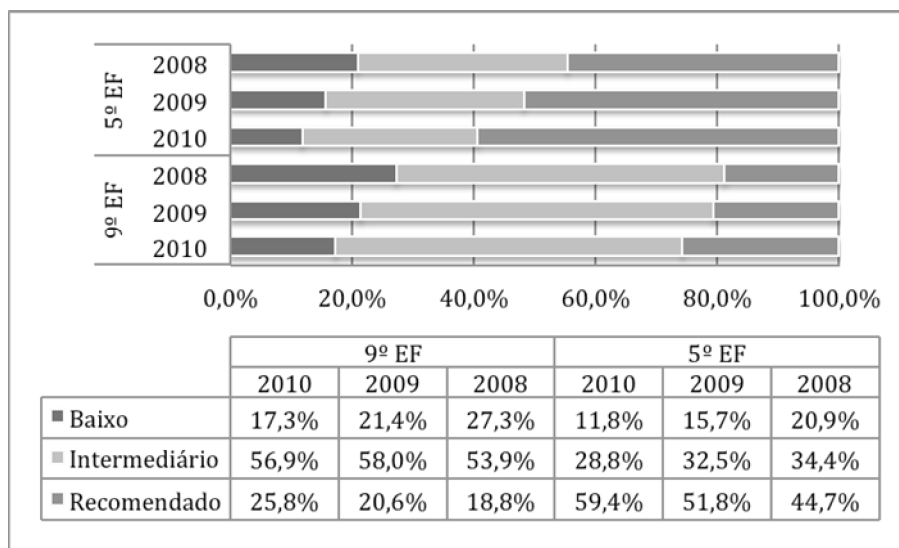
Assim como não existe nenhuma informação sobre como foram definidos os padrões de desempenho, o mesmo acontece com a associação feita entre os intervalos de proficiência e os padrões de desempenho. Mesmo assim, a Secretaria determina às escolas o que se espera delas. Dessa forma, o padrão recomendado tanto em Língua Portuguesa como em Matemática para as escolas que oferecem o 5º ano do ensino fundamental é que seus alunos apresentem um desempenho acima de 225 pontos. Mas qual o sentido disso para o processo educacional dessas crianças? E de que forma essa meta pode auxiliar o planejamento das escolas? Como o que se tem são apenas informações genéricas, as respostas a essas duas questões mostram que esse número só tem sentido para quem o define e que ele provavelmente direciona o ensino para a preparação com vistas ao teste, reforçando que o desempenho é mais importante que a aprendizagem.

Os resultados do Proeb também são apresentados na forma de porcentagem do número de alunos em cada um dos três padrões de desempenho estipulados pela Secretaria. Essa distribuição pode ser vista nos Gráficos 3.5 e 3.6.



**Gráfico 3.5 – Percentual de alunos distribuídos por padrão de desempenho para os anos 5º e 9º do ensino fundamental e para a 3ª série do ensino médio da rede estadual na área de Língua Portuguesa no Proeb, entre 2008 e 2010.**

Fonte: MINAS GERAIS, 2010b, com elaboração do autor.



**Gráfico 3.6 – Percentual de alunos distribuídos por padrão de desempenho para os anos 5º e 9º do ensino fundamental e para a 3ª série do ensino médio da rede estadual na área de Matemática no Proeb, entre 2008 e 2010.**

Fonte: MINAS GERAIS, 2010b, com elaboração do autor.

Percebe-se, nos dois casos, que, com o passar dos anos, amplia-se o percentual de alunos classificados no Padrão de Desempenho Recomendado. Esse fato é considerado pela Secretaria como uma prova do avanço da qualidade educacional da rede estadual de Minas. Porém essa qualidade tem uma abrangência muito restrita, pois não necessariamente o fato de

o percentual de alunos no padrão recomendado ter se ampliado e o percentual dos outros dois padrões estar diminuindo significa que a aprendizagem esteja ocorrendo de fato.

### ***Programa de Avaliação da Alfabetização***

O Proalfa retoma uma das ideias do Programa de Avaliação, proposta coordenada por Heraldo Vianna em 1992, de se acompanhar com dados objetivos o processo de alfabetização nos anos iniciais do ensino fundamental. A proposta de 1992 previa testes em Língua Portuguesa e Matemática, aplicados no início da antiga 3ª série, normalmente em março, para avaliar a alfabetização ao final da 2ª série.

O teste do Proalfa, inserido no Simave em 2005, foi estruturado e tem seus itens e testes elaborados pelo Centro de Alfabetização, Leitura e Escrita (Ceale), ligado à Faculdade de Educação da UFMG, e aplicados pelo CAEd. O teste, apenas de Língua Portuguesa, é aplicado a amostras de alunos do 2º e do 4º ano do ensino fundamental, e seus resultados são expressos em uma escala de proficiência que varia de zero a mil pontos. Dessa forma, não se comparam os resultados dos testes do Proeb com os do Proalfa, pois os dois estão em escalas diferentes. Como os dois testes utilizam constructos distintos, o fato de não ser possível a comparação dos resultados é salutar, apesar de talvez isso não ser compreendido pela rede.

São também testados, de forma censitária, os alunos do 3º ano e aqueles matriculados no 4º ano que tiveram desempenho inferior a 450 pontos quando eles ainda estavam cursando o 3º ano. Todos os alunos respondem ao mesmo teste, que contém 28 itens, aplicado ao final do mês de maio. Com base nos resultados, são gerados relatórios de desempenho por escola e por aluno, sendo que, neste caso, as informações são enviadas para a família.

### **A expressão dos resultados do Proalfa**

Assim como ocorre no Proeb, a Secretaria estipulou níveis de desempenho para cada um dos anos testados, conforme pode ser observado no Quadro 3.6.

**Quadro 3.6 – Níveis de desempenho no Proalfa para os 2º, 3º e 4º ano do ensino fundamental em Língua Portuguesa**

Níveis de desempenho	Etapa de escolaridade		
	2º ano	3º ano	4º ano
Baixo	abaixo de 350	abaixo de 450	abaixo de 500
Intermediário	de 350 a 450	de 450 a 500	de 500 a 600
Recomendado	acima de 450	acima de 500	acima de 600

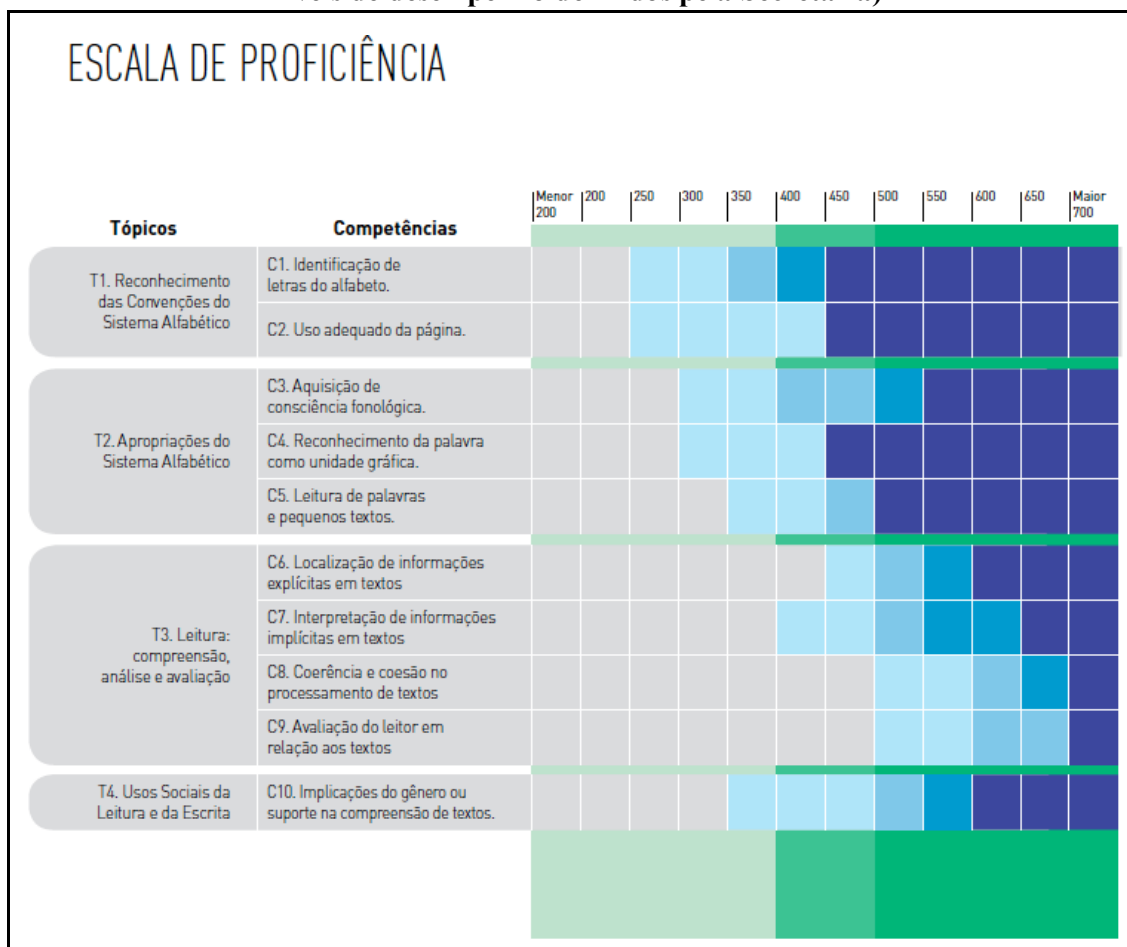
Fonte: MINAS GERAIS, 2008, elaborado pelo autor.

Os documentos disponíveis para consulta, da mesma forma como ocorreu com o Proeb, não explicitam as razões pelas quais se atribuíram os valores fixados para cada nível.

Note-se que no Proeb utiliza-se a expressão “Padrões de Desempenho” e no Proalfa utiliza-se “Níveis de Desempenho”. Pelo que se percebe pelos Quadros 3.5 e 3.6, as duas expressões tratam de expor a mesma coisa, e, portanto, seria importante unificar as expressões para não prejudicar a compreensão do leitor. Esse fato evidencia um problema nos documentos do Simave, pois, dependendo do teste, são usadas expressões diferentes para expressar a mesma coisa. Uma possível razão para isso é o fato de os materiais de cada teste terem sido desenvolvidos por instituições diferentes ó o Proeb a cargo do CAEd e o Proalfa a cargo do Ceale ó, e não ter havido a necessária compatibilização de termos comuns. Se essa é a razão, isso pode significar um problema de coordenação da Secretaria.

A partir da análise dos itens do 3º ano utilizados nos diferentes testes do Proalfa, montou-se a escala de proficiência e parte dela é mostrada no Quadro 3.7. Nele são indicadas algumas das competências testadas, de acordo com a Matriz de Referência do Proalfa; o grau de complexidade de cada uma, indicada em cinza e diversos tons de azul; e a relação delas com os níveis de desempenho, indicados em três tons de verde.

**Quadro 3.7 – Escala de proficiência do ciclo de alfabetização para o 3º ano do ensino fundamental (em azul os níveis de competência indicados por especialistas e em verde os níveis de desempenho definidos pela Secretaria)**



Fonte: MINAS GERAIS, 2011b, p. 21.

Cada competência tem uma complexidade que lhe é própria, e a variação das cores indica como se encontra o desenvolvimento de cada uma delas para o conjunto de alunos testados. A cor cinza informa que a competência, para os alunos posicionados em determinado ponto da escala, ainda não foi desenvolvida. O tom mais escuro de azul informa os pontos da escala em que a competência já foi consolidada pelo aluno; o azul mais claro de todos informa o ponto da escala em que a competência está no começo do desenvolvimento, e os demais tons de azul informam sobre os estágios intermediários de desenvolvimento de determinada competência. Nessa forma de expressar as complexidades das diferentes competências, parece existir uma preocupação pedagógica em não rotular o desempenho dos alunos, mas sim identificar em que etapa do processo de aquisição de competências eles se

encontram depois da aplicação do teste. Assim, por exemplo, no tópico 3. Leitura: compreensão, análise e avaliação, a competência C7 – Interpretação de informações implícitas em textos começa seu desenvolvimento em 400 pontos e se consolida a partir dos 650 pontos, enquanto que a competência C9 – Avaliação do leitor em relação aos textos começa seu desenvolvimento aos 450 pontos, mas só se consolida a partir dos 700 pontos. Dessa forma, procura-se expressar que existe um crescente nas aprendizagens e que cada competência tem seu próprio tempo de maturação, que se apresenta em diferentes estágios do desenvolvimento do aluno.

Apesar dessa aparente preocupação pedagógica, em nenhum dos documentos disponíveis existe explicação de como foi o processo de construção da graduação de cores da escala. Sendo assim, fica difícil contestar ou afirmar sua validade. Além disso, essa informação pode ser útil para os especialistas da área no sentido de eles compreenderem melhor os mecanismos de desenvolvimento das diferentes competências e, a partir daí, aprofundarem ou desenvolverem novas teorias. Para a professora em sala de aula, ou para as escolas, não parece que essa seja uma informação útil para orientar suas atividades, principalmente por estas serem genéricas.

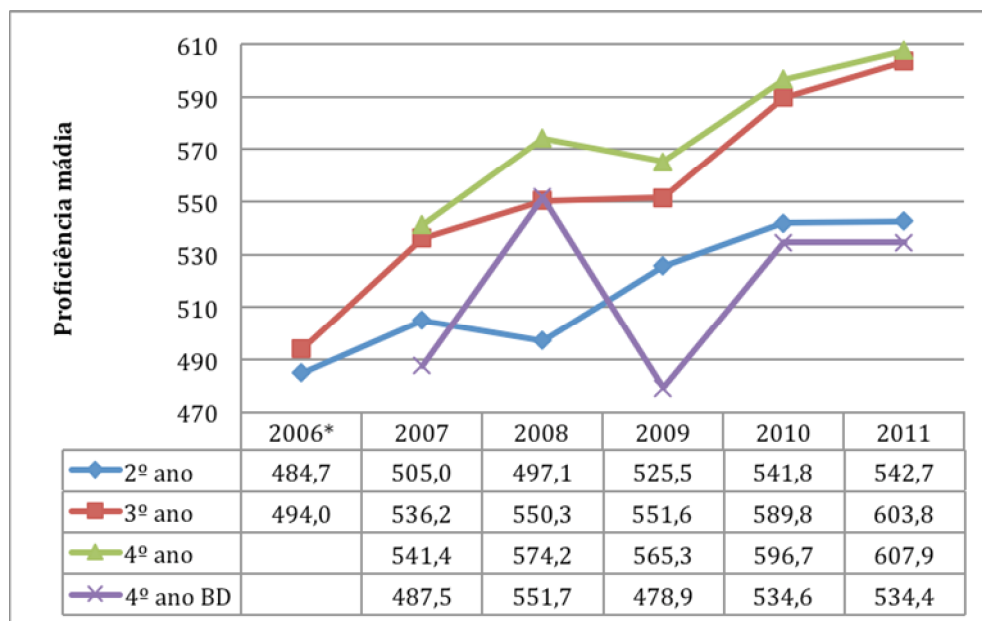
Os níveis de desempenho, indicados pelos três tons de verde, informam sobre as metas fixadas pela Secretaria, metas essas usadas para classificar as escolas. O tom de verde mais escuro, a partir dos 500 pontos, informa o nível de desempenho considerado adequado; o tom médio, entre 400 e 500 pontos, informa o nível intermediário; e menos de 400 pontos informa o nível baixo de desempenho.

Comparando-se as duas informações, o estágio de desenvolvimento das competências e os níveis de desempenho, percebem-se as possíveis dificuldades que as professoras enfrentarão para lidar com elas. Suponha uma escola que se encontra no nível 500, nível nomeado como recomendado, portanto aquele que deve ser perseguido por todas as escolas da rede. Pelo Quadro 3.7, é possível verificar que, para esse nível, as competências C1, C2, C4 e C5 estão consolidadas, enquanto que as competências C8 e C9 estão em início de desenvolvimento. Assim, o recomendado, nesse caso, é que os alunos tenham desenvolvido plenamente um grupo de competências e que outras ainda estejam em desenvolvimento? Qual o sentido pedagógico de uma informação como essa? Será que a partir dessa informação as professoras deixarão de se preocupar, em seu plano de curso, com as competências C8 e C9?

O que parece é que os instrumentos utilizados para interpretar os resultados foram construídos por equipes com pensamentos divergentes sobre como utilizá-los. Se isso de fato ocorreu, aquela que foi a responsável por fixar os níveis de desempenho deve acreditar que a

pressão por resultados contribui para a melhoria do processo educacional, enquanto aquela que ficou com a responsabilidade de expor a escala de proficiência parece acreditar que o desenvolvimento dos alunos é um processo e que existem tempos diferentes de maturação que devem ser respeitados. Por serem pensamentos aparentemente divergentes, quando colocados lado a lado, trazem mais dúvidas que esclarecimentos. Assim, quando chegam à escola, podem dificultar a compreensão dos resultados e trazendo dificuldades para trabalhar com essas informações

A cada aplicação do teste, são calculadas as proficiências médias para cada ano escolar. Os valores obtidos com o passar dos anos podem ser observados no Gráfico 3.7.



**Gráfico 3.7 – Proficiência média no Proalfa para os alunos do 2º, do 3º e do 4º ano BD da rede estadual do ensino fundamental, entre 2005 e 2011**

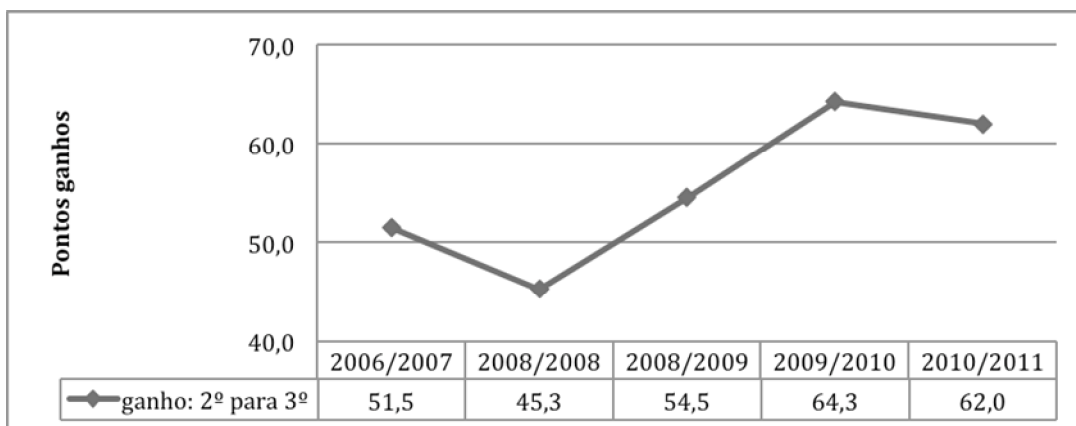
Fonte: RIANI; SOARES, 2008; MINAS GERAIS, 2008, 2009, 2011c, com elaboração do autor.

Nota: (\*) Apesar do primeiro teste do Proalfa ter sido aplicado em 2005, os resultados desse ano não foram calculados usando-se a TRI; por isso, os dados expostos começam em 2006; nesse ano, não foram testados os alunos do 4º ano, nem os alunos desse ano que tiveram no ano anterior baixo desempenho (BD) quando cursavam o 3º ano.

Antes de analisar os gráficos, alguns detalhes devem ser esclarecidos. Primeiramente, os resultados do desempenho dos alunos para os 2º e 4º anos são obtidos de forma amostral, existindo, portanto, um erro amostral associado a essas medidas, que, dependendo do ano-calendário e do ano escolar, variou de 0,3 a 7 pontos na escala. Esses erros não foram indicados no gráfico para não dificultar a visualização de sua totalidade. No caso dos testes

aplicados aos alunos dos 3º e 4º anos com baixo desempenho (BD), não existe erro amostral, pois a aplicação foi censitária.

Analisando-se o Gráfico 3.7, é possível observar que com o passar dos anos há uma tendência de as médias de proficiência dos alunos do 3º ano aumentarem, quando comparadas às do 2º ano, algo esperado, apontando um ganho de aprendizagem na passagem de um ano para outro. Assim, por exemplo, a proficiência média dos alunos do 2º ano, em 2006, foi de 484,7 pontos e quando esse grupo chega ao 3º ano, em 2007, a proficiência média sobe para 550,3 pontos. Essa variação, usada para exemplo, indica um ganho de 51,5 pontos. Fazendo-se o mesmo exercício para os outros anos-calendário, obtêm-se os ganhos de desempenho de: 45,3 pontos do 2º ano em 2007 para o 3º ano em 2008; 54,5 pontos do 2º ano em 2008 para o 3º ano em 2009; 64,3 pontos do 2º ano em 2009 para o 3º ano em 2010; 62 pontos do 2º ano em 2010 para o 3º ano em 2011. Isso pode ser verificado no Gráfico 3.8.



**Gráfico 3.8 – Ganho de proficiência média na passagem do 2º para o 3º ano da rede estadual do ensino fundamental, medido pelo Proalfa**

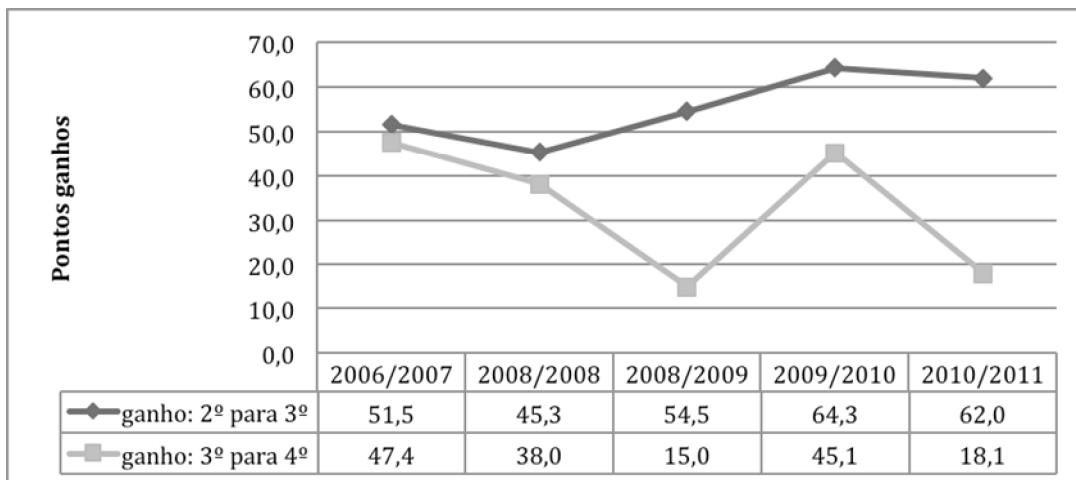
Fonte: Elaboração do autor.

Observa-se no Gráfico 3.8 que entre o 2º e o 3º ano o Proalfa mediu um ganho que variou entre 45 e 62 pontos, com tendência de crescimento com o passar dos anos, revelando dois fenômenos desejáveis: ganho de proficiência entre os anos escolares e tendência de elevação na proficiência média. No entanto, se por um lado esses dados podem refletir uma mudança importante no sistema educacional, por outro, podem estar levando outras áreas que deveriam estar sendo trabalhadas, Matemática, Estudos Sociais e Ciências, a serem deixadas de lado.

A mesma análise relativa aos ganhos obtidos na passagem do 2º para o 3º ano pode ser feita com relação à passagem do 3º para o 4º ano, obtendo-se ganhos de: 47,4 pontos do 3º



ano em 2006 para o 4º ano em 2007; 38 pontos do 3º ano em 2007 para o 4º ano em 2008; 15 pontos do 3º ano em 2008 para o 4º ano em 2009; 45,1 pontos do 3º ano em 2009 para o 4º ano em 2010; 18,1 pontos do 3º ano em 2010 para o 4º ano em 2011. Esses dados, colocados lado a lado com os ganhos do 3º para o 4ª ano, são expostos no Gráfico 3.9.



**Gráfico 3.9 – Ganho de proficiência média na passagem do 2º para o 3º e do 3º para o 4º ano do ensino fundamental da rede estadual, medido pelo Proalfa**

Fonte: Elaboração do autor.

Nota-se que os ganhos do 3º para o 4º ano são irregulares com o passar dos anos, apesar de serem menores do que os obtidos na passagem do 2º para o 3º ano. É como se, medida pela proficiência média, a aprendizagem dos alunos diminuísse de ritmo com o avanço da escolaridade. Esse é um fato sério que não é discutido nos documentos disponíveis para consulta, merecendo ser mais bem estudado.

Outro ponto a ser analisado é com relação aos alunos do 4º ano que tiveram baixo desempenho (BD) no ano anterior. Pelos resultados expostos no Gráfico 3.7, observa-se que o desempenho deles sofre muita variação ao longo dos anos. Além disso, a média de desempenho desses alunos fica abaixo da média de desempenho dos demais do 4º ano e, o que é pior, abaixo da média dos alunos do 3º ano, demonstrando que a escola apresenta graves lacunas no processo de aprendizagem. O preocupante é que, segundo informação da Secretaria de Educação, õesses estudantes que tiveram um baixo desempenho no 3º ano passaram por um treinamento diferenciado no 4º ano (MINAS GERAIS, 2009, p. 58). Essa é uma informação inquietante, pois o que seria esse õtreinamento diferenciado? Além disso, parece que esse õtreinamento diferenciadoõ não teve efeito no desempenho dos alunos. Será que esses alunos não estariam sendo, de alguma forma, excluídos pelo fato apresentarem

maiores dificuldades de aprendizagem? São fatos que também merecem estudo aprofundado, porque o direito à educação pode lhes estar sendo negado.

*Programa de Avaliação da Aprendizagem Escolar*

Em 2003, a Secretaria de Educação implantou o Projeto Escolas-Referência, que tinha como objetivo reconstruir a excelência da escola pública mineira, dentro do programa do governador Aécio Neves (2002-2006), buscando ampliar e melhorar o ensino fundamental e a universalização do ensino médio. Nesse projeto foram selecionadas, em 2004, 220 grandes escolas em todo o Estado, entre as que tradicionalmente ofereciam um ensino de qualidade reconhecido pela comunidade a que pertenciam, a partir das quais se esperava irradiar mudanças para todo o sistema. Em 2005, acrescentou-se uma nova escola ao projeto e em 2006 mais duas, formando um grupo de 223 escolas no total. Essas escolas recebiam uma série de estímulos como: programas de formação continuada de professoras, reforma e ampliação dos prédios escolares, programa de formação de gestores por meio do Progestão e financiamento diferenciado para a execução de projetos. Para expandir suas boas práticas, cada uma escolhia uma escola associada com a qual colaboraria para seu desenvolvimento. Paralelamente ao projeto, estruturou-se uma nova base curricular para os anos finais do ensino fundamental e para o ensino médio, que veio a dar origem ao Currículo Básico Comum (CBC). Além do CBC, foi desenvolvida uma forma de acompanhar o desempenho da escola e dos alunos, dando origem ao Programa de Avaliação da Aprendizagem Escolar (PAAE).

O PAAE é um sistema informatizado gerenciado pela Secretaria de Educação e desenvolvido e mantido pelo Instituto Avaliar, que o administra e constrói itens para alimentar um banco de dados, base para elaborar provas capazes de avaliar o desempenho dos alunos. O objetivo inicial do PAAE era o de fornecer dados diagnósticos sobre a aprendizagem para subsidiar o planejamento do ensino e as necessárias intervenções pedagógicas. Foi testado entre 2006 e 2007 nas Escolas-Referência e em 2008 já atingia todas as escolas de ensino médio. Em 2013, o PAAE também atinge os anos finais do ensino fundamental.

Atualmente, o sistema gera provas a serem impressas, mas também estão sendo testadas aplicações *on-line* desde 2010. Dessas provas, duas são obrigatórias: a que compõe a avaliação diagnóstica, testando o conhecimento prévio dos alunos, ou seja, aqueles adquiridos no ano escolar anterior, aplicada entre fevereiro e março, e a que compõe a avaliação anual,

que testa a aprendizagem de todo o ano letivo, aplicada entre novembro e dezembro. Além das duas, a critério da escola e da professora, podem ser realizadas outras provas entre abril e setembro, podendo seus resultados ser utilizados para compor a nota do aluno.

Por meio do sistema, é possível montar provas para todas as disciplinas que integram o CBC: Biologia, Física, Química, Geografia, História, Língua Estrangeira, Língua Portuguesa, Matemática, Filosofia, Sociologia, Artes (por eixo ou para todos os eixos) e Educação Física. Segundo dados da Secretaria, o banco contém mais de 60 mil itens de todas as disciplinas do ensino fundamental e do ensino médio. O sistema gera relatórios por turma, por disciplina, por tópicos ou habilidades e a evolução obtida da 1ª para a 2ª prova. No âmbito da escola, esses resultados são disponibilizados para a professora e para o gestor e fornecem dados sobre o desempenho das turmas, das disciplinas e seus tópicos, da escola em todas as disciplinas, da Superintendência Regional de Ensino a que a escola está ligada e de toda a rede estadual. Os dados também são disponibilizados para as Superintendências Regionais de Ensino e para a Secretaria de Educação.

Existem etapas que devem ser percorridas desde a geração da prova até a obtenção de seu gabarito. Primeiramente, a professora, com sua senha pessoal, entra no sistema e seleciona qual prova quer montar. O sistema automaticamente seleciona os itens de múltipla escolha que compõem a prova e esta é gerada para ser impressa. O gabarito da prova só é disponibilizado depois de a professora ter resolvido a prova, inserindo suas respostas no sistema, e de inserir as respostas de todos seus alunos. Os itens que a professora não acertou são disponibilizados com as justificativas de resposta para cada alternativa, de maneira que a professora, segundo a Secretaria, possa aprimorar seus conhecimentos ou, em caso de discordância, sugerir, através do próprio sistema, a mudança de gabarito.

Com o objetivo declarado de contribuir para a melhoria do aprendizado no ensino médio, por meio desse sistema a Secretaria tem, de fato, o controle sobre o que está sendo ensinado em sala de aula, sobre como o processo educacional deve ser avaliado e sobre o desempenho de cada uma das professoras da rede. Apesar de o sistema gerar um grande conjunto de dados, não se encontrou nenhum estudo da Secretaria de Educação sobre os resultados alcançados pelo PAAE nem como essas informações estão sendo utilizadas. Assim, parece que o único objetivo do teste é controlar os conhecimentos das professoras e aquilo que deve estar sendo trabalhado em sala de aula. Um processo de regulação mais sutil que a regulação institucional, mas com um poder de controle muito maior. Mesmo assim, é um controle simbólico, com o objetivo de aumentar a pressão sobre o sistema, típico dos mecanismos de regulação, na linha do pensamento de Maroy (2011a), já que pouco se sabe

sobre os usos que são feitos tanto dos resultados alcançados pelas professoras como pelos alunos.

Seria importante investigar quais os efeitos do PAEE sobre o currículo, o ensino e o trabalho da professora, algo que aparentemente a Secretaria não fez até agora.

*As aplicações dos diferentes tipos de testes que compõem o Simave*

No Quadro 3.8, apresenta-se, ano a ano, como cada aplicação do Simave se comportou. Analisando-o, percebe-se que mediante seus diferentes instrumentos, o Simave testa os alunos do 2º ao 5º ano do ensino fundamental e com a expansão do PAEE testará também todos os alunos dos anos finais do ensino fundamental e do ensino médio, sendo que os estudantes do 9º ano do ensino fundamental e da 3ª série do ensino médio são testados duas vezes, pelo PAEE e pelo Proeb. E mais, são testados todos os anos. Diante desse cenário, alguns questionamentos merecem ser feitos: O que a Secretaria tem feito com essa quantidade impressionante de dados? Que modificações ela está induzindo no processo educacional, de forma a garantir uma aprendizagem significativa? Que apoio está dando à equipe escolar no sentido de apoiá-la em suas deficiências e estimulá-la a desenvolver novas formas de ensinar? Ao que parece, o objetivo principal é conseguir que as escolas apresentem para o corpo escolar desempenhos crescentes a partir de conceitos abstratos, relacionados à níveis de proficiência que devem ser alcançados, típico do ato de regular, como definiu Barroso (2005). Assim, a escola de qualidade não é onde a aprendizagem ocorre, onde a cooperação e a solidariedade se forjam, mas sim aquela que, independentemente dos processos utilizados, atinge o desempenho determinado. Os resultados desse processo podem ser danosos ao espírito de uma sociedade democrática, papel em que a escola sempre ocupou um papel de destaque.

**Quadro 3.8 – Síntese do Simave entre 2000 e 2011: anos escolares, áreas de conhecimento e número de alunos testados**

Teste	Informações	Ano												
		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	
Proalfa	Anos escolares testados	-	-	-	-	-	2º EF	2º e 3º EF	2º; 3º e 4º EF e BD <sup>(a)</sup>	2º; 3º e 4º EF e BD <sup>(a)</sup>	2º; 3º e 4º EF e BD <sup>(a)</sup>	2º; 3º e 4º EF e BD <sup>(a)</sup>	2º; 3º e 4º EF e BD <sup>(a)</sup>	
	Disciplinas	-	-	-	-	-	1	1	1	1	1	1	1	
	Abrangência <sup>(b)</sup>	-	-	-	-	-	E; M	E; M	E; M	E; M	E; M	E; M	E; M	
	Alunos (mil)	Rede Estadual	-	-	-	-	-	11	119	158	160	173	156	143
		Rede Municipal	-	-	-	-	-	-	-	99	225	272	265	251
Proeb	Anos escolares testados	5º e 9º EF e 3ª EM	5º e 9º EF e 3ª EM	5º e 9º EF e 3ª EM	5º e 9º EF e 3ª EM	-	-	5º e 9º EF e 3ª EM	5º e 9º EF e 3ª EM	5º e 9º EF e 3ª EM	5º e 9º EF e 3ª EM	5º e 9º EF e 3ª EM	5º e 9º EF e 3ª EM	
	Disciplinas	1; 2	3; 4	1	2	-	-	1; 2	1; 2	1; 2	1; 2	1; 2	1; 2	
	Abrangência <sup>(b)</sup>	E; M	E; M	E; M	E; M	-	-	E; M	E; M	E; M	E; M	E; M	E; M	
	Alunos (mil)	Rede Estadual	487	454	467	-	-	-	459	448	445	443	439	423
		Rede Municipal	5	29	77	-	-	-	184	225	218	232	237	238
PAAE	Ano escolar testado	-	-	-	-	-	1ª EM	1ª EM	1ª EM	1ª EM	1ª EM	1ª EM	1ª EM	
	Disciplinas	-	-	-	-	-	2; 3; 4; 5	2; 3; 4; 5	2; 3; 4; 5	2; 3; 4; 5	2; 3; 4; 5	2; 3; 4; 5	2; 3; 4; 5	
	Abrangência <sup>(b)</sup>	-	-	-	-	-	E	E	E	E	E	E	E	
	Alunos (mil)	-	-	-	-	-	-	-	-	217	226	221	251	
	Escolas participantes	-	-	-	-	-	72 <sup>(c)</sup>	223 <sup>(c)</sup>	223 <sup>(c)</sup>	1.900	2.100	2.100 <sup>(d)</sup>	2.150 <sup>(d)</sup>	
	Itens utilizados	-	-	-	-	-	-	3.169	9.468	10.014	11.682	15.008	20.632	

Fonte: MINAS GERAIS, 2011a, com elaboração do autor.

Notas: (a) ó Alunos com baixo desempenho em Língua Portuguesa no ano anterior e que se encontram matriculados no 5º ano.

(b) ó Abrangência refere-se à rede que recebe o programa.

(c) ó Disponível para as Escolas-Referência.

(d) ó Projeto piloto de testes *online*.

1 ó Língua Portuguesa; 2 ó Matemática; 3 ó Ciências Humanas; 4 ó Ciências da Natureza; 5 ó Linguagens.

E ó Rede Estadual; M ó Rede Municipal; EF ó ensino fundamental; EM ó ensino médio.

Entre as diferentes ações e programas desenvolvidos pela Secretaria, três foram escolhidos por serem aqueles que se relacionam diretamente com os resultados dos testes avaliativos: o Acordo de Resultados, o Programa de Intervenção Pedagógica (PIP) e a fixação de placas nas escolas com o resultado do Ideb 2011.

### **Acordo de Resultados**

O primeiro governo Aécio Neves (2003-2007) começou prometendo um choque de gestão, com o objetivo de alavancar o desenvolvimento de Minas Gerais. Para que isso ocorresse, foi implantada na administração pública a metodologia de gestão por resultados, denominada Gestão Estratégica dos Recursos e Ações do Estado (Geraes). O ponto de partida foi a definição daqueles que seriam os Projetos Estruturadores, capazes de garantir o desenvolvimento do Estado. Durante o segundo governo Aécio Neves (2007-2010), houve um total de 57 Projetos Estruturadores, agrupados em 11 Áreas de Resultados, que deram forma ao Plano Mineiro de Desenvolvimento Integrado (PMDI), para o período 2007-2023. Todo o PMDI é baseado em 340 indicadores de desempenho, para acompanhar as metas de governo. O estabelecimento das metas e seu acompanhamento são realizados mediante contrato de gestão, com os órgãos da administração pública, conhecido como Acordo de Resultados.

*Educação de Qualidade* é uma das 11 Áreas de Resultados, com o objetivo de, entre outros, promover um salto de qualidade no ensino, orientado por padrões internacionais.<sup>123</sup>. Essa estratégia agrupa quatro Projetos Estruturadores, sendo que dois deles são de interesse desta pesquisa. O primeiro deles, o *Projeto Sistemas de Avaliação* trata do Proeb e do Proalfa, destacando que seus resultados serão a base para ordenar muitas das ações do governo no campo educacional.

O outro projeto recebe o nome de *Desempenho e Qualificação dos Professores* e prevê ações de qualificação profissional e de avaliação direcionada às professoras dos ensinos fundamental e médio. A primeira ação é desenvolvida principalmente pela Magistra, a Escola de Formação Profissional de Educadores de Minas Gerais, criada no começo de 2011, que oferece cursos, na sua maioria de 30 horas, na modalidade a distância, centrados principalmente no CBC. Também promove debates com especialistas de diversos temas,

---

<sup>123</sup> Informação disponível no site [www.geraes.mg.gov.br](http://www.geraes.mg.gov.br), acessado em 30/03/2012.

dentro do Programa Roda de Conversa, que, depois de gravados, são disponibilizados na Internet<sup>124</sup>. Pelas informações existentes, essa ação está longe de ser um programa estruturado de formação continuada de professoras, e os cursos de curta duração oferecidos e os debates disponibilizados parecem ter pouca relação com os resultados dos testes avaliativos mineiros. É como se os resultados obtidos nos testes não revelassem problemas que deveriam ser trabalhados juntamente com as professoras da rede estadual, demonstrando, assim, um descompasso entre as diversas ações desenvolvidas pela Secretaria. Quanto à avaliação das professoras, não existem informações sobre qual seria seu objetivo, nem como e quando seria realizada, mas parece que o desejo é aplicar um teste às professoras para medir suas habilidades e seus conhecimentos.

O Acordo de Resultados, pactuado entre o governo estadual e a Secretaria de Educação, estabelece o cumprimento de metas de execução quantitativas para os Projetos Estruturadores. As metas são monitoradas por meio dos indicadores de resultados e entre eles encontram-se o percentual de alunos classificados no nível recomendado definido pelo Proeb e pelo Proalfa e a proficiência média nesses testes. Além desses indicadores, utilizam-se também as taxas de distorção idade-série, de matrícula líquida e de abandono, calculadas pelo Inep. Depois de pactuar as metas com o chefe do governo estadual, a Secretaria define unilateralmente as metas de desempenho para cada escola da rede, sem que haja a indicação de um critério objetivo para tal. Portanto, todo o sistema funciona em função dos indicadores de desempenho nos testes e do fluxo escolar. A aprendizagem e o processo educacional não recebem atenção, talvez na esperança de que os testes de desempenho por si só sejam capazes de resolver todos os problemas, cabendo à Secretaria apenas apontar o caminho que as escolas devem seguir.

Com base no desempenho da Secretaria e das escolas, frente às metas acordadas, todos os servidores são candidatos a receber uma bonificação em dinheiro, que pode chegar a uma remuneração mensal, a ser paga no semestre seguinte ao da apuração dos resultados. Assim, no caso da Secretaria de Educação de Minas, o bônus embutido no Acordo de Resultados está atrelado ao desempenho dos alunos nos testes. Portanto, em Minas, a escola de qualidade é aquela em que as professoras têm seu esforço medido pelo desempenho dos alunos, recompensado com o pagamento de até uma remuneração mensal por ano.

Essa concepção pode até dar resultados durante certo tempo, mas tem vida curta. Isso porque as condições estruturais que afetam o desempenho dos alunos não estão sendo modificadas. Além disso, reforça-se a ideia de o incentivo estar contribuindo fortemente para

---

<sup>124</sup> Disponibilizado pelo canal ômagistraseemgö do You Tube, disponível em <http://www.youtube.com/playlist?list=PL40938325AB7067FA>

a redução do currículo, ao invés de garantir a aprendizagem e conseqüentemente o direito à educação.

### **Programa de Intervenção Pedagógica**

O Programa de Intervenção Pedagógica começa a ser desenvolvido em 2006, com o objetivo de atingir a meta de todas as crianças lerem e escreverem até os 8 anos de idade, sendo gradativamente expandido para agregar também o objetivo de implantar os Conteúdos Básicos Comuns nos anos finais do ensino fundamental e no ensino médio (MINAS GERAIS, 2010d). Nesse programa, estão envolvidas as equipes da Secretaria, das Superintendências Regionais e das escolas, com o objetivo de analisar os resultados das avaliações externas, produzidas pelo Proalfa e pelo Proeb, e das avaliações internas realizadas pelas escolas e, a partir daí, propor um Plano de Intervenção Pedagógica para aprimorar os resultados da escola.

A elaboração do plano e seu acompanhamento são realizados por profissionais da escola, designados como Especialistas em Educação Básica. O trabalho desses profissionais é descrito em detalhes em um guia que delinea passo a passo as atividades a serem desenvolvidas e a época em que cada uma deve ocorrer (MINAS GERAIS, s/d). O guia informa que o Plano é uma ação ligada ao projeto pedagógico da escola, devendo ser elaborado por toda escola, com a colaboração da comunidade escolar, e tem como foco a melhoria do desempenho dos alunos (MINAS GERAIS, s/d, p. 16). Para chegar até o Plano, o documento diz que algumas etapas devem ser seguidas, envolvendo a compreensão dos resultados das avaliações externas, o conhecimento das metas propostas e pactuadas pela escola e o planejamento de ações para melhorar a aprendizagem dos alunos, tendo por base a análise da realidade escolar e as informações dos *Boletins Pedagógicos*, publicações que acompanham a divulgação dos resultados dos testes, discutindo como a escola pode usá-los para melhorar o desempenho dos alunos. Apesar de o documento apontar para a melhoria da aprendizagem, o foco está claramente apontado para o desempenho dos alunos nos testes.

Elaborado o Plano, o documento recomenda que o especialista o acompanhe e sugere alguns passos para tal:

[...] Converse, discuta com o diretor, com os professores, sobre as metas pactuadas pela escola, Superintendência Regional de Ensino e Secretaria de Educação para os anos 2008, 2009 e 2010. Dê uma atenção especialíssima aos alunos avaliados pelo



Proalfa e Proeb, com desempenho abaixo do intermediário. Onde estão esses alunos? Em que turma, com quais professores? Hoje eles já estão alfabetizados? Estão recebendo algum atendimento diferenciado? Você tem a relação nominal de todos eles? Que alunos dos Anos Finais do Ensino Fundamental ainda não consolidaram o processo de Alfabetização? E do Ensino Médio? Como resolver esses casos? Identifique o problema, quantifique-o, estude com seu Diretor que tipo de atendimento será possível utilizar na Escola, como por exemplo: remanejamento temporário; participação dos alunos nas turmas do Projeto Tempo Integral; trabalho de leitura na biblioteca; atendimento em pequenos grupos; aulas de reforço e outros, a partir de sua criatividade e da sua equipe de professores (MINAS GERAIS, s/d, p. 22).

As atividades listadas são aquelas que normalmente deveriam ser desenvolvidas pela equipe pedagógica da escola. Portanto, a primeira vista, não fica claro o porquê desse novo profissional na escola. De início, antes da leitura das sugestões de atuação para esses profissionais, poder-se-ia considerar que sua presença na escola poderia ajudar na coordenação do trabalho pedagógico. A novidade é que agora esse trabalho tem como foco as avaliações educacionais e a melhoria dos indicadores. Sua atuação está focada nas áreas de conhecimento que fazem parte do teste e, além disso, seu trabalho é cuidar para que as metas estabelecidas pela Secretaria para a escola sejam cumpridas. Ao fim, pode-se argumentar que os Especialistas em Educação Básica atuam, na verdade, como fiscais e orientadores em direção ao cumprimento dessas metas. Seu trabalho, muito mais restrito do que aquele que se esperaria de um coordenador pedagógico, está a serviço não do aprimoramento da aprendizagem, apesar de ela ser uma preocupação citada no guia, mas sim do cumprimento de metas de desempenho. Em nenhum momento, o texto faz menção a alternativas de avaliação ou a necessária confrontação com outros aspectos da realidade escolar que podem estar impactando os resultados demonstrados pelos alunos, ou mesmo orientação para a realização de pesquisas que contribuam para a obtenção de dados adicionais que permitam compreender melhor a realidade da escola.

Essa preocupação que os especialistas devem ter com o cumprimento das metas fica mais evidente no trecho do documento que trata das atividades que o especialista deve desenvolver nos momentos que antecedem o dia da aplicação dos testes:

Transforme o dia da avaliação externa num momento especial para seus alunos, seus professores e sua Escola. Afinal, é com ela que saberemos se cumprimos ou não as

metas pactuadas, isto é, se cumprimos o nosso dever e compromisso de oferecer melhor ensino e maior aprendizagem a nossos alunos (MINAS GERAIS, s/d, p. 27).

O trecho, além de desnudar a função desse profissional, mostra que uma escola de qualidade e onde ocorre a aprendizagem é aquela que cumpre com as metas estabelecidas pela Secretaria de Educação.

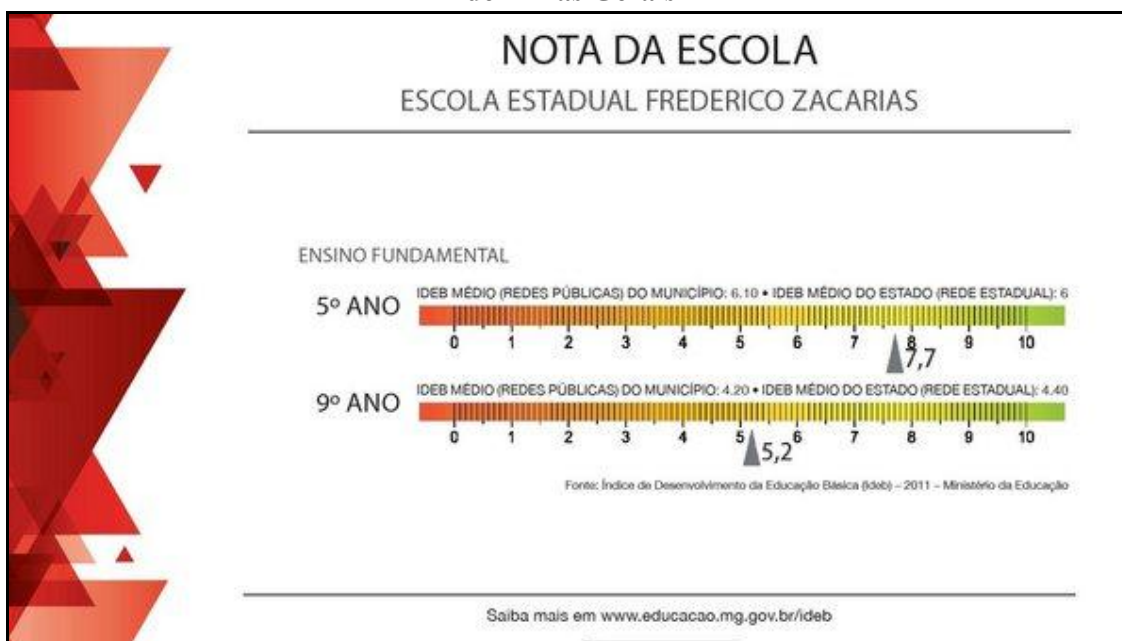
Portanto, fica evidente que a opção adotada pela Secretaria é de regular a escola, a partir da definição de metas a serem cumpridas e a indicação de um supervisor para garantir seu cumprimento.

#### **Divulgação dos resultados do Ideb 2011 nas escolas**

No fim de 2012, a Secretaria de Educação anunciou que, no início de 2013, seriam distribuídas a todas as escolas da rede estadual placas informativas com os resultados do Ideb de 2011 e do Proeb para o ensino médio. No caso do ensino médio, escolheu-se divulgar os resultados do Proeb, pois, para essa etapa da educação básica, o Saeb é aplicado apenas a uma amostra de escolas. Essa iniciativa teve a parceria do sistema Federação das Indústrias de Minas Gerais (Fiemg) e do Instituto Minas pela Paz, uma Organização da Sociedade Civil de Interesse Público (Oscip) mantida por diversas empresas com sede em Minas.

Segundo a Secretaria, essa placa deveria ser fixada, a critério da escola, em local visível para toda a comunidade. Um exemplo dessa placa, divulgado pela revista *Veja*, é mostrado no Quadro 3.9. Note-se que, nesse caso, pelo fato de a escola oferecer apenas o ensino fundamental, a placa divulga apenas os resultados do Ideb.

**Quadro 3.9 – Placa informativa contendo o Ideb de 2011 de uma escola da rede estadual de Minas Gerais**



Fonte: <<http://veja3.abril.com.br/assets/images/2012/12/112968/exemplo-de-placa-size-598.jpg?1354560206>>. Acesso em: 15 fev. 2013

Com relação ao projeto gráfico da placa, duas informações se destacam. Primeiro, o destaque dado à expressão *nota da escola*, claramente uma mensagem relacionada à classificação da escola dentro da rede estadual. Outro destaque é a gradação de cores da escala do Ideb, que começa com o vermelho e caminha até o verde, uma imagem destacando três regiões que claramente se identificam com os aspectos negativos, relacionados à cor vermelha; de alerta, relacionados ao amarelo; e positivos, relacionados ao verde.

Iniciativas semelhantes, relativas à divulgação pela escola de seus indicadores educacionais, fizeram parte da legislação mineira. Um exemplo disso é a Lei nº 11.036, de 14 de janeiro de 1993, que determina a obrigação de toda escola divulgar dados relevantes relativos a seu desempenho e afixá-los em sua Secretaria, em lugar de fácil acesso. Segundo a lei, os dados relevantes referem-se ao número de alunos matriculados e ao número de vagas por série; ao percentual de alunos aprovados, reprovados e que abandonaram a escola por série; e o percentual global de repetência e abandono escolar. No entanto, não se tem informações se a lei foi, ou não, de fato aplicada. Vinte anos depois, os indicadores de desempenho ficaram reduzidos apenas ao Ideb, e não foi necessária nenhuma lei para sua implementação pela Secretaria.

Assim, como se não bastassem as metas impostas à escola pela Secretaria dentro do Acordo de Resultados, a escola também é pressionada a expor publicamente seus resultados, a partir de um indicador elaborado pelo governo federal. Nenhuma das metas a que a escola foi submetida foi discutida, mas imposta. Além disso, como já foi discutido antes, classificar as escolas a partir do Ideb é um contrassenso, na medida em que o indicador captura apenas dois momentos muito específicos da vida escolar, relacionados ao desempenho dos alunos nos testes de Língua Portuguesa e Matemática e ao fluxo escolar. Sobre os outros aspectos da escola, nada é levado em consideração. Usam-se dados que foram desenvolvidos para captar determinado aspecto específico de uma realidade, para informar equivocadamente uma pretensa qualidade que não foi sequer discutida e muito menos medida.

Independentemente da importante discussão do que seria a qualidade de uma escola, discute-se aqui a divulgação de uma informação bastante limitada e focada em determinados aspectos, como se fosse abrangente o suficiente para captar toda a realidade. Com isso, é muito difícil escapar da armadilha de fazer com que toda a escola se mobilize no sentido de se preparar cada vez mais para os testes, uma vez que está bem claro que essa será a forma pela qual será avaliada. Além dos aspectos já discutidos com relação ao currículo, esse processo acaba trazendo inevitáveis tencionamentos entre as professoras que terão seu trabalho julgado pelos testes e aquelas que passam incólumes pelo processo, por não existirem testes aplicados à área de conhecimento em que atuam.

Com base no que foi exposto sobre o caso de Minas Gerais, é possível verificar que a trajetória do governo mineiro no processo de implantação do sistema de avaliação no Estado, começou como um verdadeiro processo avaliativo cooperativo, envolvendo os órgãos gestores, a equipe escolar e a família dos alunos. Havia a preocupação inicial de se discutir os testes antes e depois de sua aplicação, o que permitia o envolvimento daqueles que efetivamente faziam o processo educativo acontecer, pois era a escola, na figura das professoras, que analisava os resultados. O que a pesquisa encontrou é que mesmo a introdução da avaliação educacional em Minas Gerais tendo iniciado a partir de um acordo de empréstimo internacional para melhorar a gestão do sistema educacional, com claros objetivos de aumentar a performatividade das escolas, o processo coordenado pela Fundação Carlos Chagas buscou trabalhar o sentido de avaliação ligada ao desenvolvimento do currículo e à aprendizagem dos alunos, o verdadeiro significado de uma avaliação educacional. Portanto, as melhoras dos indicadores estariam conectadas às melhoras na aprendizagem. Além disso, nesse período inicial, buscou-se a cooperação entre os diferentes atores, envolvendo a elaboração dos instrumentos e sua aplicação, a análise de seus resultados

e sua divulgação, no sentido de contribuir com a aprendizagem dos envolvidos, garantindo-lhes autonomia para os passos seguintes.

O que se revelou também é que, apesar de a avaliação externa ser utilizada como um instrumento para garantir a regulação dentro do modelo de governança do Estado Avaliador, ela pode se constituir como um poderoso instrumento para favorecer a aprendizagem e contribuir para o desenvolvimento da autonomia dos agentes escolares e da cooperação entre si. No caso de Minas, como se instaurava um processo, naturalmente os resultados não seriam imediatos. Mas em tempos de regulação e como forma de demonstrar a eficiência dos governos, os resultados deveriam ser mais rápidos, ajustando-se aos tempos políticos. Dessa forma, a busca pela performatividade passou a ser uma obsessão constante, em que cada governo buscou aumentar cada vez mais a pressão sobre as escolas. Esse processo passa pela indicação de um Especialista em Educação Básica para atuar junto às escolas, com o principal atributo, a partir das orientações fornecidas pela própria Secretaria, de trabalhar para garantir o cumprimento das metas estipuladas unilateralmente pelo órgão gestor, culminando com a fixação, por livre iniciativa da Secretaria de Educação, das placas nas escolas com os resultados alcançados por elas.

### ***O Estado de São Paulo***

Existem diversos estudos sobre os testes avaliativos em São Paulo, destacando-se os estudos de SOUZA; OLIVEIRA, 2007; BROOKE; CUNHA; FALEIROS, 2011; BAUER, 2006; MACHADO, 2003; CARVALHO, 2008.

Nos tópicos a seguir, serão apresentados os primórdios da avaliação educacional paulista, com o Programa de Avaliação da Rede Estadual Paulista, passando pela apresentação das propostas do governo Mario Covas, que, pela primeira vez, delineiam o *accountability* e as políticas *high stakes* que viriam a se disseminar com força nos anos seguintes. Apresentam-se também as diretrizes do Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (Saresp) e suas modificações ocorridas com o passar do tempo. São apresentados, também, os resultados obtidos até 2011 e discutidas as políticas implementadas a partir deles.

É importante destacar que muito auxiliou a pesquisa a eficiência dos mecanismos de busca e a grande quantidade de material para consulta nos *sites* da Secretaria de Educação e do *Diário Oficial* do Estado de São Paulo.<sup>125</sup>

### **O Programa de Avaliação da Rede Estadual Paulista**

Diferentemente do que ocorreu em Minas Gerais, a Constituição do Estado de São Paulo, em seu artigo 239, § 3º, trata da avaliação apenas quando se refere às escolas privadas, repetindo a Constituição Federal, e às universidades públicas estaduais, neste caso apontada como uma forma de participação da sociedade na gestão dos recursos públicos. Nenhum preceito, fora do indicado em relação à escola privada, trata da avaliação da educação básica.

Em São Paulo, o movimento em direção à avaliação educacional teve inicialmente o objetivo avaliar programas e projetos em execução pelo governo estadual, e seus primeiros estudos datam do início dos anos 1990. Em 1992, durante o governo de Luis Carlos Fleury Filho (1991-1995), foi criado o Programa de Avaliação Educacional da Rede Estadual para avaliar o Projeto Escola-Padrão e sua relação com as melhorias no desempenho dos alunos. O projeto começou em 1991 e enunciavam que seus objetivos eram descentralizar as decisões e dar mais autonomia às escolas. Em 1992, foram avaliadas 306 escolas, aplicando-se testes de desempenho nas áreas de Língua Portuguesa (incluindo Redação), Matemática, Ciências, História e Geografia a 27.609 alunos de diferentes anos escolares. Os resultados evidenciaram que os alunos:

- apresentavam melhor desempenho nas questões de conhecimento matemático cuja solução se limitava ao emprego de técnicas operatórias, e grandes dificuldades naquelas que exigiam análise e interpretação de situações e/ou resultados;
- interpretavam as questões relativas aos processos e fenômenos naturais frequentemente com elementos extraídos do senso comum, e raramente a partir de conhecimentos científicos;
- apresentavam dificuldades nas questões que exigiam interpretação de textos;

---

<sup>125</sup> Secretaria de educação de São Paulo: [www.educacao.sp.gov.br](http://www.educacao.sp.gov.br); Diário Oficial de São Paulo: [www.imprensaoficial.com.br](http://www.imprensaoficial.com.br)

- demonstravam, na produção de textos, dificuldades quanto aos elementos estruturais da narrativa e em relação aos procedimentos de organização do texto (SEE-SP, 1996, p. 121).

Muitos dos pontos abordados são recorrentes nos resultados que vêm sendo obtidos nos diferentes testes aplicados depois dele. Portanto, os destaques negativos dados aos problemas educacionais pelas matérias jornalísticas nos anos 2000, dez anos depois, não deveriam ser novidades para os gestores da área educacional.

Como um produto importante do Programa, destaca-se a utilização que as professoras fizeram dos seus resultados para seu planejamento escolar (GATTI, 1994). A autora afirma que como as professoras tiveram oportunidade de avaliar o teste aplicado, pois receberam, depois de sua aplicação, uma análise orientada sobre os erros cometidos pelos alunos, puderam refletir sobre sua prática.

Dois anos depois, em 1994, o Programa de Avaliação foi estendido a todas as escolas da rede estadual, mas de forma amostral. Para tanto, foi selecionada uma amostra de 818 escolas e testados 152.279 alunos matriculados nas 4ª séries, nas áreas de Língua Portuguesa, incluindo Redação e Matemática, e nas 8ª séries; além dessas áreas, foram testadas Ciências, História e Geografia. Os resultados obtidos pelos alunos da 4ª série foram os seguintes:

- apresentavam um bom rendimento nas atividades de linguagem;
- demonstravam grandes dificuldades no reconhecimento da função e do uso dos elementos que estruturam as frases e os textos;
- mostravam mais facilidade para resolver questões que envolviam operações com números naturais, e mais dificuldade quando se tratava de problemas com números racionais (na forma fracionária ou decimal) (SEE-SP, 1996, p. 122).

Para os alunos da 8ª série, os resultados mostraram que os alunos:

- apresentavam dificuldades na resolução de questões que demandavam a compreensão de enunciados, a percepção de conceitos envolvidos, a seleção de informações e a análise de situações-problema;
- revelaram dificuldades para ler e interpretar tabelas e gráficos;
- apresentavam inadequações na interpretação textual e na identificação e análise dos elementos frasais, de acordo com a função que desempenham dentro das frases (SEE-SP, 1996, p. 122).

Pelos excertos, percebe-se que os problemas citados não diferiram muito dos resultados descritos em 1992, mas, em alguns casos, foram destacados pontos específicos, que traziam elementos que permitiriam trabalhar melhor o currículo escolar, quando, por exemplo, foram identificados problemas em determinados conteúdos de Matemática. Além disso, destaca-se pelo menos uma interpretação bastante genérica dos resultados quando se refere a ãapresentavam um bom rendimento nas atividades de linguagemö. Essa informação, isoladamente, pouco contribuía para compreender o que estava acontecendo com o ensino.

Também entre 1992 e 1994, dentro de um projeto financiado pelo Banco Mundial, conhecido por *Inovações* (Projeto Inovações na Educação Básica em São Paulo), foram aplicados testes de Língua Portuguesa e Matemática aos estudantes para avaliar o impacto das políticas educacionais estaduais (BITAR; FARES *et al.*, 1998; MACHADO, 2003). As políticas em questão eram o Ciclo Básico de Alfabetização, adotado em 1984, em que as crianças faziam as duas primeiras séries dos anos iniciais do ensino fundamental sem reprovação e a Jornada Única, adotada em 1988, em que a professora do Ciclo Básico regeria apenas uma classe e trabalharia por seis horas diárias. Buscava-se, então, saber se o:

[...] Ciclo Básico/Jornada Única causam algum impacto, algum diferencial no desempenho e no desenvolvimento das crianças? Há diferenciais no desenvolvimento das habilidades básicas que devem ser adquiridas? Altera-se a relação aprovação/reprovação/evasão? (GATTI, 1994, p. 75).

Portanto, muito mais do que procurar respostas sobre o que as crianças estavam aprendendo, buscavam-se respostas sobre as habilidades que estavam sendo desenvolvidas e seus efeitos sobre o fluxo escolar, verificando em que medida as políticas implementadas estavam superando os problemas existentes.

Esse estudo foi o primeiro com caráter longitudinal realizado no Brasil, acompanhando o mesmo segmento de crianças durante três anos, com o objetivo de ãanalisar mudanças de aprendizagem e de características cognitivas ao longo do tempoö (GATTI, 1996, p. 26). Por seu caráter inovador, deveria ter sido mais bem estudado e dado origem a muitas pesquisas.



Em 1995, durante a gestão do governador Mario Covas (1995-1999), a Secretaria de Educação, comandada, na época, por Rose Neubauer, divulga através do Comunicado SE, de 22 de março de 1995, as diretrizes da Secretaria durante o período de janeiro a dezembro desse ano. Esse Comunicado revelou uma prática salutar de anunciar publicamente as políticas que seriam implantadas pela Secretaria. Ações como essa são raras no campo educacional.

Depois de fazer um diagnóstico da educação básica paulista, o Comunicado, ao destacar a ineficácia e a ineficiência do sistema estadual, estabelece uma série de diretrizes para superá-las. Uma delas refere-se a mudanças no padrão de gestão com a instituição de mecanismos de avaliação, a racionalização do fluxo escolar e o aumento da autonomia administrativa, financeira e pedagógica da escola. Com relação à autonomia, o documento diz que o Estado de São Paulo deveria acompanhar as propostas implementadas com sucesso em outros Estados e destaca que

[...] a autonomia tem como contrapartida a responsabilidade e o compromisso. A administração, buscando a integração das ações escolares, deverá estabelecer um sistema criterioso de avaliação dos resultados da aprendizagem dos alunos e cuidar para que as escolas respondam por eles (SÃO PAULO, 1995, p. 16).

Assim, a avaliação dos resultados dos alunos surge como um mecanismo de regulação da Secretaria, em troca da autonomia dada à escola, sem ser esclarecido, no Comunicado, qual autonomia seria essa.

Ainda segundo o Comunicado, por meio da avaliação, seriam fornecidos dados para atuar na superação das desigualdades existentes nas escolas paulistas, possibilitando ao Estado cumprir seu papel equalizador. Além disso,

[...] os resultados do desempenho das escolas deverão ser amplamente divulgados, de forma que tanto a equipe escolar como a comunidade usuária seja capaz de identificar a posição da sua escola no conjunto das escolas de sua Delegacia [de Ensino], de seu bairro e de seu município. Isto possibilitará à escola a busca de formas diversificadas de atuação, com o objetivo de implementar a melhoria dos resultados escolares. Por outro lado, permitirá também à população acesso às informações, de modo que possa fiscalizar, participar e cobrar a qualidade do serviço que lhe deve ser prestado (SÃO PAULO, 1995, p. 16).

Parte-se do princípio que o maior problema da educação paulista estaria ligado a uma gestão ineficiente e que, para superá-la, a melhor estratégia seria publicizar os resultados das escolas, na esperança de que essa exposição levasse a uma melhoria dos resultados escolares, impulsionada pela fiscalização da população. É o *accountability* buscando a prestação de contas e a responsabilização da escola sendo delineado e utilizado como a base de regulação dentro do Estado Avaliador.

O comunicado destaca também a transparência e a participação popular como imprescindíveis para a maior produtividade dos serviços públicos, sem que seja explicitado como elas ocorreriam. Reconhece, ainda, a necessidade de correção salarial para as professoras, sinalizando com um piso inicial de cinco salários mínimos. No entanto, essa correção deveria estar atrelada à melhoria do desempenho delas, fazendo-se necessário estabelecer mecanismos de avaliação e de promoção pelo trabalho efetivamente realizado pelas professoras.

Pela análise do Comunicado, percebe-se uma mudança de paradigma nas políticas da Secretaria, pois, anteriormente, centralizava-se a gestão e descentralizava-se o projeto pedagógico e sua avaliação. A proposta, a partir de então, passa a centralizar os parâmetros de desempenho e a avaliação, pontos considerados chaves para a gestão do sistema, descentralizando-se a gestão da escola (BAUER, 2006).

As ideias para a utilização do *accountability* e para a formulação de políticas *high stakes* estavam sendo lançadas através do Comunicado SE, de 22 de março de 1995. No entanto, elas não foram prontamente colocadas em prática nesse governo, mas foram sendo aprofundadas nos próximos governos, com alguns recuos em determinados momentos e avanços em outros.

No ano seguinte, em 1996, a Secretaria divulga um estudo intitulado *Saresp – Documento de Implantação*, no qual são descritos os objetivos do Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo, quais sejam:

ó desenvolver um sistema de avaliação de desempenho dos alunos dos ensinos fundamental e médio do Estado de São Paulo, que subsidie a Secretaria da Educação nas tomadas de decisão quanto à política educacional do Estado;

ó verificar o desempenho dos alunos nas séries dos ensinos fundamental e médio, bem como nos diferentes componentes curriculares, de modo a fornecer ao sistema de ensino, às equipes técnico-pedagógicas das Delegacias de Ensino e às Unidades Escolares informações que subsidiem:

- a capacitação dos recursos humanos do magistério;

- a reorientação da proposta pedagógica desses níveis de ensino, de modo a aprimorá-la;
- ó a viabilização da articulação dos resultados da avaliação com o planejamento escolar, a capacitação e o estabelecimento de metas para o projeto de cada escola, em especial a correção do fluxo escolar (SEE-SP, 1996, p. 124-125).

Nesse trecho, é importante destacar a ênfase dada à utilização dos resultados, para subsidiar a capacitação dos recursos humanos e a reorientação da proposta pedagógica das escolas, e ao desejo de viabilizar o estabelecimento de metas para cada uma delas, voltadas especificamente para o fluxo escolar. Com relação ao primeiro ponto, constata-se uma evidente tentativa para construir caminhos que permitissem melhorar o desempenho dos alunos, fornecendo instrumental adequado às professoras a partir de sua capacitação. No ponto seguinte, percebe-se, ainda que timidamente, uma inclinação para o *accountability*, mas ainda muito distante das políticas que vieram a ser implementadas anos depois, pois o foco principal era desenvolver ações para melhorar o problema do fluxo escolar na rede estadual paulista (BITAR, FARES *et al.*, 1998). De qualquer forma, os objetivos se adequavam a uma avaliação do sistema educacional e não pretendiam, pelo menos formalmente, identificar ou ranquear escolas.

Ainda em 1996, a Secretaria edita a Resolução nº 27, que cria o Saresp, justificando sua criação pela necessidade de uma política de avaliação do rendimento escolar articulada com o Saeb, para recuperar o padrão de qualidade do ensino paulista e a importância de pautar a tomada de decisões, seja pelas Diretorias de Ensino, seja pelas escolas, a partir de resultados objetivos cientificamente apurados (SÃO PAULO, 1996b). Seriam abrangidas pelo Saresp todas as escolas da rede estadual e das redes municipais e privadas que aderissem à proposta. De forma gradativa e contínua, envolveria todas as séries do ensino fundamental nas áreas de Língua Portuguesa, incluindo Redação, Matemática, Ciências, História e Geografia e todas as séries do ensino médio nas áreas de Língua Portuguesa, incluindo Redação, Matemática, História, Geografia, Química, Física e Biologia.

Em documento de 1998, outra intenção é acrescentada ao Saresp: a de gerar uma cultura de avaliação que permitisse agilizar a tomada de decisões com relação à melhoria do ensino (SÃO PAULO, 1998). A expressão "gerar uma cultura de avaliação", que também foi utilizada na implantação dos testes avaliativos em Minas Gerais, é muito comum em diversos documentos que tratam de sistemas de avaliação educacional. Se, no caso de Minas Gerais, o objetivo dessa cultura de avaliação estava claramente expresso na expectativa de que o

contato com as práticas e os instrumentos de avaliação externa influenciasse a avaliação no interior da escola, no caso de São Paulo soa como uma proposta vaga, pois não relaciona a cultura de avaliação com o aprimoramento do processo educacional, apenas se refere a uma genérica melhoria do ensino. Destaque-se que, nos últimos tempos, essa cultura de avaliação tem-se identificado fortemente com o sentido de expandir ainda mais os testes avaliativos com o objetivo, próprio do Estado Avaliador, de regular a escola mediante seus resultados (MAROY, 2011). De toda a forma, a citada expressão transformou-se em um chavão recorrente.

Com relação à análise dos resultados, Souza e Oliveira (2007) afirmam que até 1998

[...] cada unidade escolar tinha uma equipe responsável por elaborar um relatório descritivo, inicialmente, e depois propositivo, nas aferições de 1997 e 1998. Essa demanda estava vinculada a um propósito definido: fazer com que as equipes escolares e as Diretorias de Ensino se debruçassem sobre os dados, os analisassem e fizessem proposições acerca da superação dos resultados obtidos. As Diretorias Regionais elaboravam uma síntese dos relatórios de suas escolas e enviavam um documento de trabalho à Coordenadoria de Estudos e Normas Pedagógicas (Cenp), que, por sua vez, elaborava a forma de devolução de cada relatório. Além disso, tais relatórios serviam para que a equipe técnica da Cenp diagnosticasse as dificuldades das equipes das Diretorias e das escolas na manipulação e entendimento dos dados e fizesse propostas para superar essas dificuldades. Aos poucos, a construção desses relatórios foi sendo substituída pela construção de gráficos de resultados. Atualmente [2007], não há uma diretriz da Secretaria da Educação para o trabalho com os relatórios, sendo que cada Diretoria Regional propõe uma forma diferenciada para que as equipes escolares se debruçam sobre os dados (p. 175).

Depreende-se dessa transcrição que, inicialmente, havia um envolvimento direto da escola com o tratamento dos dados gerados pela avaliação, uma forma, talvez, de permitir que a equipe escolar se apropriasse dos resultados. Depreende-se, também, que, com o passar do tempo, aquelas informações, que poderiam servir para orientar o trabalho da escola, deixam de ser prioridade, e exige-se das escolas que expressem seus resultados em gráficos, dando destaque única e exclusivamente ao resultado, independentemente do processo ocorrido.

Para melhor visualização de como vem ocorrendo a aplicação do Saresp, quais as áreas de conhecimento testadas, os anos escolares em que essa testagem ocorreu e o número de alunos testados entre 1996 e 2011, montou-se o Quadro 3.10.

**Quadro 3.10 – Síntese do Saresp entre 1996 e 2011: anos escolares<sup>126</sup>, áreas de conhecimento e número de alunos testados**

Ano	Anos do EF									Anos do EM			Nº de alunos
	1º	2º	3º	4º	5º	6º	7º	8º	9º	1ª	2ª	3ª	
1996 <sup>a</sup>				1,2,3				1,2,3 ,4,5, 6					1.078.592
1997 <sup>a</sup>					1,2,3				1,2,3 ,4, 5,6				946.170
1998 <sup>a</sup>						1,2,3				1,2,3 ,4,5, 6			947.036
1999	Não houve aplicação do Saresp												
2000 <sup>b</sup>						1,2,3 ,4		1,2,3 ,4				1,2,3 ,8	1.320.336
2001					1,2				1,2				872.086
2002 <sup>c</sup>					1,2				1,2				297.258
2003 <sup>d</sup>		1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2	4.274.404
2004		1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2	4.495.865
2005		1,2,3	1,2,3	1,2,3	1,2,3	1,2,3	1,2,3	1,2,3	1,2,3	1,2,3	1,2,3	1,2,3	4.786.410
2006	Não houve aplicação do Saresp												
2007		1,2,3	1,2,3		1,2,3		1,2,3		1,2,3			1,2,3	1.858.077
2008			1,2,3		1,2,3		1,2,3 ,4		1,2,3 ,4			1,2,3 ,7	1.836.435
2009 <sup>e</sup>			1,2,3		1,2,3		1,2,3 ,5,6		1,2,3 ,5,6			1,2,3 ,5,6	2.260.024
2010			1,2,3		1,2,3		1,2,3 ,4		1,2,3 ,4			1,2,3 ,7	1.889.088
2011			1,2,3		1,2,3 ,5,6		1,2,3 ,5,6		1,2,3 ,5,6			1,2,3 ,5,6	1.889.088

Fontes: SOUZA; OLIVEIRA, 2007; SÃO PAULO, 2000, 2004, 2008, 2009, 2010, 2011, 2012, com elaboração do autor.

Notas: (a) Entre 1996 e 1998, o teste foi aplicado no início do ano escolar seguinte ao de interesse (avaliação de entrada); nos anos seguintes, o teste foi aplicado no final do ano escolar (avaliação de saída); foi aplicado de forma censitária, mas metade da turma respondeu itens de Matemática e a outra de Leitura. Nos anos seguintes, todos os alunos responderam aos testes de todas as áreas testadas.

(b) Pesquisa de Carvalho, (2008) mostra que também foram testadas as áreas de Geografia e História, informação que não condiz com documento da Secretaria (SÃO PAULO, 2000).

(c) Em 2002, foi a única vez que os testes foram aplicados a uma amostra de alunos. Nos outros anos, a aplicação foi censitária.

(d) Em 2003, participaram além dos alunos da rede regular, os das classes de Aceleração, Recuperação de Ciclo, classes Mistas, Magistério, Cefam e classes de Flexibilização.

(e) em 2009, foram aplicados testes em braile e ampliados; exceto para o 2º ano, aplicaram-se a uma amostra de alunos cinco questões abertas de Matemática.

EF ó ensino fundamental; EM ó ensino médio.

1 ó Leitura; 2 ó Redação; 3 ó Matemática; 4 ó Ciências; 5 ó Geografia; 6 ó História; 7 ó Ciências da Natureza; 8 ó Biologia.

Observa-se pelo Quadro 3.10 que existiu grande variação no desenho da aplicação dos testes, em termos de áreas de conhecimento e anos escolares testados. No entanto, a aplicação dos testes continuou censitária em todo o período, com exceção do ano 2002, em que foi amostral. Não foi possível detectar as razões de o teste ter sido amostral em 2002, nem uma análise dos benefícios ou não dessa decisão. Uma experiência interessante ocorreu entre 1996

<sup>126</sup> Utiliza-se a divisão do ensino fundamental de nove anos instituída pela Lei 11.274/2006. É importante destacar que o ensino fundamental de nove anos só começou a ser implantado na rede estadual paulista em 2009. Portanto, para os anos anteriores foi feita a conversão.

e 1998, quando se pretendia acompanhar o desenvolvimento entre séries, mas nenhum estudo da época explica as razões para esse desenho ter sido abandonado. As únicas áreas que foram testadas em todos os anos foram Leitura e Redação.<sup>127</sup> Logo a seguir, como a segunda área mais testada, está Matemática.

Em 1999, primeiro ano da segunda gestão de Mario Covas (1999-2001), não houve aplicação do Saresp, não sendo possível apurar, pela documentação pesquisada, as razões de sua não realização.

A aplicação do Saresp é retomada em 2000, com testes nos 6º e 8º anos do ensino fundamental. Mas por que testar os alunos um ano depois do início dos anos finais do ensino fundamental e um ano antes de seu término? Não existe nenhuma explicação sobre as razões dessa escolha, além de ela se diferenciar de qualquer outro modelo utilizado durante todo o período da avaliação educacional no Brasil. Assim, parece que havia a decisão de testar os alunos, retomando a tradição do Saresp, mas, por algum motivo, resolveu-se inovar sem qualquer estudo prévio, algo bem ao gosto de muitas das mudanças de governo no Brasil em todos os níveis. O que espanta é que uma alteração como essa, que mexe com o dia a dia das escolas e de seus integrantes, tenha sido feita sem o mínimo cuidado. Iniciativas como essa só reforçam as desconfianças que se tem sobre a avaliação educacional.

Entre 2001 e 2002, voltam a ser aplicados testes no final dos anos iniciais e finais, seguindo a mesma metodologia do Saeb, mas testou-se apenas Leitura e Redação, e isso tem uma explicação, ainda que não oficial. Em 2001, a Secretaria, novamente comandada por Rose Neubauer, mediante a Resolução SE nº 124/2001, regulamentada pela Instrução Cenp de 13/11/2001, determina que todos os alunos dos 5º e 9º anos do ensino fundamental estariam obrigados a fazer um teste de Leitura e Redação e que o resultado obtido seria usado para decidir sobre a progressão, ou não, para os anos escolares seguintes (SÃO PAULO, 2001). O teste utilizado foi o Saresp. Assim, utilizou-se o Saresp, que tinha por objetivo acompanhar a evolução da educação, para atender a um objetivo totalmente diferente daquele para o qual foi desenhado, ou seja, como exame de promoção para a etapa seguinte da educação básica. Com esse procedimento, é clara a perda de autonomia das escolas, pois retira delas a decisão sobre a aprovação ou não dos alunos, fato que gerou uma resistência muito grande entre os profissionais da rede, diminuindo mais ainda sua pouca credibilidade no Saresp (SOUZA; OLIVEIRA, 2007).

---

<sup>127</sup> Apesar da literatura oficial referir-se a testes de Língua Portuguesa, optou-se por separar a área em Leitura, o mesmo que é avaliado pelo Saeb, e Redação, em que os alunos devem desenvolver um texto a partir de um estímulo dado.

Durante o período 2003-2005, no segundo governo de Geraldo Alckmin (2003-2006), sob o comando de Gabriel Chalita na Secretaria de Educação, a testagem atingiu todos os alunos da rede. Nesse período, foram testados, pela primeira vez, os alunos dos 2º e 3º anos do ensino fundamental. O modelo dos itens utilizados envolvia itens de múltipla escolha e de resposta construída, sendo que para alguns deles havia instrução de leitura pela professora. Em 2003 e 2004, testaram-se as áreas de Leitura e Redação e, em 2005, foi acrescentada a área de Leitura. É possível imaginar o esforço dispendido para dar conta do desafio de testar quase cinco milhões de alunos anualmente e a quantidade de recursos gastos nessa tarefa. Segundo Carvalho (2008), o custo de aplicação variou entre sete e dez milhões de reais somente em 2004. Apesar disso e da grande quantidade de dados que devem ter sido gerados, nenhum único documento da Secretaria foi localizado ou mesmo foram localizadas pesquisas com informações sobre o Saresp aplicados entre os anos de 2003 a 2005.

Em 2006, no governo Cláudio Lembo (2006-2007), assumido devido à renúncia do governador Geraldo Alckmin para se candidatar à presidência da República, o Saresp não foi aplicado pela segunda vez desde sua criação. Segundo Ferreira (2007), isso ocorreu porque os dados de 2005 não haviam ainda sido analisados e estavam sendo estudadas alterações nos testes.

A partir de 2007, o Saresp adquire seu formato atual, testando anualmente os alunos dos 3º, 5º, 7º e 9º anos do ensino fundamental e da 3ª série do ensino médio, com exceção de 2007, em que também se testaram os alunos do 2º ano do ensino fundamental. Até então, a escala do Saresp não tinha correspondência com a escala da Prova Brasil e do Saeb. Em 2007, foi realizada uma prova de ligação, em que as duas escalas foram colocadas na mesma métrica, possibilitando que a partir daí os resultados do Saresp fossem comparados aos resultados do Saeb.

Se, por um lado, é interessante poder comparar resultados de testes diferentes a partir de uma mesma métrica, por outro, essa comparação perde seu significado, pois cada teste utiliza uma matriz diferente da outra. A comparação então só se faz em termos numéricos, nada mais, perdendo-se portanto, seu significado pedagógico. Por outro lado, esse procedimento dificulta possíveis manipulações de resultados por parte dos governos subnacionais, algo que acontece em alguns estados dos EUA (RAVITCH, 2010).

Durante o período entre 2008 e 2011, os procedimentos do Saresp permanecem os mesmos. Além das áreas de Leitura, Redação e Matemática, testadas todos os anos, a cada ano, alternadamente, são também testadas as áreas de História e Geografia, em um ano, e Ciências e Ciências da Natureza, no ano seguinte.

Além das escolas da rede estadual, cuja participação é obrigatória, podem participar do Saresp as escolas da rede municipal e da rede privada. Para que as escolas municipais participem, é necessário que seja firmado termo de cooperação entre os governos municipal e estadual, sem envolver custos. Já as escolas da rede privada são obrigadas a participar do teste desde 2009, e seus resultados são utilizados nos processos de renovação de autorização para funcionamento, apesar de não se ter informações disponíveis de como esse processo é realizado e sobre quais os efeitos dessa ação. Novamente, utiliza-se o teste do Saresp para um objetivo diferente daquele para o qual foi criado.

Quanto à construção dos instrumentos, nos anos iniciais da implantação dos testes avaliativos, as professoras eram responsáveis pela construção dos itens (SOUZA; OLIVEIRA, 2007), apesar de, nessa época, os testes terem sido montados por instituição contratada. Depois disso, instituições contratadas, como a Fundação para o Vestibular da Universidade Estadual Paulista (Vunesp), a Fundação Carlos Chagas (FCC), a Fundação Cesgranrio e o Centro de Políticas Públicas e Avaliação da Educação (CAEd), passaram a não somente fornecer os itens para o teste, como também ficaram responsáveis pela construção dos testes, sua aplicação, análise e divulgação dos resultados. Portanto, assim como ocorre em Minas Gerais, a Secretaria não tem o controle de todo o processo, pois os instrumentos de coleta são realizados por instituições externas a ela, assim como o são os dados gerados. Alguns podem até achar adequado que a aplicação do teste fique por conta de uma instituição externa, dadas as dificuldades de logística, mas não controlar a elaboração de itens, a construção dos testes, sua análise e divulgação dos resultados é abdicar de funções essenciais de qualquer processo avaliativo. É como se, guardadas as devidas proporções, a professora não participasse do processo avaliativo de seus alunos, deixando para outrem essa tarefa. É claro que a colaboração em algumas áreas onde ainda não haja expertise na Secretaria é importante, mas desde que os profissionais da Secretaria, com conhecimentos mínimos do processo avaliativo, acompanhem todas as etapas do processo. Dessa forma, a Secretaria estaria desenvolvendo competências na área e garantindo controle sobre todo o processo.

#### *Os níveis de proficiência do Saresp e as expectativas de aprendizagem*

A Secretaria definiu, baseada nas Orientações Curriculares, que são melhor especificadas no Apêndice C, e nos resultados do Saresp diferentes níveis de proficiência:



Ó Abaixo do básico: os alunos neste nível demonstram domínio insuficiente dos conteúdos, competências e habilidades requeridas para a série escolar em que se encontram.

Ó Básico: os alunos neste nível demonstram desenvolvimento parcial dos conteúdos, competências e habilidades requeridas para a série escolar em que se encontram.

Ó Adequado: os alunos neste nível demonstram conhecimentos e domínio dos conteúdos, competências e habilidades requeridas para a série escolar em que se encontram.

Ó Avançado: os alunos neste nível demonstram conhecimentos e domínio dos conteúdos, competências e habilidades além do requerido para a série escolar em que se encontram (SÃO PAULO, 2009a, p. 2).

Uma dificuldade trazida por essas definições é com relação à área de Língua Portuguesa. As Orientações Curriculares definidas pela Secretaria para a área tratam de expectativas de aprendizagem bastante amplas e a relação delas com as competências e habilidades, muito restritas e específicas, apresentadas nas Matrizes da área, são muito tênues. Assim, a descrição dos níveis de proficiência é muito frágil, como pode ser verificado na análise realizada no Apêndice C.

Assim, não parece acertado classificar como adequados para o ano escolar em que se encontram, por exemplo, conteúdos, competências e habilidades de uma dimensão menor, extrapolando-os para uma dimensão maior apresentada nas Orientações Curriculares. No fundo, é como se a Secretaria enviasse uma mensagem para a rede de ensino recomendando que o adequado é seguir as Matrizes, mais restritas, e não as Orientações Curriculares, de caráter mais amplo, estreitando-se o currículo praticado na escola. No caso de Matemática, a correspondência entre as Orientações e as Matrizes é muito grande, portanto têm lógica as definições utilizadas para os níveis de proficiência.

Outro problema causado pelas definições dos níveis de proficiência utilizados refere-se à adjetivação dada a cada nível. Essa adjetivação pode servir de fator discriminatório para as escolas onde os alunos obtiverem desempenhos õabaixo do básicoõ e õbásicoõ. Essas escolas seriam consideradas não adequadas?

Além de uma definição para cada um dos níveis de proficiência, a Secretaria estabeleceu uma correspondência entre os níveis e a escala de proficiência interpretada, denominando essa correspondência de *expectativas de aprendizagem*. O resultado dessa operação é apresentado no Quadro 3.11.

**Quadro 3.11 – Expectativas de aprendizagem definidas a partir dos níveis de proficiência do Saresp**

<b>Língua Portuguesa (Leitura)</b>			
<b>Níveis</b>	<b>5º ano EF</b>	<b>9º ano EF</b>	<b>3ª série EM</b>
Abaixo do Básico	<150	<200	<250
Básico	Entre 150 e <200	Entre 200 e <275	Entre 250 e <300
Adequado	Entre 200 e <250	Entre 275 e <325	Entre 300 e <375
Avançado	> 250	> 325	> 375
<b>Matemática</b>			
<b>Níveis</b>	<b>5º ano EF</b>	<b>9º ano EF</b>	<b>3ª série EM</b>
Abaixo do Básico	<175	<225	<275
Básico	Entre 175 e <225	Entre 225 e <300	Entre 275 e <350
Adequado	Entre 225 e <275	Entre 300 e <350	Entre 350 e <400
Avançado	> 275	> 350	> 400

Fonte: SÃO PAULO, 2009a, p. 2.

De acordo com o quadro apresentado, provavelmente, foi feita a descrição da escala de proficiência a partir de seus itens âncora para, com a ajuda de especialistas das duas áreas do conhecimento, escolher as descrições mais próximas do considerado adequado para cada ano escolar, definindo assim o Nível Adequado. O mesmo procedimento deve ter sido usado para indicar o Nível Básico. Os outros dois níveis relacionaram habilidades mais baixas que aquelas do Básico, sendo nomeado Nível Abaixo do Básico, ou relacionaram habilidades mais complexas do que aquelas agrupadas no Adequado, sendo nomeado de Nível Avançado. Todos os níveis têm uma extensão de 50 pontos, ou um desvio-padrão dentro da escala entre 0 e 500.

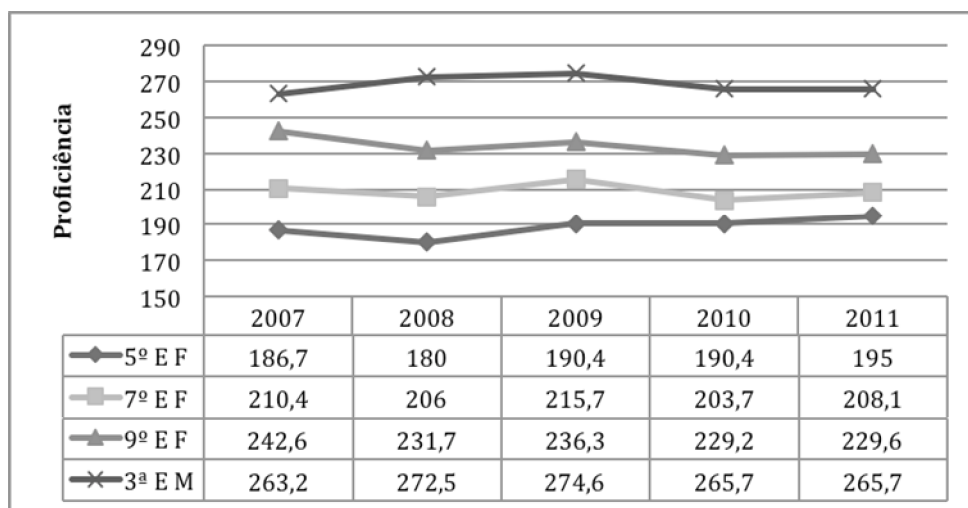
Mesmo que todo o procedimento descrito tenha sido realizado com maestria e por especialistas de renome e com grande preocupação com a educação, não se justifica que todo o processo não tenha sido descrito em detalhes e muito menos que tenha sido feito sem a participação ampla das professoras, inclusive para dar legitimidade a esse trabalho.

Além disso, essa classificação é frontalmente contrária a qualquer processo pedagógico.

#### ***Os resultados do Saresp***

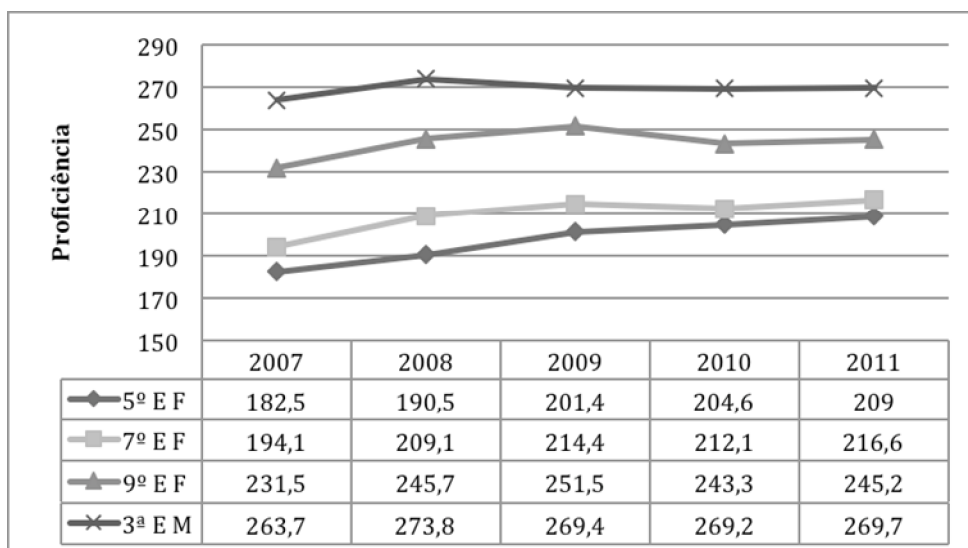
Para analisar a variação dos resultados obtidos pelo Saresp, tomaram-se os dados entre 2007 e 2011, devido à indisponibilidade de relatórios dos anos anteriores. Mesmo que estivessem disponíveis, a comparação seria impossível, pois antes de 2007 os desempenhos foram apresentados como percentuais de respostas certas do total de itens do teste, e os testes não eram comparáveis entre si.

A seguir, os Gráficos 3.10 e 3.11 mostram os valores das proficiências médias de Leitura e Matemática, respectivamente.



**Gráfico 3.10 – Proficiência média no Saresp em Leitura para os 5º, 7º e 9º anos do ensino fundamental e para a 3ª série do ensino médio – 2007 a 2011**

Fonte: SÃO PAULO, 2008, 2009, 2010, 2011; 2012, com elaboração do autor.



**Gráfico 3.11 – Proficiência média no Saresp em Matemática para os 5º, 7º e 9º anos do ensino fundamental e para a 3ª série do ensino médio – 2007 a 2011**

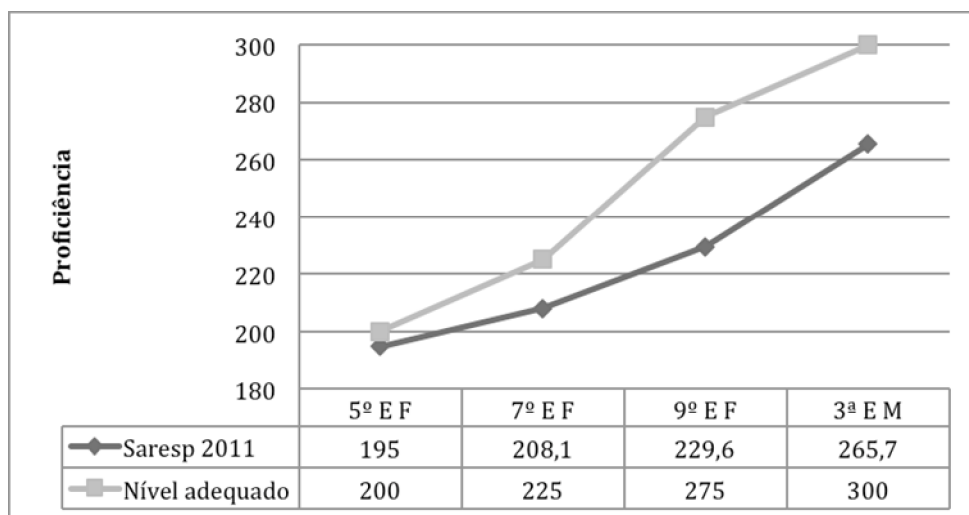
Fonte: SÃO PAULO, 2008, 2009, 2010, 2011; 2012, com elaboração do autor.

Primeiramente, nos Gráficos 3.10 e 3.11 não foram apresentados os resultados do 3º ano do ensino fundamental. Isso porque os valores medidos não permitem que sejam comparados com os dos outros anos, pois a escala utilizada é apresentada em uma métrica diferente das demais. Além disso, a forma de cálculo do desempenho é diferente. Cada item

presente no teste recebe uma pontuação de acordo com a complexidade da tarefa que o aluno deverá desenvolver, de forma que no total o teste de Língua Portuguesa atinja o valor máximo de 72 pontos e o de Matemática 100 pontos (SÃO PAULO, 2012). Como a TRI não é utilizada nesses cálculos, não faz sentido comparar os resultados das diferentes aplicações e nem do 3º ano do ensino fundamental com os demais anos escolares, pois são métricas diferentes. A Secretaria alega que, para contornar esse problema, os relatórios do período entre 2007 e 2011 comparam o percentual de alunos classificados nos diferentes níveis em que as escalas foram divididas, uma afirmação que não tem nenhum embasamento científico.

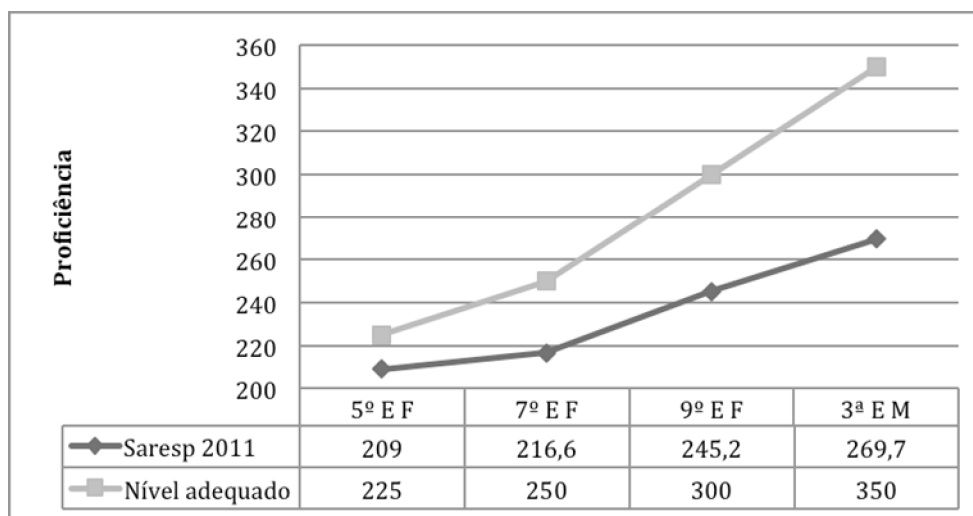
Com relação aos outros anos escolares, conforme já foi discutido sobre os resultados do Saeb, só é possível observar uma variação positiva consistente nas áreas de Leitura e Matemática, no período entre 2007 e 2011, no 5º ano do ensino fundamental. A exceção fica com Matemática no 7º ano, que apresenta um crescimento consistente. Para os outros anos, o movimento é muito irregular. Não fosse esse último caso, seria possível utilizar a mesma argumentação já utilizada quando discutido o Saeb de que a melhoria constante no 5º ano poderia ser, em parte, justificada por uma maior preparação para o teste que estaria acontecendo nas escolas. Segundo o argumento utilizado antes, essa prática seria mais efetiva nos anos iniciais por conta da unidocência, por possibilitar mais liberdade de tempo para as professoras lidarem com a organização do ensino. O desempenho crescente apenas no 5º ano e o fato de ter sido observado aumento de proficiência no 7º ano, apenas em Matemática, e uma trajetória errática para Leitura para o mesmo 7º ano deveriam ter sido investigados em detalhes.

O relatório de 2011 traz algumas informações, reproduzidas nos Gráficos 3.12, 3.13, que merecem ser discutidas e que envolvem a proficiência demonstrada pelos alunos no Saresp e o Nível de Proficiência Adequado fixado pela Secretaria.



**Gráfico 3.12 – Distanciamento das médias de proficiência aferidas no Saresp 2011, em relação à expectativa do Nível Adequado para os anos/séries testados em Leitura**

Fonte: SÃO PAULO, 2012, p. 44.



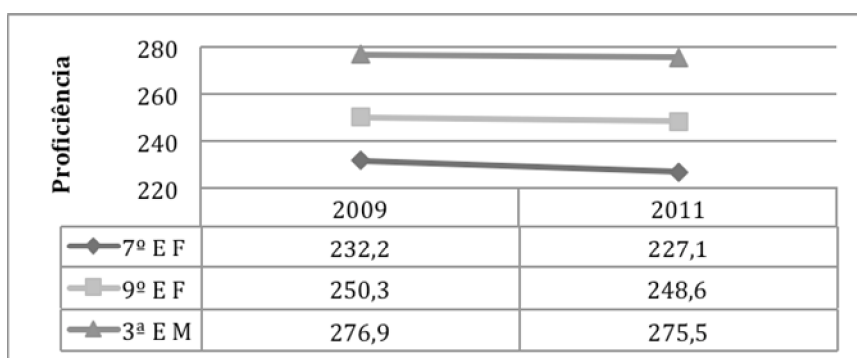
**Gráfico 3.13 – Distanciamento das médias de proficiência aferidas no Saresp 2011, em relação à expectativa do Nível Adequado para os anos/séries testados em Matemática**

Fonte: SÃO PAULO, 2012, p. 47.

Nos dois gráficos, é possível observar que as proficiências médias demonstradas pelos alunos no Saresp 2011 estão aquém das metas fixadas, a partir da definição dos Níveis Adequados para cada ano escolar. Percebe-se, também, que, com o avanço da escolaridade, aumenta o distanciamento entre a proficiência média e o Nível Adequado, distanciamento que é menor para o 5º ano do ensino fundamental e maior na 3ª série do ensino médio. Observa-se

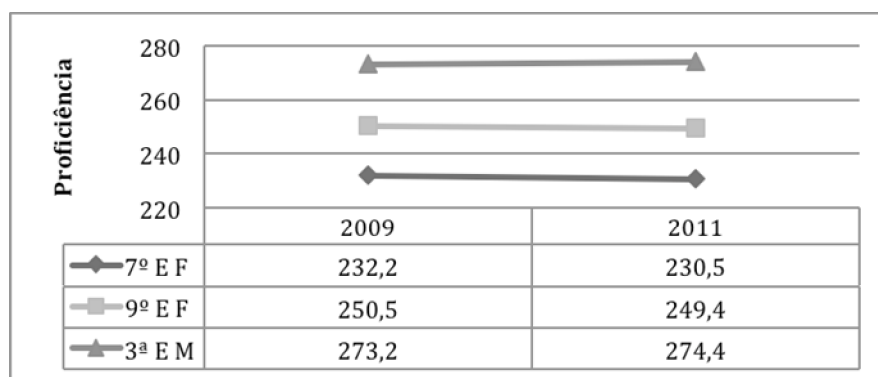
pelo gráfico que com o aumento da escolaridade aumenta o distanciamento daquilo que foi chamado de Nível Adequado. Muitas podem ser as causas desse fenômeno, entre elas estaria a não compreensão, em anos escolares anteriores, de conceitos básicos fundamentais para dominar habilidades mais complexas, apresentadas em anos escolares posteriores. Esse raciocínio é válido, caso se acredite que as metas fixadas para o nível adequado tenham sido calculadas respeitando a série histórica do Saresp. Do contrário, a comparação de desempenho com o nível adequado não faz qualquer sentido. Como não foi explicitada a lógica para a definição desses níveis, esse tema fica como indicação para novos estudos.

Para ilustrar a variação da proficiência para as áreas de Geografia, História, Ciências e Ciências da Natureza, apresentam-se os Gráficos 3.14, 3.15 e 3.16, respectivamente.



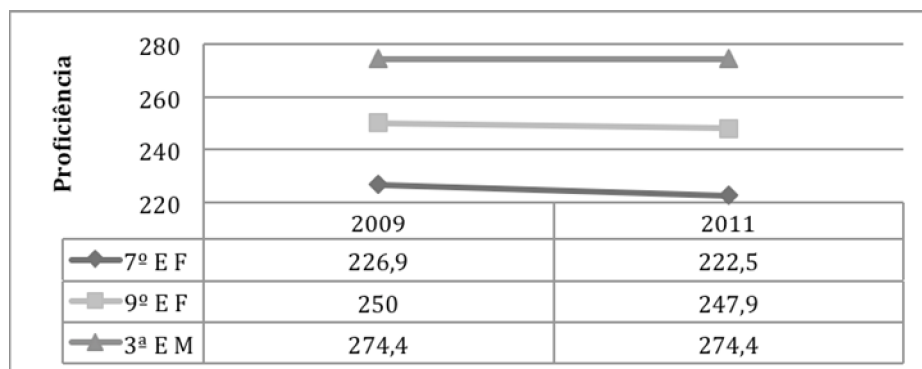
**Gráfico 3.14 – Proficiência no Saresp em Geografia para os 7º e 9º anos do ensino fundamental e para a 3ª série do ensino médio – 2009 a 2011**

Fonte: SÃO PAULO, 2010, 2012, com elaboração do autor.



**Gráfico 3.15 – Proficiência no Saresp em História para os 7º e 9º anos do ensino fundamental e para a 3ª série do ensino médio – 2009 a 2011**

Fonte: SÃO PAULO, 2010, 2012, com elaboração do autor.



**Gráfico 3.16 – Proficiência no Saresp em Ciências e Ciências da Natureza em 2008 e 2010 para os 7º e 9º anos do ensino fundamental e para a 3ª série do ensino médio**

Fonte: SÃO PAULO, 2009; 2011, com elaboração do autor.

Apesar de o período entre um teste e outro ter sido muito curto, afinal trata-se de apenas duas aplicações para cada área, observa-se que para nenhuma das áreas houve diferenças significativas no desempenho. Talvez com novas aplicações possa ser possível identificar alguma tendência que permita explicar o que pode estar acontecendo nas quatro outras áreas.

No Apêndice C, são discutidos os possíveis impactos do instrumental utilizado pelo Saresp sobre o currículo e a aprendizagem.

#### *As ações e os programas que utilizam os resultados do Saresp à luz da legislação*

São listados e comentados, a seguir, os resultados da pesquisa sobre a legislação emanada pelos diversos governos paulistas a partir do ano 2000, que levaram em consideração os resultados do Saresp. Com isso, será possível acompanhar como uma série de políticas *high stakes* vai, aos poucos, sendo implantada na rede estadual de São Paulo.

Em 2000, no governo Mário Covas (1999-2001), é aprovada a Lei Complementar nº 890, que instituiu o Bônus Gestão para os envolvidos no suporte pedagógico da escola: Dirigentes Regionais de Ensino, Supervisores de Ensino, Diretores, Vice-Diretores e Assistentes de Diretores de Escola, Coordenador Pedagógico. Para a avaliação do desempenho desses profissionais, deveriam ser levados em consideração, entre outras coisas, o desempenho no Saresp e os índices de evasão escolar (SÃO PAULO, 2000). Um decreto do governador, no ano seguinte, instituiu uma Menção Honrosa para aqueles que tenham recebido o valor máximo do Bônus Gestão (SÃO PAULO, 2001a). Essa lei pode ser considerada o

embrião para a criação, oito anos depois, da Bonificação por Resultados instituído na rede estadual paulista.

Em 2001, no primeiro governo Geraldo Alckmin (2001/2003), foram premiadas escolas com base no desempenho do Saresp do ano anterior, levando em conta os resultados obtidos e as taxas de evasão escolar. Das 4.100 escolas participantes do teste, 1.025 foram selecionadas, sendo que 345 classificadas na categoria *Destaque*, premiadas com viagens de cunho histórico/cultural para vários Estados do Brasil, e 680 consideradas *Acima da média*, que viajariam para diversos municípios paulistas (CEE, 2001).

Em 2002, sabe-se que os resultados foram utilizados para orientar o Programa de Formação Continuada *Teia do Saber*, instituído pela Secretaria naquele ano e voltado para as professoras da rede (SOUZA; OLIVEIRA, 2007). No entanto, não foi possível encontrar documentação sobre o Programa apontando os efeitos sobre o ensino e como os resultados dos testes foram utilizados pelo *Teia do Saber*.

Em 2007, no governo José Serra (2007/2010), a Secretaria, por meio da Resolução SE nº 86, institui o Programa Ler e Escrever para os anos iniciais do ensino fundamental, com o objetivo de alfabetizar, em três anos, até 2010, todos os alunos com idade de até 8 anos.

Nas considerações, a Resolução destaca a urgência em solucionar as dificuldades apresentadas pelos alunos dessa etapa de ensino, com relação às competências de ler e escrever, expressas nos resultados do Saresp 2005 (SÃO PAULO, 2007). Esse Programa, segundo o documento consultado, baseou-se em uma experiência bem-sucedida desenvolvida pela prefeitura do município de São Paulo desde 2006. Para apoiar sua implantação, foi lançada uma série de materiais didáticos destinados às professoras. Também em 2007, a Secretaria institui a função gratificada de Professor Coordenador, profissional que atuaria como gestor implementador das políticas da Secretaria na área de coordenação pedagógica, com o objetivo de *ampliar* o domínio dos conhecimentos e saberes dos alunos, elevando o nível de desempenho escolar evidenciado pelos instrumentos de avaliação externa e interna (SÃO PAULO, 2007a). Portanto, ao que parece, função similar àquela desempenhada pelo Especialista em Educação Básica em Minas Gerais, conforme discutido no capítulo anterior.

Em 2008, o Plano Plurianual (PPA) para o período 2008-2011, instituído pela Lei nº 13.123, estabelece o desafio de garantir um ensino de qualidade, a partir do cumprimento de dez metas. Entre elas: todos os alunos de 8 anos de idade deveriam ser plenamente alfabetizados e os índices de desempenho dos ensinos fundamental e médio nas avaliações nacionais e estaduais deveriam aumentar 10%. Define também seis eixos para promover as



condições de melhoria da aprendizagem, entre eles a avaliação do sistema educacional como processo indutor da melhoria da aprendizagem (SÃO PAULO, 2008c).

O ano de 2008, a Secretaria, comandada por Maria Helena Castro, instituiu uma série de importantes e profundas mudanças. As primeiras medidas foram a criação do Programa São Paulo Faz Escola, que envolveu a reformulação curricular da rede e programas de formação continuada de professoras a distância, para colocar um novo currículo em prática nas escolas e instituir aulas de reforço escolar no início do ano letivo, destinadas aos alunos da 5ª à 8ª série do ensino fundamental e para o ensino médio com duração de 42 dias, nas áreas de Língua Portuguesa e Matemática (SÃO PAULO, 2008e).

Outra alteração de fundo, instituída em 2008, foi dada pela Resolução nº 74, que criou o Programa Qualidade da Escola e o Índice de Desenvolvimento da Educação do Estado de São Paulo (Idesp). Essa resolução estabeleceu que o Idesp seria usado para o cálculo das metas de qualidade fixadas para cada escola (SÃO PAULO, 2008d).

Além disso, a Resolução nº 74 definiu o Idesp como indicador coletivo para a bonificação por desempenho dos servidores. Essa bonificação foi estabelecida pela Lei Complementar nº 1.078/2008 (SÃO PAULO, 2008f), reformulando duas Leis Complementares anteriores: a de nº 1.016/2007 e a de nº 1.017/2007, que trataram sobre a concessão de Bônus e Bônus Merecimento, respectivamente.

Também em 2008, é editada a Resolução SE nº 76, que dispõe sobre a implementação da Proposta Curricular para o Estado de São Paulo, que passa a ser o referencial obrigatório para a formulação da proposta pedagógica da escola (SÃO PAULO, 2008g).

Outra medida de 2008 foram as normas para disciplinar o estágio probatório dos professoras, estabelecidas pela Resolução SE nº 66/08. Para avaliar as professoras, são criados sete indicadores, e entre eles destacam-se: *responsabilidade*, definido como o comprometimento com os objetivos pactuados nos planos de trabalho da escola e da Diretoria de Ensino, de acordo com as metas da Secretaria; *produtividade*, ditado pelas contribuições para a melhoria do nível de desempenho dos alunos, da escola e da Diretoria de Ensino. Os dois indicadores são altamente subjetivos, procurando estabelecer uma relação entre o esforço da professora e a melhoria do desempenho dos alunos.

Em 2009, o Conselho Estadual de Educação, em sua Deliberação CEE nº 84/09, determinou que as instituições particulares de educação básica se inscrevessem no Saesp, e que seus resultados seriam considerados elementos fundamentais nos procedimentos gerais de avaliação, supervisão e regulação (SÃO PAULO, 2009d). Nesse mesmo ano, a Resolução SE nº 90 dispõe sobre a definição de perfis profissionais e de competências e habilidades

requeridos para Supervisores e Diretores de Escola: identificar o papel dos resultados do Saresp na construção do Idesp e as semelhanças e diferenças entre o Idesp e o Ideb, reconhecer as principais características dos sistemas de avaliação da educação básica e compreender os conceitos básicos que os fundamentam, em especial o Saresp (SÃO PAULO, 2009e). Ainda em 2009, é editada a Resolução SE nº 18, que dispõe sobre os estudos de recuperação e determina que as atividades a serem desenvolvidas deverão, prioritariamente, repor as estruturas lógico-matemáticas e linguísticas (SÃO PAULO, 2009g).

Ainda nesse ano, a Secretaria edita a Resolução nº 70/2009, que dispõe sobre os perfis profissionais, as competências e as habilidades requeridas dos educadores da rede pública estadual e os referenciais bibliográficos que fundamentam os exames, os concursos e os processos seletivos. Entre as competências e as habilidades requeridas encontram-se: usar os resultados do Saresp para a tomada de decisões pedagógicas; identificar as principais características do Saresp; e interpretar adequadamente o Ideb e o Idesp, como se constroem, para o que servem e o que significam para a educação escolar brasileira e paulista (SÃO PAULO, 2009h).

Em 2010, o governo do Estado edita o Decreto nº 55.650/10, instituindo o Programa Rede São Paulo de Formação Docente (Redefor), com a meta de oferecer, até 2012, a 30 mil docentes, cursos de especialização com duração entre 12 e 14 meses. Não foram obtidas informações sobre a realização de algum tipo de avaliação do programa.

Em 2011, durante o segundo governo de Geraldo Alckmin (2011-2015), é publicado no *Diário Oficial do Estado*, de 29/12/11, p. 20, o PPA para o período 2012-2015. Com relação à área de Educação, são instituídas cinco ações prioritárias, uma delas referente ao Programa Escolas Prioritárias, dentro do Programa Educação é Compromisso de São Paulo. Esse Programa é voltada para as 1.206 escolas da rede que apresentam os maiores índices de vulnerabilidade educacional, com a finalidade de, em quatro anos, elas alcançarem melhoria significativa nos índices de aproveitamento escolar. O programa também considera a continuidade de programas como Ler e Escrever, São Paulo Faz Escola, Escola da Família, Escola de Formação e Aperfeiçoamento de Professores (SÃO PAULO, 2011b).

Ainda em 2011, é publicado o Decreto nº 57.141, que reorganiza a Secretaria de Educação. Por essa norma, a Coordenadoria de Informação, Monitoramento e Avaliação Educacional, órgão da estrutura básica da Secretaria, e dentro dela o Departamento de Avaliação Educacional, é definida como a estrutura responsável pelo planejamento e análise das avaliações e pela sua aplicação. Cabe a outra estrutura, o Departamento de Desenvolvimento Curricular e Gestão da Educação Básica, ligada à Coordenadoria de Gestão

da Educação Básica, propor a definição de políticas, diretrizes e parâmetros para processos de avaliação de desempenho dos alunos da educação básica; analisar os resultados das avaliações e sugerir medidas de correção de rumo (SÃO PAULO, 2011a).

Uma ação polêmica, também em 2011, foi regulamentada pela Resolução SE nº 73, que instituiu o Prêmio Saesp 2011, para distribuir um total de 12 mil *notebooks* a alunos concluintes do ensino médio das escolas da rede que obtivessem os melhores resultados no teste (SÃO PAULO, 2011d). Apesar de não ser o foco desta pesquisa discutir o ensino médio, essa ação reproduz uma prática que vem sendo adotada em diferentes Estados com o intuito de elevar o desempenho dos alunos nos testes. Como já discutido no caso do Saeb, os alunos, prestes a concluir a educação básica, apresentam alta taxa de absenteísmo no dia dos testes e se envolvem muito pouco com ele. Com a premiação, procura-se, artificialmente, um maior envolvimento desses alunos e com isso um aumento das proficiências. Não foi possível identificar nenhum estudo sobre o impacto dessa ação.

Uma questão muito séria e que deve ser debatida com intensidade foi a publicação da Resolução SE nº 81, ainda em 2011, que estabeleceu as diretrizes para a organização curricular da educação básica. A síntese dessa resolução é a Matriz Curricular Básica, que determina as porcentagens do tempo de aula ou o número de aulas que devem ser dedicadas a cada área do conhecimento, conforme pode ser observado no Quadro 3.12.

**Quadro 3.12 – Matriz Curricular Básica para os ensinos fundamental e médio**

Disciplinas	Ano/Quantitativo de aulas											
	Ensino Fundamental									Ensino Médio		
	1º	2º	3º	4º	5º	6º	7º	8º	9º	1ª	2ª	3ª
Língua Portuguesa	60%	60%	45%	30%	30%	6	6	6	6	-	-	-
Língua Portuguesa e Literatura	-	-	-	-	-	-	-	-	-	5	5	5
Matemática	25%	25%	40%	35%	35%	6	6	6	5	5	5	5
Ciências	-	-	-	10%	10%	4	4	4	4	-	-	-
História/Geografia	-	-	-	10%	10%	-	-	-	-	-	-	-
Educação Física/Arte	15%	15%	15%	15%	15%	-	-	-	-	-	-	-
Arte	-	-	-	-	-	2	2	2	2	2	2	2
Educação Física	-	-	-	-	-	2	2	2	2	2	2	2
História	-	-	-	-	-	4	4	4	4	2	2	2
Geografia	-	-	-	-	-	4	4	4	4	2	2	2
Ensino Religioso <sup>(1)</sup>	-	-	-	-	-				1	-	-	-
Língua Estrangeira Moderna	-	-	-	-	-	2	2	2	2	2	2	2
Biologia	-	-	-	-	-	-	-	-	-	2	2	2
Física	-	-	-	-	-	-	-	-	-	2	2	2
Química	-	-	-	-	-	-	-	-	-	2	2	2
Filosofia	-	-	-	-	-	-	-	-	-	2	2	2
Sociologia	-	-	-	-	-	-	-	-	-	2	2	2
Total	100%	100%	100%	100%	100%	30	30	30	30	30	30	30

Fonte: SÃO PAULO, 2011e, com elaboração do autor.

\*Nota: (1) Caso não haja demanda para a disciplina, acrescentar uma aula para Matemática.

Percebe-se, pela análise do Quadro 3.12, a importância dada pela Secretaria para que nos anos iniciais do ensino fundamental a quase totalidade da carga horária estivesse destinada às áreas de Língua Portuguesa e Matemática, as duas que são testadas pelo Saresp dedicando, entre o 1º e 3º anos do ensino fundamental, apenas 15% da carga horária restante para Artes e Educação Física. Nesse caso, observa-se uma coincidência de objetivos entre o currículo e o Saresp. As áreas de Ciências, História e Geografia só aparecem na Matriz Curricular Básica no 4º e no 5º ano do ensino fundamental, o que evidencia a opção pela redução do currículo nos anos iniciais, provavelmente com o objetivo de garantir o cumprimento da meta de alfabetizar as crianças até os 8 anos de idade. Além da redução do currículo oficial, não se pode perder de vista que na escola pode haver nova redução curricular para se adequar às Matrizes do Saresp.

Outro ponto importante foi a paulatina inclusão de temas ligados ao Saresp e, depois de sua criação, do Idesp nos programas dos processos seletivos de Diretores e Vice-Diretores de escola, Supervisores, Coordenadores Pedagógicos e Professoras, de tal forma que esses dois temas estão cada vez mais presentes no dia a dia das escolas, reforçando de alguma forma aquilo que se cunhou como a cultura da avaliação, deixando claro que nesse caso a referência é à cultura da avaliação externa, algo vago e sem sentido.

Nos doze anos analisados, destacando-se as normas que mais impactos poderiam causar nas escolas, é possível perceber o quanto o Saresp vai se distanciando de seus objetivos iniciais, deixando de ser um instrumento que as escolas pudessem utilizar na tomada de decisões com vistas a melhorar o ensino oferecido, para, paulatinamente, se transformar em um instrumento de pressão por melhores resultados da escola, passando, inclusive, pela decisão de 2001, de promover ou não os alunos com base em seus resultados. Todo esse processo foi delineado em 1995, no comunicado da secretária Rose Neubauer, mas precisou de um longo percurso para ser implementado. Destaque-se, também, que apesar de o Estado ter sido governado por um único partido, o PSDB, isso não impediu que fossem tomadas decisões conflitantes conforme se alteravam os secretários de educação. Isso evidencia que uma mesma orientação partidária não é condição suficiente para a continuidade das ações.

#### *O Idesp, o estabelecimento de metas e o cálculo para a bonificação por resultados*

Em 2009, dentro do Programa de Qualidade da Escola (PQE), foi instituído o Índice de Desenvolvimento da Educação Básica (Idesp), um indicador que avalia a qualidade da escola (SÃO PAULO, 2009a, p. 1). Ainda segundo o documento, uma boa escola seria aquela em que a maior parte dos alunos aprende as competências e habilidades requeridas para a série, num período de tempo ideal o ano letivo (SÃO PAULO, 2009a, p. 1).

Como o Ideb, o cálculo do Idesp leva em consideração a proficiência dos alunos, nesse caso medida pelo Saresp, e o fluxo escolar, estabelecendo um número que varia entre 0 e 10.

O Idesp de cada etapa de escolarização é dado por:

$$IDESP = ID_s \times IF_s$$

Na equação, ID representa o Indicador de Desempenho, IF representa o Indicador de Fluxo e o subscrito *s* indica uma etapa de escolarização, 5º ou 9º ano do ensino fundamental

ou 3ª série do ensino médio. Portanto, apesar de os testes do Saresp serem aplicados também a outros anos escolares o Idesp só é calculado para esses três anos escolares, um contrassenso.

O ID de uma etapa de escolarização,  $ID_s$ , é calculado como a média do índice de desempenho em Matemática e Leitura:

$$ID_s = \frac{ID_{lp} + ID_{mat}}{2}$$

Portanto, apesar de o Saresp testar outras áreas, apenas as de Leitura e Matemática entram no cômputo do Idesp, outro contrassenso.

O indicador de desempenho de cada uma das duas áreas é calculado em função da distribuição percentual dos alunos da escola, em cada um dos quatro níveis de proficiência instituídos pela Secretaria, quais sejam: Abaixo do Básico, Básico, Adequado e Avançado:

$$ID_{js} = \left(1 - \frac{def_{js}}{3}\right) \times 10$$

Assim, o  $def_{js}$  refere-se à distribuição dos alunos por nível de proficiência, para a área de conhecimento  $j$  e ano escolar  $s$  e é dado por:

$$def_{js} = \frac{3 \times AB + 2 \times B + 1 \times Ad + 0 \times Av}{100},$$

sendo que,  $AB$ ,  $B$ ,  $Ad$  e  $Av$  referem-se às porcentagens de alunos da escola que se encontram nos níveis de proficiência Abaixo do Básico, Básico, Adequado e Avançado, respectivamente. O ID de uma escola pode assumir qualquer valor entre 0, quando 100% dos alunos estão no Nível Abaixo do Básico, e 10, quando 100% dos alunos estão no Nível Adequado. Portanto, diferentemente do Ideb, que se atém apenas à proficiência média dos alunos, o Idesp preocupa-se com a distribuição das proficiências.

O Indicador de Fluxo (IF), a segunda parcela do Idesp, é dado por:

$$IF_s = \frac{\sum_{i=1}^n A_i}{n},$$

onde,  $A_i$  é a taxa de aprovação no ano escolar  $i$  e  $n$  é o número de séries da etapa de escolarização considerada, ou seja,  $n = 5$  para os anos iniciais do ensino fundamental,  $n = 4$  para os anos finais e  $n = 3$  para o ensino médio. Assim, para cada ano escolar testado, leva-se em consideração a taxa de aprovação de toda a etapa de ensino a que ele pertence.

Assim como no Ideb, foram estabelecidas também metas para as escolas da rede estadual, que, no caso de São Paulo, foram fixadas para 2030, também tomando como parâmetro o desempenho dos países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) (SÃO PAULO, 2011c). As metas globais para 2030, estipuladas para cada um dos anos escolares testados, foram de 7,0 e 6,0 para o 5º ano e o 9º ano do ensino fundamental, respectivamente, e 5,0 para a 3ª série do ensino médio. Tendo por base o ano de 2009 e a meta para 2030, foram estabelecidas as metas anuais para todas as escolas. Assim como no caso do Ideb, as escolas receberam suas metas sem que houvesse nenhum processo de discussão anterior.

Estabeleceu-se também um índice de cumprimento de metas,  $IC_s$ , para cada ano escolar, calculado como:

$$IC_s = \left( \frac{IDESP_{s,atual} - IDESP_{s,anterior}}{META_{s,atual} - IDESP_{s,anterior}} \right) + \left( \frac{IDESP_{s,atual} - IDESPagregado_{s,atual}}{META_{s,2030} - IDESPagregado_{s,atual}} \right)$$

Na equação,  $s$  indica a etapa da escolarização; *atual* e *anterior* referem-se, respectivamente, aos anos atual e anterior e *IDESPagregado* refere-se à média do desempenho de todas as escolas da rede. O  $IC_s$  nada mais é que o somatório de duas parcelas, a primeira, chamada de Parcela Cumprida da Meta, calcula a parcela da meta fixada que foi cumprida pela escola no ano e a segunda, chamada de Adicional por Qualidade, indica a variação do Idesp da escola, com relação à média das escolas da rede e à meta de longo prazo. Portanto, compara-se a escola em relação ao cumprimento tanto de suas metas como em relação ao desempenho das escolas da rede.

Tanto a Parcela Cumprida da Meta como o Adicional de Qualidade serão utilizados para o cálculo da bonificação por resultados, paga aos profissionais da rede estadual, que pode chegar a 1,2 salário.

Em resumo, a avaliação no Estado de São Paulo também começou com o objetivo de avaliar programas que estavam em desenvolvimento e depois foi estendida para o conjunto da rede, mas no caso paulista não existem relatos de que o processo tenha sido participativo,

como ocorreu em Minas. Alguns momentos merecem destaque nessa trajetória. O primeiro, quando a Secretaria, em 1995, estabelece em comunicado as linhas mestras de sua gestão, destacando diretrizes claramente afinadas com a regulação segundo o modelo de governança do Estado Avaliador. Nesse comunicado, indicava-se a necessidade da avaliação dos resultados da aprendizagem, de maneira a responsabilizar as escolas pelos resultados e como forma de prestação de contas que permitisse às escolas comparar-se com as outras e, a partir daí, buscar melhorias, assim como permitisse às famílias serem informadas sobre a educação fornecida pelas escolas de seus filhos. Essa ação é uma das poucas vezes em que se observa a gestão expondo claramente os preceitos que embasarão suas ações. Outro momento importante no caso paulista ocorreu a partir de 2007, completando as ações anunciadas em 1995. Os resultados do Saresp passam a ser comparáveis ao Saeb; cria-se o Idesp, aprimorando algumas das inconsistências do Ideb; e a partir dele, estabelecem-se metas para todas as escolas, que passam a ser premiadas pelos resultados alcançados.

No capítulo seguinte será realizada a análise comparada dos três casos tratados nessa pesquisa.



## CAPÍTULO 4

### ANÁLISE COMPARADA DOS CASOS

Nesse capítulo apresenta-se a análise comparativa entre os três casos estudados envolvendo os testes avaliativos utilizados pelo governo federal e os dos Estados de Minas Gerais e São Paulo.

Com relação às condições objetivas para o desenvolvimento dos testes no Brasil, ficou evidente que isso foi possível devido à expertise alcançada graças aos grandes vestibulares dos anos 1970, que trouxeram os conhecimentos para a construção dos testes e também para a sua aplicação e apuração dos resultados. Tanto isso é real que as duas principais instituições envolvidas nos anos iniciais dos testes avaliativos foram a Fundação Carlos Chagas e a Fundação Cesgranrio, as duas envolvidas com os grandes vestibulares. Sem esses conhecimentos, teria sido mais difícil o desenvolvimento e a trajetória dos testes avaliativos.

Um segundo aspecto, que também marcou os anos iniciais dos testes avaliativos no país, foi sua utilização como ferramenta para avaliar programas educacionais em andamento, como se pode verificar no âmbito do governo federal e no governo de Minas Gerais, ambos coordenados pelo Prof.º Heraldo Vianna ligado à Fundação Carlos Chagas. No governo federal, o Projeto Edurural, implantado em 1980 e desenvolvido em parceria com o Banco Mundial, deu origem aos primeiros estudos nacionais entre 1988 e 1990 e ao primeiro ciclo do Saeb em 1990. No caso de Minas Gerais, a avaliação do Ciclo Básico de Alfabetização realizada em 1988, originou, em 1991, o Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais, financiado com recursos do Projeto Qualidade da Educação Básica de Minas Gerais, projeto desenvolvido em parceria entre o governo estadual e o Banco Mundial, que posteriormente deu origem ao Simave.

O caso de São Paulo não foi diferente. Havia sido implantado na rede estadual, em 1991, o Projeto Escola Padrão e no ano seguinte foi criado o Programa de Avaliação Educacional da Rede Estadual com o objetivo de avaliar aquele Projeto. Entretanto, trazia em seu bojo objetivos mais abrangentes, como o de trabalhar com as escolas pontos críticos do processo de aprendizagem detectados pelos testes. Verifica-se, ainda, que no período entre 1992 e 1994 que os testes tinham, também, como finalidade avaliar os impactos das políticas educacionais estaduais, constantes do Projeto Inovações na Educação Básica em São Paulo, desenvolvido em parceria com o Banco Mundial. Em 1996, depois dessas duas experiências, nasce o Saresp.

Tanto nos casos do Edurural e dos estudos nacionais desenvolvidos pelo governo federal, como nos casos de Minas Gerais e de São Paulo, durante os programas de avaliação desenvolvidos antes da implantação de seus sistemas próprios de avaliação, percebe-se claramente uma preocupação com os impactos na aprendizagem dos alunos. No entanto, é em Minas Gerais que essa preocupação se fez mais presente, pois contou no seu início com a participação dos professores na elaboração dos instrumentos, apuração dos resultados e sua análise, dentro do espírito de uma avaliação cooperativa, como determina a Constituição estadual. Mas, mesmo nesse caso, tinha características de um projeto de intervenção nas práticas escolares, pois pretendia-se, com os modelos de testes utilizados, induzir a melhoria da qualidade das provas aplicadas pelas professoras, além de contribuir para que se fortalecesse a prática de atividades conjuntas entre as professoras, a partir das atividades de correção dos testes e elaboração dos relatórios. De qualquer forma, era um processo com maior abrangência, pois esperava-se muito mais do que apenas obter resultados sobre os desempenhos nos testes.

Ao longo da trajetória de cada um dos testes avaliativos, nota-se que houve modificações significativas. O teste do governo federal não sofreu nenhuma interrupção na sua periodicidade, apesar dos diversos governos de diferentes partidos que se sucederam no poder, se iniciando na época da ditadura, passando pela eleição direta do presidente Fernando Collor de Mello e chegando ao governo Dilma Rousseff. Houve, porém, no início do primeiro governo Luiz Inácio Lula da Silva, em 2003, uma movimentação no sentido de "redebater os processos de avaliação e adequá-los à situação de um novo governo que tem um compromisso social com a área da Educação, diferente do governo anterior, que tinha compromisso com um perfil liberalizante ou mercantilista de educação" (MEC, 2003). No entanto, o que se viu nos últimos anos, em que o Partido dos Trabalhadores (PT) tem-se mantido no poder central, foi a crescente ampliação do número de testes avaliativos como a Provinha Brasil e a recente Avaliação Nacional da Alfabetização, bem como a transformação do Saeb em Prova Brasil censitária. Reforça-se, assim, o mito de que o baixo resultado nos testes pode ser solucionado com a aplicação de mais testes (RAVITCH, 2010). Com relação aos instrumentos, ao formato dos testes e às Matrizes de Referência, o Saeb sofreu alterações significativas, aumentando-se a quantidade de itens nos testes e diminuindo-se significativamente a quantidade de habilidades testadas, sem que nenhum estudo sobre o impacto dessas alterações tivesse sido realizado. A consequência dessas mudanças é que a Prova Brasil consegue captar menos nuances sobre a aprendizagem do que a dos testes do ciclo de 1997, quando as Matrizes foram criadas.

No caso dos dois Estados, ocorreram mudanças das propostas iniciais e alterações nos instrumentos utilizados. Em Minas, o Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais, que vai de 1992 a 1996, é substituído pelo Simave em 2000, mas que deixou de ser aplicado somente em 2004. Apesar de o Partido da Social Democracia Brasileira (PSDB) ter se mantido no poder, no governo estadual, desde 2003, em termos dos instrumentos utilizados nesse Estado, houve mudanças, especialmente as relacionadas à variação nas áreas de conhecimentos testadas. Porém, os anos escolares selecionados para fazer parte dos testes, as Matrizes e o número de itens dos testes não se alteraram. Um ponto a ser destacado, no caso mineiro, é o aumento no controle sobre a escola, com a criação, em 2005, do Programa de Avaliação da Aprendizagem Escolar (PAAE), o terceiro teste avaliativo a compor o Simave. O PAAE, gerenciado por um sistema *on line*, controla o que está sendo ensinado em sala de aula, determina como o processo educacional deve ser avaliado e mede o desempenho de cada uma das professoras da rede. Através dele o controle sobre a escola é mais efetivo e apresenta custo mais baixo do que as avaliações externas, com a vantagem adicional de se obter os resultados mais rapidamente. No entanto, não se tem conhecimento sobre o que é feito com os dados apurados, reforçando a suspeita de que o objetivo principal é utilizá-lo como instrumento de regulação, algo que vai além de simplesmente baixar uma norma, conforme discutido por Barroso (2005) e Maroy (2011 e 2011a).

Dos três casos analisados, São Paulo é onde o mesmo partido governa há mais tempo: o PSDB tem permanecido no governo estadual desde 1995. Com relação aos testes avaliativos, o Programa de Avaliação Educacional da Rede Estadual foi utilizado em 1991 e 1992, surgindo o Saesp em 1996, que deixou de ser aplicado em 1999 e 2006. Portanto, mesmo não tendo havido mudança de partido no governo do Estado, isso não foi suficiente para evitar interrupções na aplicação do Saesp. Essa mesma estabilidade também não impediu que o desenho inicial dos testes, que procuravam por informações longitudinais, viesse a ser mudado. Um dos acontecimentos mais marcantes na história da avaliação educacional no Brasil aconteceu em 2001, quando os resultados do Saesp foram utilizados para promover ou não os alunos para o ano escolar seguinte, interferindo diretamente na autonomia das professoras e das escolas. Apesar de esse fato não ter se repetido mais, provavelmente por causa das resistências das escolas a essa iniciativa, o uso indevido dos resultados dos testes aumentou muito a desconfiança das escolas com relação ao Saesp (SOUZA e OLIVEIRA, 2007). Apesar de terem ocorrido outras modificações na estrutura do Saesp, desde 2008 o teste avaliativo vem mantendo a mesma estrutura.

Ainda com relação aos instrumentos utilizados, é importante discutir qual o controle que tanto os testes federais como os dos dois Estados analisados têm sobre os itens elaborados para fazerem parte dos testes que desenvolvem. No caso do governo federal, desde o início da aplicação do Saeb, os itens eram desenvolvidos pela instituição vencedora da licitação para a aplicação dos testes ou então adquiridos por processo licitatório aberto a instituições com experiência na área. Desde 2011, o Inep vem trabalhando com diversas universidades que, após treinamento realizado pelo Instituto, elaboram itens para depois serem validados ou não pelos pesquisadores da autarquia. Com isso, tem-se garantido maior controle sobre o processo, gera-se conhecimento, difundem-se as técnicas de elaboração de itens de múltipla escolha e, tão importante quanto os demais fatores, aumenta-se a qualidade dos itens e a confiabilidade dos testes que os utilizarão. No caso dos dois Estados, desde a criação do Saresp, quanto do Simave, a coordenação da elaboração e escolha dos itens é feito por instituições que vencem os processos licitatórios anteriores à aplicação dos testes.

Com o passar dos anos, observa-se que os Sistemas de Avaliação, instituídos pelos dois Estados e pelo governo federal assumem o caráter de instrumento de regulação do sistema educativo, por meio das ações desenvolvidas pelas escolas e suas professoras. A preocupação deixa de ser com a aprendizagem, apesar de no discurso oficial ser destacado que essa é a principal justificativa dos três Sistemas criados, e passa a ser o desempenho das escolas nos testes. O processo de ensino perde relevância para o resultado obtido, como se esse refletisse totalmente o primeiro. Esquece-se que os testes, e os itens que os compõem, apresentam limites e não conseguem captar a totalidade da realidade educacional das escolas, de seus alunos e do ambiente que os cercam, impactando, portanto, diretamente os resultados. Nesse movimento, de ampliar a regulação, aumenta-se tanto a quantidade de alunos como o número de anos escolares testados, como pode ser verificado no Quadro 4.1 que cobre o período entre 1990 e 2012.

**Quadro 4.1 – Comparativo entre testes aplicados e tamanho da rede de escolas de ensino fundamental – União, Minas Gerais e São Paulo, 1990 a 2012.**

		União	Minas Gerais	São Paulo
Testes	Periodicidade	bianual <sup>(1)</sup>	anual <sup>(2)</sup>	anual <sup>(3)</sup>
	Abrangência	censitária <sup>(4)</sup>	censitária <sup>(5)</sup>	censitária <sup>(6)</sup>
	Anos escolares <sup>(7)</sup>	2 <sup>o(8)</sup> , 3 <sup>o(9)</sup> , 4 <sup>o(10)</sup> , 5 <sup>o(11)</sup> , 6 <sup>o(12)</sup> , 8 <sup>o(13)</sup> , 9 <sup>o(14)</sup>	2 <sup>o(15)</sup> , 3 <sup>o(16)</sup> , 4 <sup>o(17)</sup> , 5 <sup>o(18)</sup> , 6 <sup>o(19)</sup> , 9 <sup>o(20)</sup>	2 <sup>o(21)</sup> , 3 <sup>o(22)</sup> , 4 <sup>o(23)</sup> , 5 <sup>o(24)</sup> , 6 <sup>o(25)</sup> , 7 <sup>o(26)</sup> , 8 <sup>o(27)</sup> , 9 <sup>o(28)</sup>
	Áreas do conhecimento <sup>(29)</sup>	L, Rd <sup>(30)</sup> , Mt, C <sup>(31)</sup>	L, Rd, Mt, CH <sup>(32)</sup> , CN <sup>(33)</sup>	L, Rd, Mt, C, G <sup>(34)</sup> , H <sup>(35)</sup>
	Quantidade de alunos testados	18,7 milhões	7,3 milhões	69,2 milhões

Elaboração do autor, com base nos dados levantados pela pesquisa.

Notas: (1) não se computou os testes da provinha Brasil que são disponibilizados anualmente desde 2008.

(2) passou a ser anual para todos os anos escolares a partir de 2006.

(3) passou a ser anual para todos os anos escolares a partir de 2003, mas em 2006 não houve aplicação.

(4) passou a ser censitário com a Prova Brasil em 2007.

(5) não é censitária para alguns anos escolares.

(6) só não foi censitária em 2002.

(7) como o foco do trabalho é o ensino fundamental, não são considerados os anos escolares do ensino médio.

(8) só foi testado em 1990 e 1993, voltando a sê-lo em 2008 com a Provinha Brasil, mas como os dados da aplicação não são enviados ao INEP, eles não fazem parte do quadro. A partir de 2013 seus dados passarão a ser enviados ao INEP.

(9) passará a ser testado em 2013.

(10), (12) e (13) foram testados apenas em 1990 e 1993.

(11) e (14) passaram a ser testados a partir de 1995.

(15) começou a ser testado em 2005, sempre de forma amostral

(16) começou a ser testado em 2006.

(17) foi testado em 1992 e 1994; a partir de 2007 voltou a ser testado anualmente de forma amostral

(18) começou a ser testado em 2000

(19) foi testado em 1993 e 1994

(20) foi testado em 1992 e a995; a partir de 2000 voltou a ser testado anualmente.

(21) foi testado entre 2003 e 2005 e depois em 2007.

(22) e (26) foram testados a partir de 2003

(23), (25) e (27) foram testados apenas entre 2003 e 2005.

(24) e (28) foram testados em 1997; voltaram a sê-lo a partir de 2001

(29) os códigos referem-se a: C, Ciências; CH, Ciências Humanas; CN, Ciências da Natureza; G, Geografia; H, História; L, Leitura; Mt, Matemática; Rd, Redação

(30) e (31) as áreas de Redação e Ciências só foram testadas em 1990 e 1993

(32) e (33) as áreas de Ciências da Natureza e Ciências Humanas só foram testadas em 2001.

(34) e (35) as áreas de Geografia e História só foram testadas em 1996 e 1997; a partir de 2009 são testadas em anos alternados.

Analisando-se o Quadro 4.1, percebe-se que, nos Estados de Minas Gerais e São Paulo, os testes são aplicados anualmente. No caso do governo federal, a Prova Brasil é aplicada a cada dois anos e a Provinha Brasil anualmente. Com relação à abrangência, nos três casos, os testes são aplicados de forma censitária para os anos escolares escolhidos para serem testados.

No caso da quantidade de alunos testados entre 1990 e 2012, destaca-se o caso de São Paulo com quase 70 milhões de alunos. Isso se deve principalmente pelo fato de, durante o período entre 2003 e 2005, terem sido testados, de forma censitária, todos os alunos do ensino fundamental, lembrando-se que o ensino fundamental de nove anos só foi implantado no Estado em 2008. Com relação ao governo federal, com a criação da Prova Brasil, a cada ciclo têm sido testados uma média de 4,3 milhões de alunos. Ainda quanto ao governo federal, não foram computados os dois testes da Provinha Brasil encaminhados aos Estados e Municípios, um no início e outro ao final do ano, pois, como os dados até 2012 não eram encaminhados ao Inep, não há como contabilizar o número de alunos testados.

Com relação aos anos escolares testados, percebe-se que, nos três casos, privilegiam-se testar de forma censitária os alunos dos 5º e 9º anos do ensino fundamental. Dessa forma, duplicam-se os esforços, pois ainda que a Prova Brasil, o Proeb e o Saresp utilizem testes compostos por itens diferentes, a população testada é a mesma. Ainda mais, saturam-se os estudantes e as profissionais de educação.

Além desses anos escolares, em Minas Gerais são testados também os alunos dos 2º, 3º e 4º anos. Assim, nesse Estado, nos anos iniciais do ensino fundamental, só não é testado o 1º ano. Em São Paulo, além dos 5º e 9º anos, testam-se também os alunos dos 3º e 7º anos. Com a entrada em cena da Avaliação Nacional da Alfabetização em 2013, também haverá superposição de esforços, pois os alunos do 3º ano do ensino fundamental dos Estados de Minas Gerais e São Paulo serão testados duas vezes, uma pelos testes locais e outra pelo governo federal.

Quanto às áreas de conhecimento testadas, os três casos analisados testam Leitura e Matemática. Além delas, em São Paulo testam-se em anos pares Ciências e nos ímpares História e Geografia, aplicando-se também em todos os anos uma redação. Nesse caso, apesar de ser ampliada a quantidade de áreas de conhecimento testadas, ainda assim continuam de fora dos testes, pelos motivos já discutidos nesse trabalho, uma série de outras habilidades que são fundamentais para o desenvolvimento cognitivo dos alunos e para sua vida em sociedade. O resultado apurado nos testes, portanto, continua a refletir apenas uma parte do trabalho desenvolvido pela escola. Assim, usá-lo para expressar a qualidade da escola em sua totalidade não faz sentido. De outro lado, induz o foco do trabalho escolar nas áreas de Leitura e Matemática, desvalorizando as demais e contribuindo para reduzir o significado da educação formal.

Mas afinal, qual a razão de se coletar essa quantidade enorme de dados? Se o objetivo fosse avaliar políticas educacionais, bastaria apenas aplicar os testes a uma amostra, e mesmo

assim não teria sentido aplicá-los todos os anos já que os efeitos de uma política raramente são sentidos em curto espaço de tempo. Portanto, percebe-se que o objetivo perseguido não é obter informações que permitam contribuir para o processo de aprendizagem dos alunos ou aprimorar as políticas educacionais, mas aumentar a pressão sobre a escola e suas professoras para que os resultados nos testes aumentem a cada ano.

Os testes de Minas Gerais e São Paulo apresentam seus resultados na mesma escala do Saeb, pois utilizam-se procedimentos estatísticos que permitem isso. Apesar de os números que expressam as proficiências poderem ser comparados, não se sabe qual o sentido pedagógico dessa comparação já que as Matrizes utilizadas são diferentes entre si e com relação ao Saeb.

No que se refere à forma de expressar seus resultados, os testes avaliativos dos dois Estados utilizam-se de intervalos ou níveis, que são nomeados para indicar o que seria a proficiência desejável. Assim, em Minas Gerais, as faixas são classificadas como Recomendado, Intermediário e Baixo e em São Paulo como Avançado, Adequado, Básico e Abaixo do Básico. Cada uma dessas faixas apresenta uma definição geral. É interessante comparar a descrição do que é recomendado em Minas Gerais ó são capazes de realizar tarefas que exigem habilidades mais sofisticadas. Eles desenvolvem habilidades esperadas para o período de escolaridade em que se encontram (MINAS GERAIS, 2011a, p. 49) ó com a descrição do que é adequado em São Paulo ó demonstram conhecimentos e domínio dos conteúdos, competências e habilidades requeridas para a série escolar em que se encontram (SÃO PAULO, 2009a, p. 2). São definições bastante genéricas, definidas a partir dos resultados dos testes apenas de Leitura e Matemática e apresentam uma descrição impossível de ser medida pois, quais seriam as habilidades esperadas para a série? Nenhum dos currículos dos dois Estados respondem a essa questão, pois não fazem essa indicação de forma explícita e mesmo que indicassem as habilidades esperadas, deveriam fazê-lo para todas as áreas de conhecimento. Novamente, os níveis nada mais são do que balizamentos para obrigar as escolas a atingirem determinados desempenhos nos testes, com pouca ou nenhuma conexão com a aprendizagem.

No caso de Minas Gerais, os resultados também são expressos na forma de percentuais de alunos em cada nível de proficiência. Os relatórios utilizados para divulgar os resultados dão destaque ao avanço desse percentual em direção do nível recomendado, o último da escala.

Com relação ao Saeb e à Prova Brasil, são divulgados a proficiência média e o percentual de alunos distribuídos por níveis, nove para Leitura e doze para Matemática para o

ensino fundamental. Somente em 2003, os níveis foram agrupados em quatro níveis nominados como "Adequado", "Intermediário", "Crítico" e "Muito Crítico". Esse agrupamento deixou de ser usado nos ciclos seguintes. Não foram dadas, em nenhum momento, explicações sobre as razões que levaram o Inep a utilizar essa classificação e nem para deixar de usá-la.

Também percebe-se que em nenhum dos três casos foi explicitado como os níveis foram definidos, o que não contribui para a transparência dos testes avaliativos e para o caráter ético que deve cercar todo julgamento de valor. Esse é um problema sério, dado os termos da adjetivação utilizada, pois escolas em que as proficiências de seus alunos estão localizadas em níveis nominados como muito crítico, abaixo do básico ou baixo, podem passar a ser conhecidas também por esses adjetivos. Seriam escolas muito críticas, abaixo do básico ou baixas. Uma das possíveis consequências dessa adjetivação é afastar os professores dessas escolas que acabam sendo julgadas como escolas problemáticas. Com isso, se já eram escolas com dificuldades para garantir a aprendizagem de seus alunos, por qual motivo seja, a sua situação fica ainda mais fragilizada. Essa situação contraria a visão de que as mudanças educacionais deveriam estar conectadas com estratégias que privilegiassem a inclusão e a multiplicidade de soluções, já que a educação, de acordo com Sahlberg (2011) não pode ser considerada como um bem individual, mas sim um esforço público servindo a um propósito público e, sendo assim, com qualidade para todos. Nos EUA, a solução adotada em algumas cidades, como Nova York, para problemas como esse, é o fechamento da escola ou a sua transformação em escola *Charter*, como foi discutido nesse texto. Até agora nenhuma dessas duas saídas foram apresentadas aqui no Brasil, mas os gestores devem ficar alertas para os malefícios de alternativas como essa, fartamente documentadas por essa pesquisa.

Os programas que fazem uso dos resultados dos testes escolhidos para análise, dentre aqueles desenvolvidos pelos governos federal e estaduais considerados, têm como característica comum gerar algum tipo de consequência em função das proficiências obtidas. No caso do governo federal, o Plano de Desenvolvimento da Educação (PDE), congrega um conjunto de programas, alguns novos e outros antigos, alguns de caráter obrigatório outros voluntários. O PDE é operacionalizado através do Plano de Metas Compromisso Todos pela Educação, explicitando como as transferências voluntárias dos recursos serão realizadas pela União. O Plano, com 28 metas, impõe uma série de condições aos Entes Federados que aderirem a ele, como a divulgação do Ideb na escola e na comunidade, transformando-o no único indicador da qualidade educacional, apesar de todas as limitações aqui discutidas. Outro programa do governo federal analisado, o Pacto Nacional da Alfabetização na Idade Certa



(PNAIC), que dispõe de R\$ 3,3 bilhões de recursos, prevê a realização de programas de formação continuada a professoras e a realização de um teste, a Avaliação Nacional da Alfabetização (ANA). Do total de recursos, são disponibilizados R\$ 500 milhões para premiar escolas e professoras pelo desempenho de seus alunos nos testes. Assim, mesmo que a quase totalidade das escolas de ensino fundamental sejam mantidas por Estados e Municípios, o governo federal define metas a serem seguidas e estabelece premiações para aqueles que a atingirem, atitude que em nada contribui para o necessário regime de colaboração entre os Entes da Federação.

Em Minas Gerais, o Acordo de Resultados é uma ação do governo estadual que estabelece metas para todas as secretarias e serviços públicos oferecidos, entre eles a educação. Assim, para todas as escolas, são fixadas unilateralmente metas de desempenho baseadas nas proficiências apuradas pelo Simave e pelas notas do Ideb. Aquelas que as cumprirem, são candidatas a receber uma bonificação em dinheiro. Para acompanhar o cumprimento dessas metas é utilizado o Programa de Intervenção Pedagógica, que inicialmente tinha como objetivo garantir a alfabetização até os oito anos de idade, e agora reduziu-se a centralizar na figura do Especialista em Educação Básica, um profissional lotado nas escolas, o desenvolvimento de ações para garantir que as metas sejam atingidas. Além disso, o governo estadual mineiro distribuiu para as escolas, no início de 2013, placas informativas indicando seu Ideb. Dessa forma, todo o sistema se movimenta na direção da melhoria de dois indicadores que não se relacionam integralmente com a aprendizagem dos alunos. A estratégia utilizada para garantir essa melhoria é pressionar as escolas e suas professoras na direção de elevar os indicadores e oferecer uma retribuição em dinheiro para aquelas que atingirem as metas definidas unilateralmente.

Quanto a São Paulo, a secretaria estadual de educação vem, desde 2008, estruturando todas suas ações em função dos resultados do Saresp. Assim, durante o processo de seleção de professoras, diretores, supervisores e quadros das regionais de ensino, são considerados conhecimentos sobre o Saresp e sobre como utilizá-los em suas atividades. Até mesmo no processo de avaliação do estágio probatório das professoras é prevista a medida sobre o quanto elas têm utilizado os resultados dos testes para melhorar suas práticas e o de suas escolas. Junto com essas alterações, também criou um indicador próprio para acompanhar a evolução das proficiências medidas pelo Saresp, o Idesp, que, apesar de apresentar alguns aprimoramentos em relação ao Ideb, utiliza os mesmos fatores, proficiência e fluxo escolar, como base de cálculo. Com base no Idesp, também foram definidas metas para todas as escolas, sujeitas agora à do governo federal, que é amplamente divulgada, e outra do governo

estadual, que serve para o cálculo do pagamento da bonificação. O interessante é que apesar do Saresp testar outras áreas do conhecimento, só entram no cômputo do Idesp as proficiências de Leitura e Matemática. Para organizar os tempos escolares nas escolas, de forma a refletir a centralidade das áreas de Língua Portuguesa e Matemática nos anos iniciais do ensino fundamental, estipulou-se uma diretriz para a organização curricular. Por essa diretriz, nos três primeiros anos do ensino fundamental, 85% do tempo de aula devem ser dedicados às duas áreas testadas e o restante para educação física e artes. Nenhum espaço é garantido para as áreas de História, Geografia e Ciências. Portanto, em São Paulo busca-se a qualidade da educação principalmente através da fixação de metas, organização dos tempos escolares e o pagamento de prêmios em dinheiro, como se essas fossem condições essenciais para garantir a aprendizagem.

Diante a realidade apresentada, nos três casos analisados percebe-se uma visão utilitarista da educação, baseada na busca do cumprimento de metas de desempenho em testes avaliativos e na adoção de medidas pontuais onde a aprendizagem dos alunos não está presente. Dessa forma, os comemorados sucessos nos avanços dos indicadores educacionais devem ser vistos com cautela, pois podem não estar se refletindo na aprendizagem dos alunos. Esse fato é algo que não é devidamente problematizado na mídia escrita, algo causado principalmente pelos gestores públicos que focalizam somente os resultados, e que acaba levando a população a compreender de forma equivocada os indicadores educacionais.

## CONSIDERAÇÕES FINAIS

Este trabalho analisou a centralidade que os testes avaliativos foram ganhando nas ações desenvolvidas pelo governo nacional e pelos governos estaduais de Minas Gerais e São Paulo. Nesse sentido foram discutidos, em uma perspectiva comparada e em um contexto de mudanças quanto ao papel do Estado (regulação pós-burocrática), o desenvolvimento dos testes avaliativos que cada Ente da Federação desenvolveu e alguns programas educacionais que fazem uso dos resultados dos testes, com o objetivo de analisar seu efeito sobre as políticas desenvolvidas.

A partir dos dados sobre a periodicidade, a abrangência, os anos escolares e as áreas do conhecimento testadas, percebe-se a quantidade impressionante de dados gerados pelos testes. Foram aplicados, entre 1990 e 2012, testes a cerca de 95 milhões de alunos de todos os anos escolares do ensino fundamental, com exceção do 1º ano. Apesar de haver maior concentração de dados em Leitura e Matemática, todas as áreas já foram testadas mais de uma vez. Se aos três casos estudados fossem agregados dados relativos aos testes aplicados pelos outros 18 Estados, esses números subiriam ainda mais.

Com esses dados, seria possível construir um panorama detalhado da realidade educacional brasileira. Mas, apesar disso, alguns entraves ainda impedem que eles sejam utilizados para contribuir com a melhoria da qualidade da educação: pouca utilização desses dados em estudos, falta de interesse dos gestores educacionais em aprofundar as análises, raríssimas pesquisas do Inep baseadas nos resultados obtidos, a maior parte delas financiando estudos de pesquisadores externos à instituição ao invés de construir sua autonomia como uma instituição de pesquisa.

Desde os primeiros testes, geraram-se números: de início, indicava-se uma porcentagem de acertos e, depois do uso constante da Teoria de Resposta ao Item (TRI), passou-se a indicar a média de proficiência. Para a escola, esses números isoladamente não têm significado algum, pois o objetivo dela é buscar alternativas para garantir a aprendizagem de seus alunos, e essa aprendizagem depende de outros fatores que não apenas um número que reflete o resultado de um teste, teste esse limitado àquilo que pode ser medido para algumas, não para todas, áreas do conhecimento do currículo escolar e dentro de um

subconjunto de possibilidades para se elaborar um item, cada uma delas trazendo diferentes dificuldades.

Para tentar resolver esse problema, desenvolveram-se técnicas para procurar dar significado pedagógico a esses números. Para isso, tratou-se de interpretar a escala de proficiência, o que resultou em sua divisão em níveis, qualificados por adjetivos, no caso dos testes de Minas Gerais e São Paulo, e para cada um deles uma breve e genérica descrição. Como as professoras e a comunidade escolar como um todo não participaram do processo de construção dos instrumentos e não tiveram acesso aos testes, essa descrição genérica associada a um determinado adjetivo, além de não ajudar a desvelar o significado do número que expressa uma proficiência, pode fornecer informações ambíguas, como ocorreu na discussão sobre as diferenças nas descrições de níveis da Provinha Brasil. Dessa forma, é muito difícil que a expressão do resultado na forma de um número ou de um adjetivo possa ajudar a escola a compreender suas deficiências ou seus méritos, e os resultados dos testes acabam contribuindo muito pouco para melhorar a prática docente.

Estabelecer níveis e nominá-los, apesar de ser uma prática que se torna cada vez mais corriqueira para exprimir os resultados dos testes, tanto no Brasil como fora daqui, é criticável, pois parte de uma divisão arbitrária (BROWN, 2000) e pouco informa sobre o real progresso dos alunos. Além disso, deve-se levar em conta que as proficiências referem-se a uma probabilidade baseada nas respostas dadas a um conjunto de itens, itens esses que têm propriedades suficientes para medir apenas parte das habilidades que deveriam ser desenvolvidas na escola. Portanto, apesar de existir um esforço no sentido de dar um significado pedagógico ao número, ele ainda é insuficiente para traduzi-lo em uma linguagem compreensível para a equipe escolar. Como as professoras não participam da construção dos itens e não têm acesso a todo o conjunto de itens do teste, fica muito difícil associar a descrição à prática escolar e relacioná-la ao currículo.

Uma estratégia bastante utilizada para procurar superar o problema de compreensão dos resultados é a realização de oficinas com especialistas, para auxiliar as equipes escolares em sua interpretação, mas isso esbarra no tamanho das redes. Somente informar como é possível interpretar não significa que no momento de estarem com seus próprios dados em mãos as equipes escolares consigam dar conta de uma tarefa com tamanha complexidade. No caso do governo federal, a situação é mais complexa ainda, e mesmo que se imaginasse uma capacitação para as equipes das Secretarias de Educação dos Estados, que ficariam responsáveis por multiplicá-la, o problema da quantidade de escolas e da dificuldade de elas utilizarem os resultados para melhorar sua prática docente permanece.

Assim, como parece pouco provável que a totalidade das escolas consigam se apropriar dos resultados, apesar dos mais de vinte anos de tentativas para melhor expressá-los, provavelmente o problema esteja com o uso inadequado dos resultados dos testes avaliativos. Isso porque, inicialmente, esses instrumentos foram desenhados para avaliar políticas educacionais, evoluíram para avaliar sistemas educacionais e, sem que se alterassem seus instrumentos nem a forma como eram desenvolvidos, passou-se a querer que as escolas utilizassem seus resultados para orientar seu trabalho pedagógico. Talvez os testes avaliativos possam ser úteis para a comunidade escolar, caso haja ativa participação dela no desenvolvimento dos instrumentos e na análise dos resultados obtidos. Isso foi o que se procurou fazer durante os primeiros ciclos do Programa de Avaliação do Sistema Estadual de Ensino de Minas Gerais, no início dos anos 1990. Como o processo foi interrompido, não foi possível saber se a estratégia utilizada seria suficiente para contribuir para a melhoria da prática docente em Minas.

Apesar de sua pouca efetividade para auxiliar o trabalho da escola, não significa que os testes avaliativos e seus resultados não tenham qualquer utilidade.

Os testes, e principalmente os censitários, podem ser utilizados para mapear regiões onde seria prioritária a ação do governo e, em cada uma delas, as escolas com maiores dificuldades para garantir a aprendizagem dos alunos, sem esquecer, nesse caso, que os testes captam uma parcela muito pequena da realidade. Os questionários respondidos pelos alunos, professoras e diretores podem auxiliar nesse processo de aproximação com a realidade. Mas esses dois instrumentos juntos não seriam suficientes para fornecer um quadro real da situação e seriam necessários novos estudos de caráter qualitativo. Não se identificou, para os três casos analisados, nenhum esforço dos governos para agregar novos dados aos resultados dos testes.

Os resultados dos testes revelam as dificuldades dos alunos em determinadas áreas do conhecimento e em determinadas situações; portanto, os dados podem contribuir para a identificação de alternativas de aprimoramento de habilidades que se mostrem difíceis de ser desenvolvidas. Mas é necessário pensar em como fazer isso juntamente com as professoras. No entanto, o que se observou foram ações no formato de cursos de curta ou de média duração, presenciais ou a distância, para resolver problemas pontuais.

Ressalte-se que os testes têm sido utilizados de forma desenfreada, e sua importância tem sido superestimada. Um único teste aplicado em um dado momento do período escolar não consegue captar as sutilezas de todo o processo de aprendizagem. No máximo, tira uma foto com uma lente de longa distância. E uma foto não é um filme, por não revelar o que está

acontecendo por dentro do processo, nem captar seus movimentos. Por isso, conclusões definitivas e abrangentes com base nos resultados dos testes tornam-se preocupantes, principalmente quando se quer associá-los à qualidade do ensino.

Pode-se também constatar que os três casos estudados partem do princípio de que o *accountability*, baseado nos resultados dos testes, é a forma adequada para conseguir a melhoria da qualidade da escola, uma qualidade definida pelas proficiências apuradas nos testes.

Existe uma forte mensagem, que tem sido veiculada nos últimos anos, mostrando que aumentar o número de testes é o melhor e único caminho para aprimorar as aprendizagens. A prova disso é o aumento crescente do número de Estados que estão testando seus alunos, quase todos repetindo o mesmo formato dos testes usados pelo governo federal e desenvolvendo indicadores baseados no Ideb para estipular metas para suas escolas. Por que repetir os mesmos procedimentos sobre as mesmas escolas? A única justificativa é a de aumentar a pressão em busca de desempenhos maiores.

A decisão do governo federal de testar a totalidade das escolas públicas brasileiras tem pouco contribuído com o regime de colaboração, já que Estados e municípios participam apenas da organização da aplicação e não da construção dos instrumentos, de sua análise e da forma de divulgação de seus resultados. Além disso, os resultados da Prova Brasil têm sido utilizados para ranquear as escolas, tomando por base o Ideb, e alimentar estudos que identificam escolas que, apesar de atenderem uma população de baixo nível socioeconômico, conseguem resultados expressivos, aumentando com isso a crença de que soluções estruturadas e bem gerenciadas são capazes de reverter problemas causados, principalmente, pelas grandes diferenças sociais existentes.

Em nenhum momento, a discussão sobre os resultados apresentados significa uma posição contrária à realização de testes patrocinados pelo governo federal, que, de acordo com a legislação, tem a competência de assegurar o processo nacional de avaliação do rendimento escolar, nem à existência de um indicador para medir os aspectos da qualidade educacional. O que se discute é a utilidade da coleta de dados com abrangência censitária para a formulação de políticas de âmbito nacional.

Outrossim, todas essas ações fazem sentido dentro da perspectiva da regulação e do modelo do Estado Avaliador, em que se busca centralizar ações, ter acesso direto às escolas, fixar metas para cada uma e apontar aquelas que apresentam superação de seus resultados. Busca-se, em nível nacional, responsabilizar a escola e sua equipe, apesar de o governo federal ter poucos mecanismos para ajudá-las a superar seus problemas, pois não tem a

responsabilidade constitucional direta de mantê-las, além de não contratar suas professoras e nem definir seus currículos. De fato, espera-se que a escola se mobilize por si só para superar suas dificuldades. O que tem sido assegurado no âmbito dos programas federais de educação refere-se ao aumento do valor do repasse da redistribuição de recursos federais às escolas com baixo Índice de Desenvolvimento da Educação Básica (Ideb). Ou então, como no caso do Pacto Nacional pela Alfabetização na Idade Certa (Pnaic), dando prêmio em dinheiro às escolas e a suas professoras, a partir do desempenho delas nos testes, igual à ação desenvolvida por muitos governos estaduais, que também distribuem prêmios em função do desempenho nos testes. Mas, novamente, é questionável o sentido de uma ação federal como essa.

No caso das redes de Minas Gerais e São Paulo, a preocupação com a responsabilização é a mesma. Nos dois casos, apesar de o discurso revelar a existência de uma integração dos resultados com uma série de outras ações, como o desenvolvimento de currículos baseados nos testes, a formação de professoras para melhorar o desempenho dos alunos, o material didático de apoio à aprendizagem, na prática, a pesquisa notou o desenvolvimento de ações isoladas, não delineando outras perspectivas que não sejam melhorar as proficiências. Desenvolvem-se ações pontuais, como as diretrizes para a organização curricular da educação básica de São Paulo, delimitando os tempos que devem ser dedicados a cada área do conhecimento, ou, como em Minas Gerais, com o Programa de Intervenção Pedagógica, em que os especialistas em educação básica atuam como fiscais e orientadores, em direção ao cumprimento das metas estabelecidas pela Secretaria.

No caso de São Paulo, a pesquisa constatou maior organicidade, a partir de 2008, em direção à centralidade que assumem os testes, colocando-os, como parte do processo seletivo das professoras, no centro do processo de avaliação do estágio probatório delas, como elemento do perfil de diretores e supervisores escolares. Em Minas Gerais, ações como essas também vêm sendo tomadas, mas, principalmente, com relação às equipes gestoras. Nas duas redes, a bonificação por resultados é prática comum.

A pesquisa notou também a influência exercida por diversas instituições que tiveram ou que têm influências sobre as decisões dos governos. No final dos anos 1980 e começo dos anos 1990, percebe-se a influência do Banco Mundial nas decisões de usar medidas objetivas, como os testes de desempenho, para avaliar programas educacionais nos três estudos de caso apresentados. Mas só essa influência não justifica que o Saeb tenha sobrevivido a tantas mudanças de governo no âmbito federal. O Saeb começa na época da ditadura militar, passa pela redemocratização e alcança o governo Luiz Inácio Lula da Silva, que começa com uma

proposta de rediscutir o instrumento e adequá-lo a um novo governo, ôdiferente do governo anterior, que tinha compromisso com um perfil liberalizante ou mercantilista de educaçãoô (MEC, 2003). E foi no governo Luiz Inácio Lula da Silva que se instituiu a Prova Brasil e o Ideb, como indicador de qualidade da educação. O que acabou prevalecendo foi a crença nos mecanismos de regulação do Estado, como elemento central do processo educacional, algo que ficou transparente no comunicado à comunidade escolar, durante a primeira gestão do governador Mário Covas, em 1995, e que foi sendo implantado gradativamente, com o passar dos anos, em todos os três casos analisados, chegando-se a ponto de, a partir de 2013, o governo federal testar de forma censitária três dos cinco anos da primeira etapa do ensino fundamental.

Também aparecem, ganham visibilidade e influência atores inicialmente ligados ao mundo empresarial, que expandem seus quadros congregando gestores educacionais e pesquisadores. Juntando *expertise*, recursos financeiros e poder de influência, instituições, como o movimento Todos Pela Educação e a Fundação Lemann, entre outras, têm conseguido transformar algumas de suas ideias em ações de governo, com ênfase nos testes avaliativos.

Nos três casos analisados nesta pesquisa, e em outros testes desenvolvidos no Brasil e no exterior, percebe-se grande ênfase em centrar os testes nas áreas de Matemática e Língua Portuguesa, como se, a partir delas, fosse possível outras áreas se desenvolverem autonomamente. O aprendizado não ocorre por si só, e o ôaprender a aprenderô não se faz sozinho. Para que o aluno consiga compreender o mundo a partir de seus olhos e ser independente, o ensino tem de ser amplo e fornecer uma série de conhecimentos e ferramentas que permitam ao aluno estabelecer um considerável número de conexões, possibilitando a ele realizar diferentes interpretações da realidade e também diferentes caminhos para modificá-la. Como querer que, sem isso, o aluno viva e produza em um mundo globalizado, o tema preferido daqueles que defendem a importância de se testar cada vez mais? Para ampliar os horizontes dos alunos, é necessário muito mais que apenas as duas áreas, porque é muito mais difícil estabelecer conexões com algo que não se conhece. No caso de São Paulo, desde 2008, estão sendo testadas, em anos alternados, outras áreas, mas a ênfase continua sendo em Matemática e Língua Portuguesa.

Como o desempenho é o foco das ações dos governos, o próprio direito à educação, envolvendo a garantia de acesso e a permanência para todos, e de mesma qualidade também para todos, independentemente da origem socioeconômica e da localização da escola, acaba se apequenando. Nesses novos tempos de regulação, procura-se transmutar o direito à educação em direito à aprendizagem, reinterpretando-se a Lei de Diretrizes e Bases da



Educação Nacional (LDB) e a própria Constituição, como se a aprendizagem não estivesse garantida no direito à educação. Outrossim, ao se enfatizar o direito à aprendizagem, a regulação se torna ainda mais efetiva, pois, caso os resultados dos testes mostrem que ela não está ocorrendo, fica mais fácil responsabilizar os agentes que seriam os principais culpados pela sua não concretização: as professoras e as escolas.

Além das especificidades aqui discutidas, relativas à utilização dos testes e ao processo de *accountability* instalado, outros aspectos começaram a entrar em cena, todos provavelmente não previstos pelos seus formuladores.

Um dos aspectos é que a busca para melhorar as proficiências tem estimulado os gestores estaduais e municipais, principalmente, a procurarem soluções alternativas, pois, com o destaque que o Ideb ganhou nos meios de comunicação, a capacidade de gestão deles está sendo julgada, o que não é coisa ruim de todo, uma vez que fornece aos eleitores condições mais favoráveis para a escolha dos dirigentes. O problema é que se tem procurado soluções que permitam resultados imediatos. Mas, em educação, as mudanças são lentas, e não existe fórmula mágica ou bala de prata capaz de eliminar todos os problemas de uma única vez e com um único tiro. Essa avidez por soluções rápidas tem estimulado dois fenômenos.

O primeiro é o aumento do uso dos sistemas apostilados de ensino, que, além de prometerem eficiência, pois seu uso, conforme divulgado na mídia, aumentaria a proficiência dos alunos, vêm acompanhados de materiais instrucionais, treinamento para as professoras e cursos de formação continuada, sobre temas relacionados às deficiências evidenciadas pelos testes que acompanham o sistema apostilado. Com isso, estimula-se a criação de um grande mercado educacional de venda de soluções prontas para melhorar a qualidade do ensino. Desenvolve-se, também, o mercado editorial, com livros que visam auxiliar as professoras a conseguir melhores proficiências de seus alunos, e também o mercado para empresas e instituições que desenvolvem e aplicam testes avaliativos, cada vez mais demandados pelos governos e também pelas escolas privadas. Para dar uma dimensão do potencial desse mercado de testes, é bom lembrar que nos EUA ele gira entre US\$ 20 e US\$ 50 bilhões.

O segundo fenômeno está ligado à preparação para o teste. A pesquisa demonstrou que esse fenômeno ocorre devido ao incentivo dos governos e à iniciativa das próprias escolas. O governo federal disponibiliza no *site* do MEC um simulado para a Prova Brasil. Quanto aos governos estaduais, a pesquisa revelou a existência de uma prática de realização de simulados preparatórios, que ocorrem normalmente nos meses anteriores à aplicação da Prova Brasil, acompanhados de treinamento para as professoras de Matemática e Língua Portuguesa, que discutem estratégias para lidar com o baixo rendimento dos alunos. Ficou

claro que a preocupação dessas iniciativas é mais elevar o desempenho do Ideb que garantir a aprendizagem dos alunos. Tanta ênfase nos resultados pode estar levando ao estreitamento dos currículos, adaptando-os às Matrizes de Referência dos testes, algo que as pesquisas americanas já assinalaram (RAVITCH, 2012c; CEP, 2007; PEDULLA *et al.*, 2003) e que precisa ser também pesquisado no Brasil.

No caso das escolas, essa competição por performatividades cada vez maiores, as tem levado a utilizar também a preparação para o teste por meio de simulados. Apesar de a pesquisa detectar poucos relatos, por meio de notícias na mídia escrita, não é improvável que essa estratégia esteja ocorrendo como uma prática corriqueira, mesmo porque notícias sobre escolas individuais são mais difíceis de serem divulgadas. Isso revela o caráter da regulação autônoma, quando cada escola, comparando-se com as demais ou mesmo com seu próprio desempenho, começa um processo de autorregulação e acaba assumindo a responsabilidade que lhe é atribuída, buscando, geralmente de forma individualizada, formas para melhorar seu desempenho e superar suas metas (BARROSO, 2005; OLIVEIRA, 2011). Nesses casos, não importa a aprendizagem, mas sim o desempenho nos testes, o que ainda é mais danoso.

Apesar de tudo levar a crer que essa prática das escolas e de suas redes está se generalizando, a discussão tem que ser relativizada. Isso porque devem existir escolas que conseguem superar as dificuldades impostas pelos instrumentos de divulgação dos resultados dos testes e serem bem-sucedidas, quando os utilizam em benefício da aprendizagem dos alunos, além de redes que conseguem se mobilizar para apoiar suas escolas no processo de aprimoramento da aprendizagem de seus alunos, também utilizando os resultados dos testes. No caso das escolas, é provável que o trabalho seja apoiado em seu projeto pedagógico, num forte espírito de trabalho coletivo que envolva as professoras das áreas que não foram testadas e outras profissionais da escola e, principalmente, com uma equipe de baixa rotatividade, rotatividade essa que, na média, atinge quase 50% na rede estadual, em torno de 70% na rede municipal e impressionantes 100% na rede privada.

O que a pesquisa evidenciou é que com o passar do tempo foi se perdendo a preocupação social e pedagógica da avaliação, sendo substituída pelo ranqueamento com efeitos utilitaristas e que envolve ações como a retirada ou o aumento de determinados conteúdos, o estabelecimento de números de horas dedicadas a cada um dos componentes curriculares ou a colocação de placas para informar resultados. Além disso, a avaliação tem sido utilizada para balizar decisões relativas aos gastos públicos. Pode-se perceber também que, nesse percurso, a avaliação ficou banalizada, reduzindo-se simplesmente à medida. Com

isso, os governos deixam de formular políticas orgânicas, que tratem da educação como um todo, e centram suas ações e programas para melhorar os desempenhos nos testes avaliativos.

A pesquisa contrapõe, também, dois modelos: um, que vem sendo implantado nos EUA e seguido, em parte, por diversos países, o Brasil inclusive, e outro, utilizado na Finlândia. Mesmo sabendo que são realidades culturais, econômicas e sociais completamente distintas, a análise dos dois é importante para indicar que existem alternativas ao modelo dominante. No primeiro caso, aposta-se na competição e no individualismo, fixam-se metas ó pois desconfia-se que o sistema por si só não será capaz de buscar a excelência ó e estabelece-se o ensino e a aprendizagem estandardizados ó o contrário da própria filosofia empresarial que os inspirou, filosofia essa que defende a inovação e a criatividade como estratégias necessárias para competir em um mundo globalizado. No caso da Finlândia, aposta-se na colaboração e na solidariedade, na confiança de que as professoras não são meras técnicas de ensino, mas sim profissionais da educação que compreendem o processo educacional como um todo e que conhecem e dominam as ferramentas para garantir uma aprendizagem significativa. Além disso, aposta-se também no ensino voltado para a realidade de cada aluna e de cada aluno e em um currículo voltado para o desenvolvimento da personalidade do indivíduo, muito além do que o mero ensino estandardizado e centrado em Leitura, Matemática e Ciências.

Nesse contexto, é importante o alerta de Pasi Sahlberg sobre o ògermeõ que se espalha por todo o mundo, que tem como sintomas a competição e a responsabilização das escolas. Esse germe tem minado a educação pública, trazendo como efeitos danosos e indesejados o aumento do ensino focado nos testes e o estreitamento do currículo, e aproximado a Pedagogia de uma instrução mecânica. Espera-se que esta pesquisa ilumine o caminho dos gestores educacionais sobre os efeitos danosos desse germe.

Para compreender melhor os efeitos dos testes avaliativos sobre a realidade educacional brasileira, sugere-se alguns temas de pesquisa: estudos sobre o desenvolvimento dos testes avaliativos no Brasil e os conceitos pedagógicos utilizados inicialmente e como foram sendo alterados com o tempo, a partir de entrevistas com os principais envolvidos nesse processo, já que muitos documentos se perderam com o passar dos anos; uma análise detalhada sobre as quedas nas proficiências médias do Saeb em todos os anos escolares testados entre os anos 1997 e 2001 buscando comprovar, ou não, a tese de que foi causada pela inclusão na escola das camadas mais pobres da população; pesquisar a diminuição do ritmo de aumento da proficiência média com o avanço dos anos escolares, principalmente como observado no Proalfa na passagem do 2º para o 3º ano, mas que também se manifesta

em diferentes testes avaliativos e para diferentes anos escolares; pesquisar o poder de influência exercido pelos movimentos sociais e fundações privadas sobre a definição da política educacional e sua agenda junto aos governos nacional e subnacionais; estudar a influência de empresas privadas, ou instituições como o CAEd e a Cesgranrio, que produzem, aplicam e analisam os resultados de testes avaliativos sobre as decisões dos governos subnacionais em aplicar testes em suas redes; analisar as razões da alta rotatividade das professoras nas escolas, tanto públicas como privadas, e mensurar seu impacto no desempenho dos alunos; realizar estudos qualitativos para mensurar os efeitos da redução dos currículos escolares às Matrizes de Referência dos testes junto às escolas e aos professores.

A principal recomendação que essa pesquisa pode fornecer aos governos é que seja cumprido um dos princípios básicos de qualquer processo avaliativo: a ética. Um processo ético envolve a divulgação de dados e informações sobre todas as etapas do processo, abrangendo os estudos para o desenvolvimento das Matrizes dos testes e da relação entre os níveis de proficiência e os adjetivos utilizados para nomeá-los. Além disso, deve-se indicar os itens que deixaram de ser utilizados para os cálculos das proficiências médias e daqueles que fizeram parte, ou não, das interpretações das escalas. Também é fundamental esclarecer o público em geral, e as escolas e os meios de comunicação em particular, que todas as medidas baseiam-se em cálculos probabilísticos e como tal não podem ser consideradas como números exatos, além de trazerem embutidas erros, que devem ser explicitados. Por fim, é importante esclarecer os limites dos indicadores educacionais baseados nos testes de desempenho e indicar explicitamente que apontam para aspectos limitados da realidade educacional e, portanto, não podem ser vistos como as únicas referências para a qualidade da educação. Para isso, é fundamental que os relatórios técnicos dos testes sejam muito bem elaborados e apresentem todas as informações e justificativas de forma clara e consistente. No caso brasileiro um modelo de relatório que poderia ser seguido e aprimorado é o relatório técnico do Saeb de 2001 (Inep, 2002, 2002a, 2002b, 2002c, 2002d). Outros relatórios que poderiam servir de modelo são os relatórios do Pisa (OCDE, 2011).

## REFERÊNCIAS

- ACHIEVE & THE AMERICAN DIPLOMA PROJECT NETWORK. *Achieve*, 2012.  
Disponível em: <<http://www.achieve.org/files/About%20AchieveADP-Apr2012.pdf>>. Acesso em: 30 ago. 2012.
- AFONSO, Almerindo J. Estado, mercado, comunidades e avaliação: esboço para uma rearticulação crítica. *Educação e Sociedade*, Campinas, 20, n. 69, p. 139-164, 1999.
- \_\_\_\_\_. *Avaliação educacional: regulação e emancipação*. São Paulo: Cortez, 2005.
- ALAVARSE, Ocimar M.; BRAVO, Maria Helena; MACHADO, Cristiane. Avaliação como diretriz das políticas educacionais dos governos federal, estaduais e municipal: o caso brasileiro. In: CONGRESSO IBERO-AMERICANO DE ADMINISTRAÇÃO E POLÍTICA EDUCACIONAL, 3., 2012, Zaragoza, Espanha. Disponível em: <[http://www.anpae.org.br/iberoamericano2012/Trabalhos/OcimarM.Alavarse\\_res\\_int\\_GT3.pdf](http://www.anpae.org.br/iberoamericano2012/Trabalhos/OcimarM.Alavarse_res_int_GT3.pdf)>. Acesso em: 17 maio 2013.
- ALVES, Mary L. B. L. Os alunos da 8ª série do ensino fundamental em Minas Gerais: desempenho em História. *Estudos em Avaliação Educacional*, São Paulo, v., n. 9, p. 127-136, jan. 1994.
- AMREIN, Audrey; BERLINDER, David. High-stakes testing, uncertainty, and student learning. *Educational Policy and Analysis Archives*, v. 10, n. 8, 2002. 1-74. Disponível em: <<http://epaa.asu.edu/ojs/article/viewFile/297/423>>. Acesso em: 8 ago. 2012.
- \_\_\_\_\_; \_\_\_\_\_. Re-analysis of NAEP math and reading scores in states with and without high-stakes tests: response to Rosenshine. *Educational Policy and Analysis Archives*, v. 11, n. 25, august, 2003. Disponível em: <<http://epaa.asu.edu/ojs/article/download/253/379>>. Acessado em: 15 ago. 2012.
- ANTUNES, Ana L. Avaliação da 8ª série do ensino fundamental em Minas Gerais. *Estudos em Avaliação Educacional*, São Paulo, v. 1, n. 9, p. 17-23, jan./jul. 1994.

ANTUNES, Ana. L.; XAVIER, Edir P. C.; FREITAS, Maria Auxiliadora. Avaliação do rendimento do aluno da escola estadual de Minas Gerais: o Ciclo Básico de Alfabetização. *Estudos em Avaliação Educacional*, São Paulo, v. 1, n. 6, p. 29-41, jul. 1992.

ARAUJO, Carlos H.; CONDÉ, Frederico N.; LUZIO, Nildo. Índice de qualidade da educação fundamental (IQE): proposta para discussão. *Revista Brasileira de Estudos Pedagógicos*, Brasília, v. 85, n. 209/210/211, p. 126-131, 2004.

ARROYO, Miguel G. *Currículo, território em disputa*. Petrópolis: Vozes, 2011.

BAKER, Bruce D. Five Ridiculously Reformy Copy & Paste Policies & Why They're Misguided. *School Finance Blog*, 2012a. Disponível em: <[http://schoolfinance101.wordpress.com/2012/06/18/ridic\\_reformy\\_policie](http://schoolfinance101.wordpress.com/2012/06/18/ridic_reformy_policie)>. Acesso em: 30 jul. 2012.

\_\_\_\_\_. Still searching for miracle schools and superguy: updates on Houston and New York City *School Finance Blog*, 2012b. Disponível em: <<http://schoolfinance101.wordpress.com/2012/08/15/still-searching-for-miracle-schools-and-superguy-updates-on-houston-and-new-york-city/?blogsub=confirming#subscribe-blog>>. Acesso em: 30 ago. 2012.

\_\_\_\_\_. What do the available data tell us about NYC charter school teachers & their jobs? *School Finance Blog*, 2012c. Disponível em: <<http://schoolfinance101.wordpress.com/2012/08/28/what-do-the-available-data-tell-us-about-nyc-charter-school-teachers-their-jobs>>. Acesso em: 30 ago. 2012.

\_\_\_\_\_.; SCIARRA, David; FARRIE, Danielle. *Is school funding fair?* A National Report Card. Newark: Educational Law Center, 2012.

BAKER, Frank B. The basics of item response theory. Maryland: ERIC, 2001.

BARROSO, João. O estado, a educação e a regulação de políticas públicas. *Educação & Sociedade*, Campinas, v. 26, n. 92, p. 725-751, out. 2005.

BARROSO, João. Da política baseada no conhecimento às práticas baseadas em evidências. In: OLIVEIRA, D. A.; DUARTE, A. *Políticas públicas e educação: regulação e conhecimento*. Belo Horizonte: Fino Traço, 2011.

BAUER, Adriana. *Usos dos resultados do Saresp: o papel da avaliação nas políticas de formação docente*. Dissertação (Mestrado em Educação) ó Faculdade de Educação, Universidade de São Paulo, São Paulo, 2006.

\_\_\_\_\_. Estudos sobre sistemas de avaliação educacional: um retrato em branco e preto. *Revista @mbieteeducação*, São Paulo, v. 5, n. 1, p. 7-31, jan./jun. 2012.

BEATON, Albert E. e ALLEN, Nancy L. Interpreting scales through scale anchoring. *Journal fo educational statistics*, Washington, DC, v. 17, n. 2, p. 191-204, 1992.

BECALLI, Fernanda Z. *O ensino de leitura No Programa de Formação de Professores Alfabetizadores (Profa)*. Dissertação (Mestrado em Educação) ó Centro de Educação, Universidade Federal do Espírito Santo, Vitória, 2007.

BERADAY, George Z. F. *Método comparado em educação*. São Paulo: Nacional; Ed. USP, 1972.

BITAR, Hélia. *et al.* O sistema de avaliação escolar do Estado de São Paulo: implantação e continuidade. *Idéias*, São Paulo, n. 30, p. 9-20, 1998.

BONAMINO, Alícia C. *Tempos de avaliação educacional: o Saeb, seus agentes, referências e tendências*. Rio de janeiro: Quartet, 2002.

BONAMINO, Alícia C.; SOUZA, Sandra Z. Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola. *Educação e Pesquisa*, São Paulo, 38, n. 2, p. 373-388, 2012.

BOURQUE, Mary L. A history of NAEP achievement levels: issues, implementation, and impact 1989-2009. National Assessment Governing Board, 2009. Disponível em: <<http://www.nagb.org/who-we-are/20-anniversary/bourque-achievement-levels-formatted.pdf>>. Acesso em: 3 jul. 2012.

BRASIL. Projeto de Lei nº 1.258, de 1988. *Diário do Congresso Nacional*, Brasília, 29 nov. 1988. Seção I, p. 4267-4271.

\_\_\_\_\_. Projeto de Lei do Senado nº 67, de 1992. *Diário do Congresso Nacional*, Brasília, 21 maio 1992. Seção I, p. 3702-3727.

\_\_\_\_\_. Redação final do Projeto de Lei nº 1.258-C, de 1988. *Diário do Congresso Nacional*, Brasília, 14 maio 1993. Seção I ó Suplemento ao nº 80.

\_\_\_\_\_. Lei nº 9.394, de 20 de dezembro de 1996. *Diário Oficial da União*, Brasília, 23 dez. 1996, Seção 1, p. 27833-27841.

\_\_\_\_\_. Decreto Presidencial nº 6.094, de 24 de abril de 2007. *Diário Oficial da União*, Brasília, 25 abr. 2007, Seção 1, p. 5.

BROOKE, Nigel e CUNHA, Maria Amália. *A avaliação externa como instrumento da gestão educacional nos Estados*. São Paulo: Fundação Victor Civita, 2011.

BROOKE, Nigel; CUNHA, Maria Amália A.; FALEIROS, Matheus. *A avaliação externa como instrumento da gestão educacional dos estados*. São Paulo: Game/FAE/UFMG, 2011.

BROOKE, Nigel; SOARES, José Francisco. *Pesquisa em eficácia escolar: origens e trajetórias*. Belo Horizonte: Ed. UFMG, 2008.

BROWN, William. Reporting NAEP by achievement levels: an analysis of policy and external reviews. In: BOURQUE, Mary L.; BYRD Sheila. *Student performance standards on the National Assessment of Educational Progress*. Washington: National Assessment Government Board, 2000.

CAMPBELL, David T. Assessing the impact of planned social change. 1976. Disponível em: <<https://www.globalhivmeinfo.org/CapacityBuilding/Occasional%20Papers/08%20Assessing%20the%20Impact%20of%20Planned%20Social%20Change.pdf>>. Acesso em: 12 set. 2012.

CARDOSO, Terezinha G. Os alunos da 8ª série do ensino fundamental em Minas Gerais: desempenho em Geografia. *Estudos em Avaliação Educacional*, São Paulo, v. 1, n. 9, p. 137-146, jan. 1994.

CARNOY, Martin. Medindo os efeitos dos professores sobre os resultados dos alunos com valor agregado. Apresentação oral. In: ENCONTRO TEMÁTICO ABAVE/GERES, 1., 2010, Belo Horizonte. Disponível em: <[http://www.abave.org.br/adm/pdf\\_upload/22SETEMBRO201010.37.24278.pdf](http://www.abave.org.br/adm/pdf_upload/22SETEMBRO201010.37.24278.pdf)>. Acesso em: 30 jul. 2012.



\_\_\_\_\_; LOEB, Susan. Tiene efectos la responsabilidad externa en los indicadores educacionales de los alumnos? Un análisis entre los estados de los EE.UU. PREAL, 2004. Disponível em: <http://m.preal.org/detalhe.asp?det=32>. Acesso em: 6 maio 2012.

CARVALHO, Lilian R. D. S. *Saresp 2005: as vicissitudes da avaliação em uma escola estadual*. Dissertação (Mestrado em Educação) ó Faculdade de Educação, Universidade de São Paulo, São Paulo. 2008.

CARVALHO, Luís M. Governando a educação pelo espelho do perito: uma análise do PISA como instrumento de regulação. *Educação & Sociedade*, Campinas, 30, n. 109, p. 1009-1036, set./dez. 2009.

CONSELHO ESTADUAL DE EDUCAÇÃO DO ESTADO DE SÃO PAULO. Deliberação nº 19, de 19 de dezembro 2001. *Diário Oficial do Estado*, São Paulo, 27 dez 2001, Seção I, p. 23-26.

CENTER ON EDUCATION POLICY. Choices, changes and challenges: curriculum and instruction in the NCLB era. CEP, 2007. Disponível em: <[http://www.cep-dc.org/cfcontent\\_file.cfm?Attachment=McMurrer%5FFullReport%5FCurricAndInstruction%5F072407%2Epdf](http://www.cep-dc.org/cfcontent_file.cfm?Attachment=McMurrer%5FFullReport%5FCurricAndInstruction%5F072407%2Epdf)>. Acesso em: 15 ago. 2012.

CENTER FOR RESEARCH ON EDUCATION OUTCOMES. Multiple choice: charter school performance in 16 states. Stanford, 2009. Disponível em: <[http://credo.stanford.edu/reports/MULTIPLE\\_CHOICE\\_CREDO.pdf](http://credo.stanford.edu/reports/MULTIPLE_CHOICE_CREDO.pdf)>. Acesso em: 30 jul. 2012.

\_\_\_\_\_. Charter school performance in Pennsylvania. Stanford, 2012. Disponível em: <[http://credo.stanford.edu/reports/PA%20State%20Report\\_20110404\\_FINAL.pdf](http://credo.stanford.edu/reports/PA%20State%20Report_20110404_FINAL.pdf)>. Acesso em: 30 jul. 2012.

CENTRO DE SELEÇÃO E PROMOÇÃO DE EVENTOS. *Relatório da Oficina de Interpretação das Escalas – SAEB 2005*. Brasília: Cespe, 2006.

COELHO, Maria Ines. Vinte e cinco anos de avaliação da educação básica no Brasil: aprendizagens e desafios. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, 16, n. 59, p. 229-258, abr./jun. 2008.

COLEMAN, James S. *Equality of Educational Opportunity Study (EEOS)*. Ann Arbor, MI: U.S. Department of Health, Education and Welfare, 1966.

CRONBACH, Lee J. Course improvement through evaluation. In: STUFFLEBEAM, Daniel L., MADAUS, George F., KELAGHAN, Thomas. *Evaluation Model*. New York: Spring, 2002.

CRONIN, John *et al.* The accountability illusion. Washington, DC: Thomas B. Fordham Institute, 2009. Disponível em:  
<[http://www.edexcellencemedia.net/publications/2009/200902\\_accountabilityillusion/2009\\_AccountabilityIllusion\\_WholeReport.pdf](http://www.edexcellencemedia.net/publications/2009/200902_accountabilityillusion/2009_AccountabilityIllusion_WholeReport.pdf)>. Acesso em: 22 nov. 2012.

CURY, Carlos R. J. A educação básica como direito. *Cadernos de Pesquisa*, São Paulo, v. 38, n. 134, maio/ago. 2008.

DANTAS, Lys M. V. *As contribuições das políticas e avaliação em larga escala: o caso da avaliação da aprendizagem na Bahia*. Tese (Doutorado em Educação) ó Faculdade de Educação, Universidade Federal da Bahia, Salvador, 2009.

DARLING-HAMMOND, Linda. *The flat world and education: how America's commitment to equity will determine our future*. New York: Teacher College Press, 2010.

DORN, Sherman. *Accountability Frankenstein: understanding and taming the monster*. Charlotte: Information Age Publishing, 2007.

DUARTE, Natália. *Política social: um estudo sobre educação e pobreza*. Tese (Doutorado em Política Social) 1996 ó Departamento de Serviço Social, Universidade de Brasília, Brasília, 2012.

DURSO, Catherine S. An analysis of the use and validity of test-based teacher evaluations reported by the Los Angeles Times: 2011. *NEPC*, Boulder, 2012. Disponível em:  
<<http://nepc.colorado.edu/files/rb-latimesii.pdf>>. Acesso em: 22 nov. 2012.

ELLIOTT, John. La reforma educativa en el Estado Evaluador. *Perspectivas*, v. XXXII, n. 3, 1-20, set. 2002.

ESCORZA, Tomás E. Desde los tests hasta la investigación evaluativa actual. Um siglo, el XX, de intenso desarrollo de la evaluación em educación. *Revista Electrónica de*

*Investigación y Evaluación Educativa*, 2003. Disponível em:

<[www.uv.es/RELIEVE/v9n1/RELIEVEv9n1\\_1.htm](http://www.uv.es/RELIEVE/v9n1/RELIEVEv9n1_1.htm)>. Acesso em: 15 jun. 2012.

FARIAS, Danielle. *Relatório final individual: análise de reportagens sobre avaliação educacional*. Brasília: TEDis, set. 2012.

FERNANDES, Reynaldo. *Índice de desenvolvimento da educação básica (Ideb)*. Brasília: Inep, 2007.

FERREIRA, Roseli. *O sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (Saresp): uma análise das provas de leitura e escrita da 4ª série do ensino fundamental*. Dissertação (Mestrado em Educação) ó Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista, Presidente Prudente-SP, 2007.

FERRER, Guillermo. *Sistemas de evaluación de aprendizages en América Latina*. Santiago: Preal, 2006.

FONSECA, Marília; TOSCHI, Mirza S.; OLIVEIRA, João F. (Orgs.). *Escolas gerenciadas: planos de desenvolvimento e projetos político-pedagógicos em debate*. Goiânia: Universidade Católica de Goiás, 2004.

FRANCO, Creso. O Saeb ó Sistema de Avaliação da Educação Básica: potencialidades, problemas e soluções. *Revista Brasileira de Educação*, v. 1, n. 17, p. 127-133, 2001.

FREITAS, Dirce N. T. *A avaliação da educação básica no Brasil: dimensão normativa, pedagógica e educativa*. Campinas: Autores Associados, 2007.

FREITAS, Luiz C. (Org.). *Avaliação: construindo o campo e a crítica*. Florianópolis: Insular, 2002.

FREITAS, Luiz C. Eliminação adiada: o caso das classes populares no interior da escola e a ocultação da (má) qualidade do ensino. *Educação e Sociedade*, Campinas, v. 28, n. 100, p. 965-987, out. 2007.

FREITAS, Luiz C. A caixa preta da avaliação, 13 mar. 2013. Disponível em:

<[www.educacaoepesquisa.blog.br/?p=26](http://www.educacaoepesquisa.blog.br/?p=26)>. Acesso em: 20 mar. 2013.

FRIEDMAN, Thomas L. Average is over, part II. *New York Times*, 2012. Disponível em: <[www.nytimes.com/2012/08/08/opinion/friedman-average-is-over-part-ii-.html](http://www.nytimes.com/2012/08/08/opinion/friedman-average-is-over-part-ii-.html)>. Acesso em: 14 ago. 2012.

FULBECK, Eleanor. Beyond anecdotes: the evidence about financial incentives and teacher retention. *Shanker Blog*, 2012. Disponível em: <<http://shankerblog.org/?p=4729>>. Acesso em: 15 ago. 2012.

FUNDAÇÃO GETÚLIO VARGAS; FUNDAÇÃO LEMANN. O impacto dos métodos estruturados de ensino na proficiência dos alunos da rede pública municipal do Estado de São Paulo. Centro de Políticas Sociais do Instituto Brasileiro de Economia, Fundação Getúlio Vargas. São Paulo, abril 2009. Seminário Interno. Disponível em: <[http://www.cps.fgv.br/ibrecps/rede/seminario/pdf/Maria\\_Carolina\\_v.pdf](http://www.cps.fgv.br/ibrecps/rede/seminario/pdf/Maria_Carolina_v.pdf)>. Acesso em: 7 maio 2013.

FUNDAÇÃO LEMANN; FUNDAÇÃO GETÚLIO VARGAS. Sistemas estruturados de ensino e redes municipais do Estado de São Paulo. Seminário Interno. Fundação Lemann. São Paulo, dezembro de 2008. Disponível em: <[http://epoca.globo.com/edic/648/648\\_Sistemas\\_Ensino.pdf](http://epoca.globo.com/edic/648/648_Sistemas_Ensino.pdf)>. Acesso em: 7 maio 2013.

FUNDAÇÃO LEMANN; ITAÚ BBA. *Excelência com equidade: as lições das escolas brasileiras que oferecem educação de qualidade a alunos de baixo nível socioeconômico*. São Paulo: Fundação Lemann, 2012.

FUNDAÇÃO VICTOR CIVITA. Alunos em primeiro lugar: como Nova York renovou seu sistema público de ensino. 2011. Disponível em: <[http://www.fundacaoitausocial.org.br/\\_arquivosstaticos/FIS/pdf/especial\\_NY.pdf](http://www.fundacaoitausocial.org.br/_arquivosstaticos/FIS/pdf/especial_NY.pdf)>. Acesso em: 14 Mai 2013.

FUNDO DAS NAÇÕES UNIDAS PARA A INFÂNCIA; INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS; MINISTÉRIO DA EDUCAÇÃO; UNIÃO DOS DIRIGENTES MUNICIPAIS DE EDUCAÇÃO. *Redes de aprendizagem: boas práticas de municípios que garantem o direito de aprender*. Brasília: UNICEF, 2008.

GALL, Norman; GUEDES, Patricia M. A reforma de Nova York: possibilidades para o Brasil., 2009. Disponível em:

<[http://www.fundacaoitaisocial.org.br/\\_arquivosstaticos/FIS/pdf/reforma\\_educacional\\_ny.pdf](http://www.fundacaoitaisocial.org.br/_arquivosstaticos/FIS/pdf/reforma_educacional_ny.pdf)>. Acesso em: 14 Mai 2013.

GARRIDO, Jose Luis G. *Fundamentos de educación comparada*. Madrid: Dykinson, 1996.

GATTI, Bernadete A. Avaliação educacional no Brasil: experiências, problemas, recomendações. *Estudos em Avaliação Educacional*, São Paulo, p. 67-80, jul./dez. 1994.

\_\_\_\_\_. Desenvolvimento de projetos de avaliação educacional no Estado de São Paulo. *Estudos em Avaliação Educacional*, São Paulo, v. 1, n. 13, p. 19-26, jan./jun. 1996.

\_\_\_\_\_. Avaliação educacional: pontuando uma história de ações. *EccoS revista científica*, São Paulo, v. 4, n. 1, p. 17-41, jun. 2002.

\_\_\_\_\_. Avaliação de sistemas educacionais no Brasil. *Sísifo, Revista de Ciências da Educação*, Lisboa, v. 1, n. 9, p. 7-18, maio/ago. 2009.

\_\_\_\_\_.; VIANNA, Heraldo M.; DAVIS, Claudia. Problemas e impasses da avaliação de projetos e sistemas educacionais: dois estudos de caso. *Avaliação Educacional*, São Paulo, v. 1, n. 4, p. 7-26, jul./dez. 1991.

GOHN, Maria da Glória. Movimentos sociais na contemporaneidade. *Revista Brasileira de Educação*, Rio de Janeiro, v. 16, n. 17, p. 333-361, maio/ago. 2011.

GOULART, Iris B. O Ciclo Básico de Alfabetização: lições de uma avaliação. *Estudos em Avaliação Educacional*, São Paulo, n. 6, p. 3-8, jan./dez. 1992.

GRILLO, Jerry. One size does not fit all. *Georgia Trend Magazine*. Atlanta, 2011.  
Disponível em: <<http://www.georgiatrend.com/December-2011/One-Size-Does-Not-Fit-All>>.  
Acesso em: 30 jul. 2012.

GUIA NETO, Walfrido S. M. Educação e desenvolvimento: conscientização ó vontade política ó participação. *Estudos em Avaliação Educacional*, São Paulo, v. 1, n. 9, p. 9-16, jan./jun. 1994.

HALADYNA, Thomas M. *Writing test items to evaluate higher order thinking*. Boston: Allyn and Bacon, 1997.

\_\_\_\_\_. *Developing and validating multiple-choice test items*. Mahwah: Lawrence Erlbaum Associates, Publishers, 2004.

HANUSHECK, Eric A.; KAIN, Jonh F.; RIVKIN, Steven G. Do higher salaries buy better teachers? *National Bureau of Economic Research*, 1999. Disponível em: <<http://www.nber.org/papers/w7082.pdf>>. Acesso em: 24 abr. 2013.

\_\_\_\_\_. et al. Globally challenged: are U.S. students ready to compete? Cambridge: Harvard Kennedy School, 2011. Disponível em: <[http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG11-03\\_GloballyChallenged.pdf](http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG11-03_GloballyChallenged.pdf)>. Acesso em: 30 jul. 2012.

\_\_\_\_\_; PETERSON, Paul E.; WOESSMANN, Ludger Achievement growth: international and US trends in student performance. *Harvard Kennedy School*, 2012. Disponível em: <[http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG12-03\\_CatchingUp.pdf](http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG12-03_CatchingUp.pdf)>. Acesso em: 30 jul. 2012.

\_\_\_\_\_; RAYMOND, Margaret. Improving educational quality: how best to evaluate our schools? *Education in the 21st Century: Meeting the Challenges of a Changing World. Stanford University*, 2003. Disponível em: <<http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%2BRaymond%202003%20Educ21stCent.pdf>>. Acesso em: 15 ago. 2012.

HANUSHEK, Eric A.; WOESSMANN, L. Education quality and economic growth. World Bank, 2007. Disponível em: <[http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079934475/Edu\\_Quality\\_Economic\\_Growth.pdf](http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079934475/Edu_Quality_Economic_Growth.pdf)>. Acesso em: 30 jul. 2012.

HIATT, D. B. An interview with Ralph Tyler: no limit to the possibilities. *The Phi Delta Kappan*, Bloomington, v. 75, n. 10, Jun 1994. 786-789.

HOFFMANN, B. *The tyranny of testing*. republicação do original de 1962. Mineola: Dover Publications, 2003.

HORTA NETO, João Luiz. *Avaliação externa: a utilização dos resultados do Saeb 2003 na gestão do sistema público de ensino fundamental no Distrito Federal*. Dissertação (Mestrado em Educação) ó Faculdade de Educação, Universidade de Brasília, Brasília, 2006.

\_\_\_\_\_. Um olhar retrospectivo sobre a avaliação externa no Brasil: das primeiras medições em educação até o Saeb de 2005. *Revista Ibero-Americana de Educación*, Madrid, abr. 2007. Disponível em: <[www.rieoei.org/deloslectores/1533Horta.pdf](http://www.rieoei.org/deloslectores/1533Horta.pdf)>. Acesso em: 6 maio 2012.

\_\_\_\_\_. Avaliação externa de escolas e sistemas: questões presentes no debate sobre o tema. *Revista Brasileira de Estudos Pedagógicos*, Brasília, 91, n. 227, p. 84-104, jan./abr. 2010.

\_\_\_\_\_. Os desafios da utilização dos resultados de avaliações nacionais para o desenvolvimento de políticas educacionais por um governo subnacional no Brasil. *Revista Iberoamericana de Educación*, Madrid/Buenos Aires, v. 1, n. 53, p. 65-82, maio/ago., 2010a. Disponível em: <<http://www.rieoei.org/rie53a03.pdf>>. Acessado em 6 maio 2012.

\_\_\_\_\_. Educação como um direito e o Programa Nacional de Alimentação Escolar. In: SIMPÓSIO BRASILEIRO, 25; CONGRESSO IBERO-AMERICANO DE POLÍTICA E ADMINISTRAÇÃO DA EDUCAÇÃO, 2., 2011, São Paulo. Políticas Públicas e Gestão da Educação: construção histórica, debates contemporâneos e novas perspectivas. São Paulo: Anpae, 2011. Disponível em <<http://www.anpae.org.br/simposio2011/cdrom2011/PDFs/trabalhosCompleto/comunicacoesRelatos/0256.pdf>>. Acesso em 13 mai 2013.

\_\_\_\_\_. A avaliação educacional cada vez mais distanciada do processo educacional. In: CONGRESSO IBERO-AMERICANO DE ADMINISTRAÇÃO E POLÍTICA EDUCACIONAL, 3., 2012, Zaragoza, Espanha, 2012. Disponível em: <[www.anpae.org.br/iberoamericano2012/Trabalhos/JoãoLuizHortaNeto\\_int\\_GT2.pdf](http://www.anpae.org.br/iberoamericano2012/Trabalhos/JoãoLuizHortaNeto_int_GT2.pdf)>. Acesso em: 13 mar. 2013.

\_\_\_\_\_. Avaliação sem impacto na aprendizagem. In: YANNOULAS, Silvia C. *Política educacional e pobreza: múltiplas abordagens para uma relação multideterminada*. Brasília: Liber Livros, 2013. p. 111-135.

\_\_\_\_\_; YANNOULAS, Silvia C. A emergência da avaliação externa como organizadora da política educacional. In: XII Encontro Estadual da Anpae-SP, 12.; Encontro Regional da ANPAE Sudeste, 8., 2012, Campinas-SP. Gestão de sistemas, redes de ensino e de escolas: desafios para o campo de pesquisa e para os profissionais da educação. São Paulo: Anpae, 2012. p. 166-180.

INGERSOLL, Richard M. Teacher turnover and teacher shortages: an organization analysis. *American Educational Research Journal*, Washington, DC, v. 38, n. 3, p. 499-534, 2001.

INSTITUTO BRASILEIRO DE OPINIÃO PÚBLICA E ESTATÍSTICA; FUNDAÇÃO SM; TODOS PELA EDUCAÇÃO. A educação na agenda do próximo governo. São Paulo, maio 2010. Disponível em: <<http://www.cenpec.org.br/biblioteca/educacao/estudos-e-pesquisas/a-educacao-na-agenda-do-proximo-governo>>. Acesso em: 7 maio 2013.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS. *Relatório do Sistema Nacional de Avaliação da Educação Básica: ciclo 1990*. Brasília: Inep, 1992.

\_\_\_\_\_. *Relatório do 2º ciclo do Sistema Nacional de Avaliação da Educação Básica ó Saeb 1993*. Brasília. 1995.

\_\_\_\_\_. *Estudo comparativo dos resultados do Saeb 1995/1997*. Brasília: Inep, 1997.

\_\_\_\_\_. *Saeb 97: primeiros resultados*. Brasília: Inep, 1997a.

\_\_\_\_\_. *Relatório final Saeb/95*. Brasília: Inep, 1998.

\_\_\_\_\_. *Matrizes curriculares de referência para o Saeb*. Brasília: Inep, 1998a.

\_\_\_\_\_. *Resultados do Saeb/ 99: Relatório*. Brasília: Inep, 2000.

\_\_\_\_\_. *Saeb 2001: novas perspectivas*. Brasília: Inep, 2002.

\_\_\_\_\_. *Saeb 2001: relatório nacional*. Brasília: Inep, 2002a.

\_\_\_\_\_. *Saeb 2001: relatório Matemática*. Brasília: Inep, 2002b.

\_\_\_\_\_. *Saeb 2001: relatório Língua Portuguesa*. Brasília: Inep, 2002c.

\_\_\_\_\_. *Saeb 2001 ó Relatório de Intrepretação de Escalas*. Inep: Brasília, 2002d.

\_\_\_\_\_. *Relatório nacional Saeb 2003*. Brasília: Inep, 2006.

\_\_\_\_\_. *Perfil do magistério da educação básica ó Censo do professor 2007*. Brasília: Inep, 2007.

\_\_\_\_\_. *Provinha Brasil: Guia de correção e interpretação dos resultados ó Teste 1, Primeiro semestre 2008*. Brasília: Inep, 2008a.



\_\_\_\_\_. *Provinha Brasil: Guia de correção e interpretação de resultados ó Teste 2, segundo semestre 2008*. Brasília: Inep, 2008b.

\_\_\_\_\_. *Provinha Brasil: Guia de correção e interpretação de resultados ó Teste 1, primeiro semestre 2009*. Brasília: Inep, 2009a.

\_\_\_\_\_. *Provinha Brasil: Guia de correção e interpretação de resultados ó Teste 2, segundo semestre 2009*. Brasília: Inep, 2009b.

\_\_\_\_\_. *Guia de elaboração e revisão de itens*. Brasília: Inep, 2010. Disponível em: <[http://download.inep.gov.br/outras\\_acoes/bni/guia/guia\\_elaboracao.pdf](http://download.inep.gov.br/outras_acoes/bni/guia/guia_elaboracao.pdf)>. Acesso em: 5 maio 2013.

\_\_\_\_\_. *Saeb/Prova Brasil 2011 ó primeiros resultados*. 2012. Disponível em: <[http://download.inep.gov.br/educacao\\_basica/prova\\_brasil\\_saeb/resultados/2012/Saeb\\_2011\\_primeiros\\_resultados\\_site\\_Inep.pdf](http://download.inep.gov.br/educacao_basica/prova_brasil_saeb/resultados/2012/Saeb_2011_primeiros_resultados_site_Inep.pdf)>. Acesso em: 10 mar. 2012.

\_\_\_\_\_. Nota informativa do Ideb 2011, 2012a. Disponível em: <[http://download.inep.gov.br/educacao\\_basica/portal\\_ideb/o\\_que\\_e\\_o\\_ideb/nota\\_informativa\\_ideb\\_2011.pdf](http://download.inep.gov.br/educacao_basica/portal_ideb/o_que_e_o_ideb/nota_informativa_ideb_2011.pdf)>. Acesso em: 5 maio 2013.

KANE, Thomas; STAIGER, Douglas. Estimating teacher impacts on student achievement: an experimental evaluation. NBER Working Paper nº 14607. National Bureau of Economic Research. 2009. Disponível em: <<http://www.nber.org/papers/w14607>>. Acesso em: 15 ago. 2012.

KIM, Kyung H. Explaining the Decline of Creativity in American Children: A Reply to Readers, 2010. Disponível em: <<http://www.britannica.com/blogs/2010/12/explaining-the-decline-of-creativity-in-american-children-a-reply-to-readers/>>. Acesso em: 15 Mai 2013.

KLEIN, Ruben. Utilização da Teoria da Resposta ao item no sistema de Avaliação da educação Básica (Saeb). *Ensaio*, Rio de Janeiro, v. 11, n. 40, p. 283-296, jan./set. 2003.

KORETZ, Daniel. *Measuring up: what educational testing really tell us*. Cambridge: Harvard University Press, 2008.

KORTE, Robert C.; NASH, Beverly E. *Review of the Sistema Nacional de Avaliação da Educação Básica (Saeb)*. Brasília: Inep, 1995.

LIJPHART, Arend. Comparative Politics and the Comparative Method. *American Political Science Review*, v. 65, n. 3, p. 683-93, 1971.

LUCIO, Elizabeth O. *Tecendo os fios da rede: o Programa Pró-Letramento e a tutoria na formação continuada de professores alfabetizadores da educação básica*. Dissertação (Mestrado em Educação) ó Faculdade de Educação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.

LUCKESI, Cipriano C. *Avaliação da aprendizagem escolar*. São Paulo: Cortez, 2009.

MACHADO, Cristiane. *Avaliar as escolas estaduais para quê? Uma análise do uso dos resultados do Saresp 2000*. Tese (Doutorado em Educação) ó Faculdade de Educação, Universidade de São Paulo, São Paulo. 2003.

MACHADO, Magali E. *A contribuição e os reflexos do sistema de avaliação básica na sala de aula: o caso do Sistema Mineiro de Avaliação da Educação Pública (Simave)*. Dissertação (Mestrado em Educação) ó Faculdade de Educação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.

MARGOLIS, Jesse. O quanto estamos aprendendo?, 2010. Disponível em: <<http://pt.braudel.org.br/eventos/arquivos/downloads/o-quanto-estamos-aprendendo.pdf>>. Acesso em: 14 Mai 2013.

MAROY, Cristian. Em direção a uma regulação pós-burocrática dos sistemas de ensino na Europa? In: OLIVEIRA, D. A.; DUARTE, A. (Org.). *Políticas públicas e educação: regulação e conhecimento*. Belo Horizonte: Fino Traço, 2011. p. 19-46.

MAROY, Cristian. Regulação dos sistemas educacionais. In: ZANTEN, A. V. *Dicionário de educação*. Petrópolis: Vozes, 2011a.

MENDES, Maria Helena. Textos produzidos pelos alunos do CBA em Minas Gerais ó 1992: análise qualitativa. *Estudos em Avaliação Educacional*, São Paulo, n. 6, Jul 1992. 95-136.

MINISTÉRIO DA EDUCAÇÃO. *Dados preliminares do Saeb/95*. Brasília: Inep, 1996.

\_\_\_\_\_. Seminário discute políticas de avaliação da educação brasileira. *Universia Notícias*, São Paulo, 3 de abril de 2003. Disponível em: <<http://noticias.universia.com.br/vida-universitaria/noticia/2003/04/03/532909/seminario-discute-politicas-avaliacao-da-educacao-brasileira.html>>. Acesso em: 25 abr. 2013.

\_\_\_\_\_. Toda Criança Aprendendo. *Revista Brasileira de Estudos Pedagógicos*, Brasília, v. 84, n. 206/207/208p. 197-201, jan./dez 2003a.

\_\_\_\_\_. *Sistema Nacional de Formação Continuada e Certificação de Professores: Matrizes de Referência ó anos iniciais do ensino fundamental*. Brasília: MEC, 2003b.

\_\_\_\_\_. *O Plano de Desenvolvimento da Educação: razões, princípios e programas*. Brasília: MEC, 2011. Disponível em: <<http://portal.mec.gov.br/arquivos/livro>>. Acesso em: 16 abr. 2013.

\_\_\_\_\_. *Pacto Nacional pela Alfabetização na Idade Certa: avaliação no ciclo de alfabetização ó reflexões e sugestões*. Brasília: MEC, 2012.

\_\_\_\_\_. *Pacto Nacional pela Alfabetização na Idade Certa: currículo na alfabetização ó concepções e princípios*. Brasília: MEC, 2012a.

\_\_\_\_\_. *Pacto Nacional pela Alfabetização na Idade Certa: currículo no ciclo de alfabetização ó consolidação e monitoramento do processo de ensino e de aprendizagem*. Brasília: MEC, 2012b.

\_\_\_\_\_. *Pacto Nacional pela Alfabetização na Idade Certa: currículo inclusivo ó o direito à aprendizagem*. Brasília: MEC, 2012c.

\_\_\_\_\_. Elementos conceituais e metodológicos para a definição dos direitos de aprendizagem e desenvolvimento do ciclo de alfabetização (1º, 2º e 3º anos) do ensino fundamental. MEC, 2012d. Disponível em: <[http://portal.mec.gov.br/index.php?option=com\\_docman&task=doc\\_download&gid=12827&Itemid=>](http://portal.mec.gov.br/index.php?option=com_docman&task=doc_download&gid=12827&Itemid=>)>. Acesso em: 13 maio 2013.

\_\_\_\_\_; INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS; FUNDO DAS NAÇÕES UNIDAS PARA A INFÂNCIA. *Aprova Brasil: o direito de aprender*. Brasília: Unicef, 2007.

\_\_\_\_\_;\_\_\_\_\_;\_\_\_\_\_. *Aprova Brasil: o direito de aprender ó boas práticas em escolas públicas avaliadas pela Prova Brasil*. Brasília: Inep, 2007.

MINAS GERAIS. *Resolução SEE n° 104*, de 14 de julho de 2000. Belo Horizonte. 2000.

- \_\_\_\_\_. Escola Sagarana: educação para a vida com dignidade e esperança. Secretaria de Estado da Educação: Belo Horizonte. 2001.
- \_\_\_\_\_. *Proalfa 2008*: Boletim de resultados. Secretaria de Estado da Educação. Juiz de Fora. 2008.
- \_\_\_\_\_. *Proalfa 2009*: Boletim de resultados. Juiz de Fora: Secretaria de Estado da Educação, 2009.
- \_\_\_\_\_. *Simave/Proeb 2010* Juiz de Fora: Secretaria de Estado da Educação: Juiz de Fora, 2010.
- \_\_\_\_\_. *Simave/Proeb 2010*. V. 1. Juiz de Fora: Secretaria de Estado da Educação, 2010a.
- \_\_\_\_\_. *Proeb 2010*. V. 2: Boletim de Resultados. Secretaria de Estado de Minas Gerais. Juiz de Fora, 2010b.
- \_\_\_\_\_. *Proeb 2010*. V. 3: Boletim Pedagógico da Escola. Juiz de Fora: Secretaria de Estado da Educação, 2010c.
- \_\_\_\_\_. *Guia de orientação para a reorganização e implementação do Plano de Intervenção Pedagógica 2010/2011*. Belo Horizonte: Secretaria de Estado da Educação, 2010d.
- \_\_\_\_\_. Governo de Minas divulga resultados de 2011 do programa de avaliação de desempenho dos alunos da rede estadual. *Agência Minas*, 3 de abril de 2011. Disponível em: <<http://www.agenciaminas.mg.gov.br/noticias/governo-de-minas-divulga-resultados-de-2011-do-programa-de-avaliacao-de-desempenho-dos-alunos-da-rede-estadual/>>. Acesso em: 10 mar. 2013.
- \_\_\_\_\_. *Simave 2011*: revista do sistema. Belo Horizonte: Secretaria de Estado da Educação, 2011a.
- \_\_\_\_\_. *Proalfa 2011*: Revista pedagógica ó 3º ano do ensino fundamental. Belo Horizonte: Secretaria de Estado da Educação, 2011b.
- \_\_\_\_\_. *Proalfa 2011*: Revista do gestor. Juiz de Fora: Secretaria de Estado da Educação, 2011c.
- \_\_\_\_\_. *Guia do Especialista em Educação Básica*. Belo Horizonte: Secretaria de Estado da Educação, s/d.

NATIONAL CONFERENCE OF STATE LEGISLATURES. Parente trigger in the States. Washington, DC: NCSL, 2012. Disponível em: <<http://www.ncsl.org/issues-research/educ/state-parent-trigger-laws.aspx>>. Acesso em: 30 jul. 2012.

NATIONAL COMMISSION ON TEACHING AND AMERICA'S FUTURE. No dream denied: a pledge to America's children. Washington, DC: NCTAF, 2003. Disponível em: <[http://nctaf.org/wp-content/uploads/2012/01/no-dream-denied\\_summary\\_report.pdf](http://nctaf.org/wp-content/uploads/2012/01/no-dream-denied_summary_report.pdf)>. Acesso em: 15 ago. 2012.

NEAVE, Guy. On the cultivation of quality, efficiency and enterprise: an overview of recent trends in higher education in Western Europe, 1986-1988. *European Journal of Education*, v. 23, n. 1/2, p. 7-23, 1988.

NOGUEIRA, Maria Alice. A sociologia da educação no final dos anos 60/início dos anos 70: o nascimento do paradigma da reprodução. *Em Aberto*, Brasília, v. 9, n. 46, p. 49-58, abr./jun. 1990.

NOWAKOWSKI, Jeri R. On education evaluation: a conversation with Ralph Tyler. 1981. Disponível em: <<https://www.globalhivmeinfo.org/CapacityBuilding/Occasional%20Papers/13%20An%20Interview%20with%20Ralph%20Tyler.pdf>>. Acesso em: 3 jul. 2012.

OLIVEIRA, Ana P. M. *A Prova Brasil como política de regulação da rede pública do Distrito Federal*. Dissertação (Mestrado em Educação) ó Faculdade de Educação, Universidade de Brasília, Brasília. 2011.

OLIVEIRA, Dalila A. *Educação básica: gestão do trabalho e da pobreza*. Petrópolis: Vozes, 2000.

OLIVEIRA, Dalila A. Regulação das políticas educacionais na América Latina e suas consequências para os trabalhadores docentes. *Educação & Sociedade*, Campinas-SP, p. 763-775, 2005.

OLIVEIRA, Regina T. C. *A LDB e o contexto nacional: o papel dos partidos políticos na elaboração dos projetos ó 1988 a 1996*. Tese (doutorado em Educação) ó Faculdade de Educação, Universidade Estadual de Campinas, Campinas-SP. 1997.

OLIVEIRA, Romualdo P.; ARAUJO, G. C. Qualidade do ensino: uma nova dimensão da luta pelo direito à educação. *Revista Brasileira de Educação*, São Paulo, n. 28, jan./fev./março/abr. 2005.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO. *Pisa 2009 at a glance*. OCDE, 2011. Disponível em: <<http://www.oecd.org/pisa/46660259.pdf>>. Acesso em: 30 jul. 2012.

PASQUALI, Luiz. *Psicometria: teoria e aplicações*. Brasília: UnB, 1998.

PAYNE, Charles. *So much reform, so little change: the persistence of failures in urban schools*. Cambridge: Harvard Education Press, 2010.

PEARSON FOUNDATION, COUNCIL OF CHIEF STATE SCHOOL OFFICERS. International conference on teacher quality and student support. Pearson Foundation, 2009. Disponível em: <[http://www.pearsonfoundation.org/downloads/CCSSO\\_Helsinki2009\\_report.pdf](http://www.pearsonfoundation.org/downloads/CCSSO_Helsinki2009_report.pdf)>. Acesso em: 5 maio 2013.

PEDULLA, Joseph J. et al. Perceived effects of state-mandated testing programs on teaching and learning: findings from a national survey of teachers. *National Board on Testing and Public Policy*, 2003. Disponível em: <<http://www.bc.edu/research/nbetpp/statements/nbr2.pdf>>. Acesso em: 15 ago. 2012.

PESTANA, Maria Ines. O Sistema de avaliação da educação básica. *Estudos em Avaliação Educacional*, São Paulo, v. 1, n. 5, p. 81-83, jan./jun. 1992.

PETERS, Guy. *Comparative Politics: theory and methods*. New York: New York University Press, 1998.

PETRILLI, Michael. NCBL2: The carrot that feels like a stick. Thomas Fordham Institute, 2009. Disponível em: <<http://www.educationgadfly.net/flypaper/2009/07/the-race-to-the-top-the-carrot-that-feels-like-a-stick/>>. Acesso em: 3 jul. 2012.

PHAM, Vinn H. Computer Modeling of the Instructionally Insensitive Nature of the Texas Assessment of Knowledge and Skills (TAKS) Exam. *University of Texas at Austin*, 2009. Disponível em:

<<http://repositories.lib.utexas.edu/bitstream/handle/2152/10602/phamv53572.pdf?sequence=2>>. Acesso em: 30 jul. 2012.

PILATTI, Orlando. *Saeb – Sistema Nacional de Avaliação da Educação Básica no Brasil: análise e proposições*. Brasília. 1995.

RAGIN, Charles. *Constructing social research: the unity and diversity of method*. Thousand Oaks, CA: Pine Forge Press, 1994.

RAVELA, Pedro. *Para compreender las evaluaciones educativas: fichas didacticas*. Santiago: Preal, 2005.

RAVITCH, Diane. *The death and life of the great american school system: how testing and choice are undermining education*. New York: Basic Books, 2010.

\_\_\_\_\_. Federal study: charters and special education. *Diane Ravitch's Blog*, 2012b. Disponível em: <<http://wp.me/p2odLa-t5>>. Acesso em: 30 jul. 2012.

\_\_\_\_\_. Credibility of Florida's accountability system near collapse. *Diane Ravitch's Blog*, 2012c. Disponível em: <<http://dianeravitch.net/2012/07/19/credibility-of-floridas-accountability-system-near-collapse/>>. Acesso em: 30 jul. 2012.

\_\_\_\_\_. Can anyone teach? This teacher says no. *Diane Ravitch's Blog*, 2012d. Disponível em: <<http://wp.me/p2odLa-yq>>. Acesso em: 30 jul. 2012.

\_\_\_\_\_. Kindergarten madness. Disponível em: *Diane Ravitch's Blog*, 2012e. Disponível em: <<http://dianeravitch.net/2012/07/21/kindergarten-madness>>. Acesso em: 30 jul. 2012.

\_\_\_\_\_. This is what people will say in 20 years, I predict. *Diane Ravitch's Blog*, 2012f. Disponível em: <<http://dianeravitch.net/2012/08/04/this-is-what-people-will-say-in-20-years-i-predict>>. Acesso em: 30 jul. 2012.

\_\_\_\_\_. School letter grades are preposterous. *Diane Ravitch's Blog*, 2012g. Disponível em: <<http://dianeravitch.net/2012/07/22/school-letter-grades-are-preposterous>>. Acesso em: 30 jul. 2012.

\_\_\_\_\_. Arne Duncan is imposing NCLB on teacher education. *Diane Ravitch's Blog*, 2012h. Disponível em: <<http://dianeravitch.net/2012/08/01/arne-duncan-is-imposing-nclb-on-teacher-education>>. Acesso em: 30 jul. 2012.

\_\_\_\_\_. Denial as reform strategy. *Diane Ravitch's Blog*, 2012j. Disponível em: <<http://dianeravitch.net/2012/08/21/denial-as-a-reform-strategy>>. Acesso em: 30 jul. 2012.

\_\_\_\_\_. Tell the truth about american education (Math). *Diane Ravitch's Blog*, 2012l. Disponível em: <<http://dianeravitch.net/2012/07/25/tell-the-truth-about-u-s-education-math>>. Acesso em: 30 jul. 2012.

\_\_\_\_\_. Mayor Bloomberg is angry. *Diane Ravith's Blog*, 2012m. Disponível em: <<http://dianeravitch.net/2012/06/26/mayor-bloomberg-is-angry>>. Acesso em: 30 jul. 2012.

\_\_\_\_\_. What can we learn from Finland. *Diane Ravitch's Blog*, 2012n. Disponível em: <[http://blogs.edweek.org/edweek/Bridging-Differences/2011/10/what\\_can\\_we\\_learn\\_from\\_finland.html](http://blogs.edweek.org/edweek/Bridging-Differences/2011/10/what_can_we_learn_from_finland.html)>. Acesso em: 30 jul. 2012.

\_\_\_\_\_. How choice may kill public education. *Diane Ravitch's Blog*, 2012o. Disponível em: <<http://dianeravitch.net/2012/06/24/how-choice-may-kill-public-education>>. Acesso em: 30 jul. 2012.

REGUNBERG, Aaran. Corporate agenda behind public charters schools. *RI Future*, 2012. Disponível em: <<http://www.rifuture.org/the-agenda-behind-public-charter-schools.html>>. Acesso em: 15 ago. 2012.

RIANI, Juliana L. R.; SOARES, Tufi M. Programa de Alfabetização de Minas Gerais (Proalfa). SEMINÁRIO CEDEPLAR, 2008, Diamantina-MG. Disponível em: <[http://www.cedeplar.ufmg.br/seminarios/seminario\\_diamantina/2008/D08A081.pdf](http://www.cedeplar.ufmg.br/seminarios/seminario_diamantina/2008/D08A081.pdf)>. Acesso em: 30 mar. 2013.

RISTOFF, Dilvo. Avaliação institucional: pensando princípios. In: BALZAN, N.; DIAS SOBRINHO, J. *Avaliação institucional: teoria e experiências*. São Paulo: Cortez, 1995. p. 37-51.

RODRIGUES, Margarida M. *Avaliação educacional sistêmica na perspectiva dos testes de desempenho e de seus resultados: estudo do Saeb*. Tese (Doutorado em Psicologia) – Faculdade de Psicologia, Universidade de Brasília, Brasília. 2007.

ROTHMAN, Robert et al. Benchmarking and alignment of standards and testing. *Achieve*, 2002. Disponível em: <<http://www.achieve.org/files/TR566.pdf>>. Acesso em: 30 ago. 2012.



RUBINSTEIN, Gary. Examining NYC DOE's only egg basket. *Teach for Us*, 2012. Disponível em: <<http://garyrubinstein.teachforus.org/2012/05/01/examining-nyc-does-only-egg-basket>>. Acesso em: 30 jul. 2012.

RUSTIQUE-FORRESTER, Elle. Accountability and the pressures to exclude: a cautionary tale from England. *FACSO*, 2005. Disponível em: <[http://www.facso.uchile.cl/psicologia/epe/\\_documentos/GT\\_cultura\\_escolar\\_politica\\_educativa/recursos%20bibliograficos/articulos%20relacionados/rustique\(2005\)accountabilityandthepressurestoexclude.pdf](http://www.facso.uchile.cl/psicologia/epe/_documentos/GT_cultura_escolar_politica_educativa/recursos%20bibliograficos/articulos%20relacionados/rustique(2005)accountabilityandthepressurestoexclude.pdf)>. Acesso em: 15 ago. 2012.

RYCHEN, Dominique S.; SALGANICK, Laura H. (Orgs.). *Defining and seleting key competencies*. Göttingen: Hogrefe & Huber Publishers, 2001.

SÃO PAULO. *Comunicado SE*, de 22 de março de 1995. São Paulo, 1995.

\_\_\_\_\_. Documento de implantação do Saesp. São Paulo, 1996a.

\_\_\_\_\_. *Resolução SE nº 27*, de 29 de março de 1996. São Paulo, 1996b.

\_\_\_\_\_. *Saesp 98: Manual de orientação*. São Paulo, 1998.

\_\_\_\_\_. *Lei Complementar nº 890*, de 28 de dezembro de 2000. São Paulo, 2000.

\_\_\_\_\_. *Saesp 2000: Caracterização dos estudantes da rede estadual paulista*. São Paulo, 2000.

\_\_\_\_\_. *Resolução SE nº 124*, de 13 de novembro de 2001. São Paulo, 2001.

\_\_\_\_\_. *Decreto nº 45.656*, de 5 de fevereiro de 2001. São Paulo, 2001a.

\_\_\_\_\_. *Saesp ó Sistema de avaliação de rendimento escolar do Estado de São Paulo*. São Paulo, 2004.

\_\_\_\_\_. *Resolução SE nº 86*, de 19 de dezembro de 2007. *Diário Oficial do Estado de São Paulo*, São Paulo, 21 dez. 2007. Seção Executivo I, p. 23.

\_\_\_\_\_. *Resolução SE nº 88*, de 19 de dezembro de 2007. *Diário Oficial do Estado de São Paulo*, São Paulo, 21 dez. 2007a. Seção Executivo I, p. 23-24.

\_\_\_\_\_. *Unificação de dispositivos legais relativos ao ensino fundamental e médio*. São Paulo, 2008.

\_\_\_\_\_. *Orientações Curriculares do Estado de São Paulo: Língua Portuguesa e Matemática ó Ciclo I*. São Paulo, 2008a.

\_\_\_\_\_. *Saresp ó Matrizes de Referência para a avaliação..* São Paulo. 2008b.

\_\_\_\_\_. Lei nº 13.123: Plano Plurianual 2008-2011, de 8 de julho de 2008. *Diário Oficial do Estado de São Paulo*, São Paulo, 31 dez. 2008c. Seção Executivo I, p. 1.

\_\_\_\_\_. Resolução SE nº 74, de 6 de novembro de 2008. *Diário Oficial do Estado de São Paulo*, São Paulo, 7 nov. 2008d. Seção Executivo I, p. 19.

\_\_\_\_\_. Parecer CEE nº 668/2008. *Diário Oficial do Estado de São Paulo*, São Paulo, 31 dez. 2008e. Seção Executivo I, p. 43.

\_\_\_\_\_. Lei complementar nº 1.078, de 17 de dezembro de 2008. *Diário Oficial do Estado de São Paulo*, São Paulo, 18 dez. 2008f. Seção Executivo I, p. 1.

\_\_\_\_\_. Resolução SE nº 76, de 7 de novembro de 2008. *Diário Oficial do Estado de São Paulo*, São Paulo, 11 ago. 2008g. Seção Executivo I, p. 29.

\_\_\_\_\_. *Saresp 2007 ó Sumário Executivo*. São Paulo, 2008h.

\_\_\_\_\_. *Saresp 2008 ó Sumário Executivo*. São Paulo. 2009.

\_\_\_\_\_. *Programa de Qualidade da Escola: nota técnica*. São Paulo, março de 2009, 2009a.

\_\_\_\_\_. *Escala de Proficiência de Língua Portuguesa (Leitura)*. São Paulo, 2009b.

\_\_\_\_\_. *Escala de Proficiência de Matemática*. São Paulo, 2009c.

\_\_\_\_\_. Deliberação CEE nº 84/09. *Diário Oficial do Estado de São Paulo*, São Paulo, 20 mar. 2009d. Seção Executivo I, p. 67.

\_\_\_\_\_. Resolução SE nº 90, de 3 de dezembro de 2009. *Diário Oficial do Estado de São Paulo*, São Paulo, 4 dez. 2009e. Seção Executivo I, p. 28.

\_\_\_\_\_. *Matrizes de referência para a avaliação Saresp: documento básico*. São Paulo: Secretaria de Estado da Educação, 2009f.

\_\_\_\_\_. Resolução SE nº 18, de 4 de março de 2009. *Diário Oficial do Estado de São Paulo*, São Paulo, 5 mar. 2009g. Seção Executivo I, p. 36.

\_\_\_\_\_. Resolução SE nº 70/2009, de 26 outubro de 2009. *Diário Oficial do Estado de São Paulo*, São Paulo, 27 out. 2009h. Seção Executivo I, p. 25.

\_\_\_\_\_. *Saresp 2009* ó Sumário Executivo. São Paulo, 2010.

\_\_\_\_\_. *Currículo do Estado de São Paulo: Ciências Humanas e suas Tecnologias*. São Paulo, 2010a.

\_\_\_\_\_. *Currículo do Estado de São Paulo: Ciências da Natureza e suas Tecnologias*. São Paulo, 2010b.

\_\_\_\_\_. *Currículo do Estado de São Paulo: Linguagens, Códigos e suas Tecnologias*. São Paulo, 2010c.

\_\_\_\_\_. *Currículo do Estado de São Paulo: Matemática e suas Tecnologias*. São Paulo, 2010d.

\_\_\_\_\_. *Saresp 2010* ó Sumário Executivo ó V. 1. São Paulo, 2011.

\_\_\_\_\_. Decreto nº 57.141, de 18 de julho de 2011. *Diário Oficial do Estado de São Paulo*, São Paulo, 19 jul. 2011a. Seção Executivo I, p. 1-10.

\_\_\_\_\_. PPA 2012-2015: Lei nº 14.676, de 27 de dezembro de 2011. *Diário Oficial do Estado de São Paulo*, São Paulo, 29 dez. 2011b. Seção Executivo I, p. 3.

\_\_\_\_\_. Programa de Qualidade da Escola: nota técnica, março de 2011c, São Paulo, 2011.

\_\_\_\_\_. Resolução SE nº 73, de 18 de novembro de 2011. *Diário Oficial do Estado de São Paulo*, São Paulo, 19 nov. 2011d. Seção Executivo I, p. 22.

\_\_\_\_\_. Resolução SE nº 81, de 16 de dezembro de 2011. *Diário Oficial do Estado de São Paulo*, São Paulo, 17 dez. 2011e. Seção Executivo I, p. 28.

\_\_\_\_\_. *Saresp 2011* ó Sumário Executivo ó V. 1. São Paulo, 2012.

SACKS, Peter. *Standardized minds: the high price of america's testing culture and what we can do to change it*. New York: Da Capo Press, 2000.

SACRISTÁN, José Gimeno. *Educar por competências. O que há de novo?* Porto Alegre: Artmed, 2011.

SAHLBERG, Pasi. *Finnish lessons: what can the world learn from educational change in Finland?* New York: Teacher College Press, 2011.

\_\_\_\_\_. How GERM is infecting schools around the world? Pasi Shlberg's Blog, 2012.

Disponível em: <<http://www.pasisahlberg.com/blog/?p=234>>. Acesso em: 15 ago. 2012.

SANFELICE, José Luís.; PEDROSA, Larisse D. Minas aponta o caminho: o processo de reforma da educação mineira. SEMINÁRIO NACIONAL ESTADO E POLÍTICAS SOCIAIS NO BRASIL, 2., 2005, Cascavel-PR. Disponível em: <<http://cac.php.unioeste.br/projetos/gpps/midia/seminario2/trabalhos/educacao/medu21.pdf>>. Acesso em: 26 mar. 2013.

SANTOS, Maria da Glória A. Os alunos da 8ª série do ensino fundamental em Minas Gerais: desempenho em Língua Portuguesa. *Estudos em Avaliação Educacional*, São Paulo, v. 1, n. 9, jan. p. 59-76, 1994.

SASS, Tim. The stability of value-added measures of teacher quality and implications for teacher compensation policy. *Urban*, 2008. Disponível em: <[http://www.urban.org/uploadedpdf/1001266\\_stabilityofvalue.pdf](http://www.urban.org/uploadedpdf/1001266_stabilityofvalue.pdf)>. Acesso em: 15 ago. 2012.

SCHULTZ, Theodore W. Capital formation by education. *Journal of Political Economy*, Chicago, v. 68, n. 6, p. 571-583, dez. 1960.

SCRIVEN, Michael. The methodology of evaluation. In: TYLER, Ralph W.; GAGNÉ, Robert M.; SCRIVEN, Michael (Orgs.). *Perspectives of curriculum evaluation*. Chicago: Rand McNally, 1967. p. 39-83.

SECRETARIA DE ESTADO DA EDUCAÇÃO DE SÃO PAULO. Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo. *Estudos em Avaliação Educacional*, São Paulo, v. 1, n. 13, p.121-136, jan./jun. 1996.

SIGUETA, Elzo. A seleção de recursos humanos e a contribuição da Fundação Carlos Chagas: uma perspectiva histórica. *Educação e Seleção*, São Paulo, v. 1, n. 11, p. 35-45, jan./jun. 1985. Disponível em:

<<http://www.fcc.org.br/pesquisa/publicacoes/es/artigos/97.pdf>>. Acesso em: 5 maio 2013.

SILVÉRIO, Ailton D. O.; XAVIER, Edir P. C.; PORTINHO, Dalva D. F. Os alunos da 8ª série do ensino fundamental em Minas Gerais: desempenho em Matemática. *Estudos em Avaliação Educacional*, São Paulo, n. 9, p. 101-113, jan. 1994.

SILVA, Maria Juliana A. *Regulação educativa: o uso dos resultados das avaliações do Proeb por diretores escolares em Minas Gerais*. Tese (Doutorado em Educação) ó Faculdade de Educação, Universidade Federal de Minas Gerais, Belo Horizonte, 2011.

SIMÕES, Anilce M. Avaliação dos textos de alunos do Ciclo Básico de Alfabetização do estado de Minas Gerias: um exercício de análise. *Estudos em Avaliação Educacional*, São Paulo, . n. 14, p. 141-197, jul. 1996.

SOARES, José Francisco. Índice de Desenvolvimento da Educação Básica de São Paulo (Idesp). *São Paulo Perspectiva*, São Paulo, v. 23, n. 1, p. 29-41, jan./jun. 2009.

\_\_\_\_\_. Enem no Ideb? Oportunidade ou manipulação? 2012. Disponível em: <[http://www.schwartzman.org.br/sitesimon/?p=3855□=pt-br&utm\\_source=feedburner&utm\\_medium=email&utm\\_campaign=Feed%3A+SimonsSite+%28Simon%27s+Site%29](http://www.schwartzman.org.br/sitesimon/?p=3855□=pt-br&utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+SimonsSite+%28Simon%27s+Site%29)>. Acesso em: 12 set. 2012.

SOARES, Magda B. Por uma alfabetização até os oito anos de idade. In: TODOS PELA EDUCAÇÃO. *De Olho nas Metas 2010*. São Paulo: Todos pela Educação, 2010. p. 14-17.

Disponível em:

<[http://www.todospelaeducacao.org.br//arquivos/biblioteca/sumario\\_de\\_olho\\_nas\\_metas\\_\\_\\_2010\\_pdf\(final\).pdf](http://www.todospelaeducacao.org.br//arquivos/biblioteca/sumario_de_olho_nas_metas___2010_pdf(final).pdf)>. Acesso em: 18 abr. 2013.

SOUZA, Maria Alba. A avaliação da escola pública de Minas Gerais. *Estudos em Avaliação Educacional*, São Paulo, v. 1, n. 12, p. 25-32, jul. 1995.

\_\_\_\_\_. A experiência de avaliação educacional em Mnas Gerais: 1992 a 1998. *Estudos em Avaliação Educacional*, São Paulo, v. 1, n. 19, p. 57-76, jan./jun. 1999.

\_\_\_\_\_. Avaliação do rendimento do aluno da escola pública estadual de Minas Gerais no período 1991-1998: a experiência e seus ensinamentos. *Estudos em Avaliação Educacional*, São Paulo, 18, v. 18, n. 37, p. 42-90, maio/ago 2007.

SOUZA, Sandra Z.; OLIVEIRA, Romualdo P. *Sistemas de Avaliação Educacional no Brasil: carcterísticas, tendências e uso dos resultados*. São Paulo: Universidade de São Paulo, 2007.

SPRENG, Connor P. Policy options for interventions on failing schools. Tese (Doutorado em Public Policy Analysis) ó Pardee Rand Graduate School. Santa Monica, CA, 2005.

Disponível em: <[http://www.rand.org/pubs/rgs\\_dissertations/2005/RAND\\_RGSD187.pdf](http://www.rand.org/pubs/rgs_dissertations/2005/RAND_RGSD187.pdf)>.

Acesso em: 8 ago. 2012.

STRAUSS, Valerie. Probe: widespread cheating on tests detailed in Atlanta. *The Washington Post*, 2011. Disponível em: <[http://www.washingtonpost.com/blogs/answer-sheet/post/probe-widespread-cheating-on-tests-detailed-in-](http://www.washingtonpost.com/blogs/answer-sheet/post/probe-widespread-cheating-on-tests-detailed-in-atlanta/2011/07/05/gHQAURaczH_blog.html#pagebreak)

[atlanta/2011/07/05/gHQAURaczH\\_blog.html#pagebreak](http://www.washingtonpost.com/blogs/answer-sheet/post/probe-widespread-cheating-on-tests-detailed-in-atlanta/2011/07/05/gHQAURaczH_blog.html#pagebreak)>. Acesso em: 15 ago. 2012.

STUFFLEBEAM, Daniel L. Introduction: recommendations for improving evaluations in U. S. public schools. *Studies in Educational Evaluation*, v. 20, n. 1, p. 3-21, 1994.

SUCHMAN, Edward A. Evaluative research: principles and practice in public service and social action programs. *Teacher Center for Educational Research*, 1976. Disponível em:

<[http://www.tcer.org/research/documents/teacher\\_turnover\\_full.doc](http://www.tcer.org/research/documents/teacher_turnover_full.doc)>. Acesso em: 15 ago.

2012.

TAVARES, Cristina Z. *Teoria da resposta: uma análise crítica dos pressupostos epistemológicos*. Relatório (Pós Doutorado em Educação) ó Faculdade de Educação, Universidade de São Paulo, São Paulo. 2012.

TEACHER CENTER FOR EDUCATIONAL RESEARCH TCER. The cost of teacher turnover. TCER, 2000. Disponível em:

<[http://www.tcer.org/research/documents/teacher\\_turnover\\_full.doc](http://www.tcer.org/research/documents/teacher_turnover_full.doc)>. Acesso em: 8 ago.

2012.

TEMKIN, Debora; ROELLKE, Cristopher. Federal educational control in No Child Left Behind: implications of two court challenges. In: RICE, Jenifer K.; ROELLKE, Cristopher (Orgs.). *High stakes accountability: implications for resources and capacity*. Charlotte: IAP, 2009.

TODOS PELA EDUCAÇÃO. *Compromisso Todos Pela Educação ó Bases Éticas, Jurídicas, Pedagógicas, Gerenciais, Político-Sociais e Culturais*. São Paulo, agosto de 2006. Disponível em: <<http://www.todospelaeducacao.org.br//arquivos/biblioteca/d1bf9ccb-028a-495e-890e-8d16fb9c54a7.PDF>>. Acesso em: 19 abr. 2013.

\_\_\_\_\_. *Todos pela Educação 2006-2009*. São Paulo, 2009.

\_\_\_\_\_; EDITORA MODERNA. *Anuário brasileiro da educação básica*. São Paulo: Moderna. 2013.

TYLER, Ralph W. *Basic Principles of Curriculum and Instruction*. Chicago: The University of Chicago Press, 1969.

UNIVERSIDADE FEDERAL DE MINAS GERAIS. Relação entre pobreza e aprendizado escolar deve ser revista, defende pesquisadora da UnB. *UFMG online*, 6 maio 2013. Disponível em: <<https://www.ufmg.br/online/arquivos/028207.shtml>>. Acesso em: 6 maio 2013.

UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION. *A view inside primary schools: a world educational indicators (WEI) cross-national study*. Montreal: UNESCO Institute of Statistics, 2008.

UNITED STATES OF AMERICA. *Race to the Top Grantee Frequently Asked Questions*. U.S. Department of Education, 20 maio 2010. Disponível em: <<http://www2.ed.gov/programs/racetothetop/faq.pdf>>. Acesso em: 3 jul. 2012.

VALLE, Raquel C. A construção e a interpretação das escalas de conhecimento ó considerações gerais e uma visão do que vem sendo feito no Saesp. *Estudos em Avaliação Educacional*, São Paulo, v. 1, n. 23, p. 71-92. jan./jun. 2001.

VIANNA, Heraldo M. Avaliação do rendimento de alunos de escolas de 1º grau da rede pública: um estudo em 20 cidades. *Educação e Seleção*, São Paulo, jan./jun. 1989.

\_\_\_\_\_. *Avaliação do rendimento de alunos de escolas públicas da rede pública: um estudo em 15 capitais e 24 outras cidades*. São Paulo: Fundação Carlos Chagas, 1990.

\_\_\_\_\_. Avaliação do Ciclo Básico de Alfabetização em Minas Gerais. *Estudos em Avaliação Educacional*, São Paulo, v. 1, n. 5, p. 91-94, jan./jun. 1992.

\_\_\_\_\_. Programas de avaliação em larga escala: algumas observações. *Estudos em Avaliação Educacional*, v. 1, n. 23, p. 93-104, jan./jun., 2001.

\_\_\_\_\_. *Avaliações em debate: Saeb, Enem, Provão*. Brasília: Plano Editora, 2003.

\_\_\_\_\_. Fundamentos de um programa de avaliação educacional. *Estudos em Avaliação Educacional*, São Paulo, v. 1, n. 28, p. 23-37, jul./dez. 2003a.

\_\_\_\_\_. *Fundamentos de um programa de avaliação educacional*. Brasília: Liber Livro, 2005.

\_\_\_\_\_; ANTUNES, Ana L.; SOUZA, Maria Alba. Desenvolvimento de um Programa de Avaliação do Sistema Estadual de Ensino: o exemplo de Minas Gerais. *Estudos em Avaliação Educacional*, São Paulo, v. 1, n. 8, p. 5-37, jul./dez. 1993.

\_\_\_\_\_; FRANCO, Guilherme T. *Avaliação da aprendizagem: instrumento para a eficiência e qualidade do ensino*. Brasília: Ipea, 1991.

\_\_\_\_\_; GATTI, Bernadete A. Avaliação do rendimento de alunos de escolas de 1º grau da rede pública: uma aplicação experimental em 10 cidades. *Educação e Seleção*, São Paulo, jan./jun. 1988.

\_\_\_\_\_; SQUÁRCIO, Nilza do Carmo; VILHENA, Maria das Graças. As escolas de Minas Gerais e o Ciclo Básico da Alfabetização. *Estudos de Avaliação Educacional*, São Paulo, v.1, n. 6, p. 137-159, jul. 1992.

VIEIRA, Evaldo. A política e a base do direito educacional. *Cedes*, Campinas, v. 21, n. 55, p. 9-29, nov. 2001.

VILLAS BOAS, Benigna M. F. *Virando a escola do avesso por meio da avaliação*. Campinas: Papyrus, 2008.

YANNOULAS, S. C. *Política educacional e pobreza: múltiplas abordagens para uma relação multideterminada*. Brasília: Liber Livros, 2013.

YANNOULAS, Silvia C.; SOUZA, Camila F.; ASSIS, Samuel G. Políticas educacionais e Estado avaliador: uma relação conflitante. *Sociedade em Debate*, Pelotas-RS, v. 15, n. 2, p. 55-67, jul./dez. 2009.

YANNOULAS, Silvia C. ; OLIVEIRA, T. S. de . Avatares de Prometeu: duas décadas de avaliação e regulação das políticas educacionais. *Linhas Críticas*, Brasília, v. 1, n. 19, p. 71-88, 2013.

ZANARDINI, J. B. *Ontologia e avaliação da educação básica no Brasil (1990-2007)*. Tese (Doutorado em Educação) ó Faculdade de Educação, Universidade Federal de Santa Catarina, Florianópolis, 2008.



## APÊNDICE A

### O processo de construção da interpretação das escalas do Saeb e os problemas surgidos

A interpretação das escalas do Saeb é uma prática que vem ocorrendo desde o ciclo de 1995. Neste apêndice é apresentada uma discussão sobre as diversas escalas interpretadas e os problemas que surgiram nesse processo, indicando-se, a seguir, uma sugestão para resolvê-los. Para fazer isso, apresenta-se um pequeno histórico, com base em publicações do Inep, mostrando como o processo de interpretação foi realizado nos diferentes ciclos do Saeb.

O relatório *Saeb 1997 – Primeiros Resultados* (INEP, 1997a) diz que a metodologia utilizada envolve a identificação dos itens âncora e a interpretação deles por um Painel de Especialistas. O documento define que para um item ser escolhido como âncora, ele deve satisfazer a três condições simultaneamente: 65% ou mais de todos os alunos situados em determinado nível acertam o item; menos de 50% dos alunos posicionados no nível anterior acertam o item; a diferença entre os percentuais de alunos posicionados nesse nível e os alunos posicionados no nível imediatamente superior que acertaram o item deve ser maior que 30%.

Para a interpretação, foram escolhidos os seguintes pontos da escala: 175, 250, 325 e 400. Nesse documento (p. 7-10), há a descrição das interpretações de Matemática, Língua Portuguesa e Ciências. Não são citados seus autores, mas provavelmente deve ter sido de autoria do Consórcio Carlos Chagas/Cesgranrio, responsáveis pela aplicação dos testes. Não se sabe também os nomes dos participantes do Painel de Especialistas.

O relatório *Saeb – Resultados 99* (INEP, 2000) mostra que a interpretação, em vez de ser realizada apenas a partir de determinados pontos da escala, seria realizada para níveis delimitados por alguns dos pontos da escala, ou seja, seria descrito não apenas o ponto, mas o intervalo. Segundo o documento,

[...] não é possível atribuir significado pedagógico às médias de desempenho que se situam entre dois pontos interpretados da escala. Sob esta perspectiva, portanto, havia lacunas na interpretação das escalas de desempenho, dificultando a análise e utilização dos resultados pelos agentes educacionais (p. 20).

Foram definidos níveis diferentes para cada área<sup>128</sup>. Nas páginas 54 a 64 do relatório, apresenta-se a interpretação de Matemática e a de Língua Portuguesa aparece nas páginas 66 a 79.

No ciclo de 2001, foi editada uma publicação específica sobre a metodologia utilizada na interpretação da escala: *Saeb 2001 ó Relatório de Interpretação das Escalas* (INEP, 2002d), na qual se afirma que o painel de especialistas ocorreu em abril de 2002. O documento esclarece, em linhas gerais, a metodologia utilizada, na época, pela Fundação Carlos Chagas e pela Fundação Cesgranrio, enumerando os participantes do painel de especialistas<sup>129</sup>, mas não se fornecem informações detalhadas sobre como ele ocorreu.

No documento afirma-se que nos ciclos anteriores foi utilizada a metodologia descrita em Beaton e Allen (1992), mas não se comentam as diferenças metodológicas com relação aos dois ciclos anteriores. São enumerados quatro problemas com relação à definição do critério para a escolha dos itens âncora:

- a) para se ter muitos itens âncora, os níveis selecionados têm que ser bem espaçados;
- b) para se ter mais itens para auxiliar a interpretação, é necessário ampliar o conceito de nível âncora para quase-âncora;
- c) houve dificuldade dos membros dos painéis de especialistas de utilizarem o conceito de discriminação entre níveis, só usando praticamente itens considerados dominados pelos alunos em torno do nível;
- d) a não inclusão, na descrição da escala, de alguns descritores cobertos por itens que, embora de boa qualidade técnica e pedagógica, não foram classificados como itens âncora. A solução para isso foi identificar estes itens e verificar em que níveis os alunos estariam dominando estes assuntos através do conhecimento dos percentuais de acerto em diversas partes da escala (p. 3-4).

O documento continua informando sobre a metodologia utilizada:

A modificação introduzida em 1999 e mantida em 2001 foi selecionar os pontos da escala de 100 a 425, em intervalos de 25, o que inclui o ponto 250, média arbitrada da distribuição de proficiências da 8ª série em Língua Portuguesa e Matemática.

Para cada um desses pontos ou níveis, foi considerado um intervalo de comprimento 25 e centrado no ponto.

---

<sup>128</sup> Para Matemática: até 160, > 160 até 175, > 175 até 225, > 225 até 275, > 275 até 325, > 325 até 375, > 375 até 425 e > 425 até 475; para Língua Portuguesa: até 150, >150 até 200, >200 até 250, >250 até 300, > 300 até 350, > 350 até 400 e >400.

<sup>129</sup> Em Língua Portuguesa: Anna Maria Marques Cintra (SP), Lillian Guiuro Passarelli (SP), Nilvia Therezinha da Silva Pantaleoni (SP), Ana Lucia Tinoco Cabral (SP), Carlos Alberto Faraco (PR) e Maria Irandé Costa Morais Antunes (PE); em Matemática: Edda Cury (SP), Lillian Nasser (RJ), Maria Alice Gravina (RS), Paulo Cezar Pinto Carvalho (RJ) e Ruy César Pietropaolo (SP).

Para cada item foi estimado o percentual de acerto dos alunos em cada nível, calculando-se o percentual de acerto dos alunos com proficiência no intervalo que contém o nível.

Nessa abordagem, um item é dito âncora em um nível se:

1. O número de alunos no nível que respondeu ao item é maior que 50;
2. O percentual de acertos do item nos níveis anteriores é menor que 65%;
3. O percentual de acertos do item no nível considerado e nos níveis acima é maior que 65%;
4. O ajuste da curva TRI é bom.

Dessa maneira, todo item considerado bom é utilizado, pois vai ser âncora em algum nível, a não ser que seja muito difícil. Esse novo procedimento também dará flexibilidade para o painel de especialistas escolher alguns níveis para resumir a interpretação da escala (p. 4).

Portanto, houve uma mudança nos critérios para a escolha dos itens âncora, sem que se apresentasse uma justificativa para isso. O critério ficou menos rigoroso, talvez para permitir que mais itens participassem da interpretação da escala.

Além disso, das páginas 7 a 13 são apresentadas tabelas que indicam os níveis, o número do item, o bloco em que ele estava e sua ordem nesse bloco. Apresenta-se também a programação do painel, com duração de três dias, um para cada etapa da educação básica avaliada, e acrescenta-se que depois desse painel a equipe do Consórcio, acompanhada por técnicos do Inep, procedeu à revisão e ao refinamento das escalas interpretadas (p. 15).

Apesar de terem sido estabelecidos 8 níveis para Língua Portuguesa (125, 150, 175, 200, 250, 300, 350 e 375) e 10 níveis para Matemática (125, 150, 175, 200, 250, 300, 350, 375, 400 e 425), o documento não esclarece as razões para isso.

O documento apresenta a interpretação da escala em dois formatos. No primeiro, como tabela, indica a interpretação comum a todas as séries e faz uma análise a partir de tópicos (Língua Portuguesa) e temas (Matemática). Essa análise parece que procura agregar informações que aprimorem a simples descrição do item. O segundo formato é uma matriz que apresenta a interpretação por série, relacionando os tópicos ou temas, as habilidades, os níveis da escala e o percentual de alunos em cada nível. Esse segundo formato, visualmente interessante para se divulgar a interpretação, traz informações mais resumidas, que apenas descrevem os itens. Tal formato foi apresentado nas publicações dos relatórios técnicos usados para divulgar os resultados do ciclo de 2001 (INEP, 2002b; 2002c).

Para o ciclo de 2003, foi publicado o documento *Relatório Nacional Saeb 2003* (INEP, 2006) e nele não se faz referência a como a interpretação da escala foi construída, quem participou do processo e qual instituição liderou o processo. A única informação disponível é a apresentação da matriz por série, no formato de 2001, que nesse ciclo foi chamada de processo de construção de competências. O que chama a atenção é o fato de a matriz por série de Língua Portuguesa ir da página 64 até a 87 e a de Matemática ocupar apenas três páginas, dando a impressão de que foram interpretados mais itens de Língua Portuguesa do que de Matemática. Esse formato por si só traz dúvidas sobre a qualidade do trabalho produzido.

É desse documento a classificação dos níveis com os adjetivos *crítico*, *intermediário* e *adequado*. Essa classificação serviu de base para todas as matérias jornalísticas que trataram da divulgação dos resultados desse ciclo. Alguns trabalhos, notadamente aqueles produzidos pelo movimento Todos Pela Educação, de alguma forma, também classificam os diferentes níveis com adjetivos. Percebe-se, ainda, que em algumas avaliações estaduais adjetivos semelhantes vêm sendo utilizados. No caso do Inep, em nenhuma outra publicação da instituição isso foi feito.

Para o ciclo de 2005, foi gerado o *Relatório da Oficina de Interpretação das Escalas* (CENTRO DE SELEÇÃO E PROMOÇÃO DE EVENTOS, 2006), sob a responsabilidade de Mauro Rabelo, Josemberg Andrade e Girlene de Jesus. No documento, chama a atenção o fato de que o Painel de Especialistas foi substituído por uma Oficina de Interpretação de Itens. Essa mudança, menos do que indicar uma simples troca de nomes, pode ter resultado em uma forma diferenciada de trabalho, apesar de o relatório deixar explícito: *decidiu-se manter os procedimentos utilizados nos anos anteriores* (p. 70).

No entanto, o estabelecimento do item âncora baseou-se na escolha dos itens que, para um dado nível, pelo menos 65% dos estudantes acertaram. Nada foi comentado sobre as outras quatro condições, enumeradas de *a* a *e* na publicação *Saeb 2001 ó Relatório de Interpretação das Escalas* (INEP, 2002d). Além disso, foi proposta uma forma de diferenciar o item pelo seu grau de discriminação. Para tanto, foram indicados dois tipos de itens:

Item *Itálico*: quando o item é acertado por pelo menos 65% dos indivíduos daquele nível de habilidade, e a diferença entre esse percentual e o percentual de indivíduos do nível de habilidade anterior que o acertam é inferior a 15%.

Item *negrito*: o item é acertado por pelo menos 65% dos indivíduos daquele nível de habilidade, e a diferença entre esse percentual e o percentual de indivíduos do nível de habilidade anterior que o acertam é inferior a 50%.

Essa categorização indica que praticamente todos os itens poderão ajudar na descrição da escala. No entanto, os itens da classe negrito discriminarão melhor e representarão mais precisamente os níveis, (p. 5).

Pelo que se percebe, foram utilizados dois critérios para a escolha dos itens âncora. O do item itálico, o menos rigoroso de todos até então utilizados, e o do item negrito, que repete o critério do ciclo de 1997. No entanto, não se indica como trabalhar com dois critérios tão conflitantes. Se o método previa que, na ausência de itens dentro do critério mais rigoroso, seriam substituídos por itens escolhidos segundo um critério menos rigoroso, qual o impacto disso para a interpretação da escala? No documento não existe nenhuma justificativa para a escolha desse duplo critério, muito menos uma comparação que demonstre a superioridade desse método em relação aos outros utilizados para a determinação dos itens âncora.

Ainda mais, o documento informa que essa metodologia foi a mesma utilizada em 2003, com a diferença de que agora foram utilizadas as duas classes de itens. Quanto ao grupo de especialistas envolvidos<sup>130</sup>, percebe-se que o trabalho do ciclo de 2005 esteve restrito a um grupo de especialistas ligados diretamente à Universidade de Brasília (UnB). Todas essas informações estão descritas com mais detalhes nas páginas 3 a 11 desse documento.

As interpretações foram apresentadas na forma de uma escala comum às três séries avaliadas e às escalas específicas para cada uma das séries, todas na forma de tabelas. Como informado antes, o formato de matriz, utilizado em 2001, para divulgar a interpretação por anos e série, foi abandonado. Com isso, a qualidade da informação diminuiu bastante na medida em que aparecem informações que parecem conflitantes.

A escala comum entre séries de Língua Portuguesa é composta por 11 níveis, começando no nível 150 e avançando a intervalos de 25 pontos até o nível 400. A de Matemática é composta por 14 níveis, começando em 125 e avançando a intervalos de 25 pontos até o nível 450. Comparando com os dados apresentados neste documento, a escala de interpretação de 2001 utilizou níveis diferentes da escola de 2005.

A publicação apresenta também, como anexo, uma classificação de itens diferenciados por cores (vermelho, verde e azul), que não foram utilizados no teste (CENTRO DE SELEÇÃO E PROMOÇÃO DE EVENTOS, 2006, p. 71). Não se explicita qual a razão para essa nova classificação.

---

<sup>130</sup> Em Língua Portuguesa: Lucília Garcez, Maria Irandé Antunes, Maria Luiza Coroa, Norma Lima, Orlene Carvalho, Vilma Corrêa; em Matemática: Ana Nascimento, Cleyton Gontigo, Haydée Poubel, Maria Terezinha Gaspar, Pedro Garcia, Regina Pina e Romildo Pina.

O relatório técnico de 2005 não foi publicado, assim como os dos ciclos de 2007, 2009 e 2011. Desde então, não é possível identificar nenhum outro documento que trate da interpretação da escala de itens.

Diante o exposto, é necessário compatibilizar todas as interpretações realizadas até então. Uma das formas de fazer isso é normatizar e publicizar os critérios utilizados no processo de interpretação e, a seguir, realizar a interpretação com a maior quantidade de itens utilizados nos testes aplicados, com o objetivo de construir a escala. Até que esse trabalho esteja concluído, recomenda-se utilizar a interpretação realizada em 2001, por ser aquela que parece ter tido maior rigor metodológico.

## APÊNDICE B

### Rotatividade de professoras no ensino fundamental

Os Quadros B1 a B4 mostram a rotatividade das professoras do ensino fundamental nas escolas públicas, de acordo com a rede, entre 2007 e 2011.

**Quadro B1 – Rotatividade de professoras dos anos iniciais do ensino fundamental na rede estadual – 2007 a 2012**

Unidade da Federação		Docentes no Ensino Fundamental – Anos Iniciais – Rede Estadual															
		2007	2008			2009			2010			2011			2012		
		Total	Total	Novos	Δ(%)	Total	Novos	Δ(%)	Total	Novos	Δ(%)	Total	Novos	Δ(%)	Total	Novos	Δ(%)
11	RO	2.429	2.497	806	33	2.547	940	38	2.445	651	26	2.469	686	28	2.481	696	28
12	AC	1.360	1.413	586	43	1.342	513	36	1.296	302	23	1.263	500	39	1.265	355	28
13	AM	3.472	3.344	850	24	3.262	858	26	3.253	861	26	3.232	941	29	3.013	1.341	45
14	RR	1.522	1.297	529	35	1.191	452	35	1.084	406	34	936	292	27	799	310	39
15	PA	3.201	3.346	977	31	2.892	981	29	2.802	1.004	35	2.770	669	24	2.683	930	35
16	AP	2.272	2.184	700	31	1.971	580	27	2.040	743	38	2.004	619	30	1.767	462	26
17	TO	2.319	2.234	718	31	2.216	654	29	2.228	682	31	2.297	786	35	2.012	809	40
21	MA	2.400	2.282	891	37	2.083	888	39	2.168	1.076	52	1.903	605	28	1.072	372	35
22	PI	3.295	3.576	2.046	62	3.497	1.617	45	3.026	1.566	45	2.528	880	29	1.842	585	32
23	CE	524	457	223	43	314	139	30	238	90	29	221	83	35	190	72	38
24	RN	2.798	2.674	987	35	2.453	970	36	2.313	802	33	2.216	1.001	43	2.071	793	38
25	PB	3.709	3.674	1.379	37	3.469	1.220	33	3.388	857	25	3.401	1.189	35	3.118	840	27
26	PE	3.724	3.321	1.196	32	2.894	1.067	32	2.562	859	30	2.042	614	24	1.407	357	25
27	AL	1.742	1.498	347	20	1.412	403	27	1.184	376	27	993	350	30	747	193	26
28	SE	2.125	2.179	886	42	2.131	731	34	1.934	522	24	1.786	650	34	1.737	398	23
29	BA	2.672	2.462	844	32	1.784	619	25	1.042	209	12	908	266	26	685	256	37
31	MG	28.683	28.016	7.795	27	27.050	6.972	25	26.347	6.629	25	24.892	5.809	22	22.402	5.677	25
32	ES	2.195	2.258	805	37	2.389	1.012	45	2.640	1.008	42	2.623	922	35	2.902	1.224	42
33	RJ	5.091	4.544	1.486	29	3.092	894	20	2.561	653	21	1.646	314	12	942	184	20
35	SP	37.558	37.574	9.850	26	37.095	8.645	23	33.621	6.236	17	32.277	5.607	17	32.461	6.619	20
41	PR	1.349	1.337	682	51	1.130	534	40	819	306	27	615	291	36	472	224	47
42	SC	8.526	8.552	2.953	35	7.940	2.583	30	7.110	2.452	31	8.566	3.919	55	7.527	2.916	39
43	RS	14.479	12.608	2.423	17	11.947	2.503	20	11.573	2.483	21	11.814	2.580	22	12.317	2.727	22
50	MS	3.365	3.551	1.218	36	3.345	1.009	28	3.259	802	24	3.417	917	28	3.374	977	29
51	MT	4.141	4.126	1.854	45	3.768	1.446	35	3.427	1.435	38	3.626	1.516	44	3.881	1.664	43
52	GO	3.268	2.745	1.183	36	2.410	861	31	2.121	1.002	42	2.000	1.043	49	1.628	857	53
53	DF	6.052	6.565	2.719	45	6.402	2.091	32	6.666	2.054	32	6.755	2.448	37	6.793	1.761	26
<b>Totais</b>		<b>154.271</b>	<b>150.314</b>	<b>46.933</b>	<b>30</b>	<b>142.026</b>	<b>41.182</b>	<b>27</b>	<b>133.147</b>	<b>36.066</b>	<b>25</b>	<b>129.200</b>	<b>35.497</b>	<b>27</b>	<b>121.588</b>	<b>33.599</b>	<b>28</b>

Fonte: MEC/INEP; dados fornecidos pela DEED/INEP, com elaboração do autor.

**Quadro B2 – Rotatividade de professoras dos anos iniciais do ensino fundamental na rede municipal – 2007 a 2012**

Unidade da Federação		Docentes no Ensino Fundamental – Anos Iniciais – Rede Municipal															
		2007		2008		2009			2010			2011			2012		
		Total	Total	Novos	Δ(%)	Total	Novos	Δ(%)	Total	Novos	Δ(%)	Total	Novos	Δ(%)	Total	Novos	Δ(%)
11	RO	3.264	3.226	1.229	38	3.271	1.295	40	3.348	1.135	35	3.444	1.126	34	3.602	1.274	35
12	AC	1.100	1.095	556	51	1.144	616	56	1.127	516	45	1.174	605	54	1.157	479	41
13	AM	8.896	9.230	3.607	41	9.429	3.907	42	9.230	3.332	35	9.848	3.780	41	9.985	3.625	36
14	RR	959	1.189	650	68	1.332	786	66	1.342	660	50	1.629	823	61	1.884	854	45
15	PA	20.590	22.001	7.856	38	22.947	9.239	42	23.929	8.659	38	24.886	8.340	35	26.210	8.746	33
16	AP	1.200	1.272	591	49	1.347	542	43	1.389	563	42	1.412	494	36	1.507	576	38
17	TO	3.868	3.844	1.408	36	3.812	1.756	46	3.799	1.503	39	3.840	1.517	40	4.124	1.588	39
21	MA	24.376	25.654	10.776	44	26.000	12.248	48	25.993	10.410	40	25.530	9.307	36	26.721	10.043	38
22	PI	11.202	11.848	4.888	44	12.329	5.888	50	11.897	4.651	38	12.243	4.989	42	12.634	4.845	38
23	CE	24.191	24.910	8.530	35	24.504	8.694	35	24.336	7.489	31	24.285	7.072	29	25.331	7.692	30
24	RN	7.551	7.526	2.366	31	7.520	2.756	37	7.188	2.033	27	7.343	2.053	29	7.592	2.077	27
25	PB	10.765	10.878	3.852	36	10.610	4.052	37	10.140	2.970	28	10.051	3.063	30	10.145	2.901	29
26	PE	20.905	21.540	7.881	38	21.273	8.626	40	21.271	7.134	34	20.924	6.609	31	21.682	6.793	31
27	AL	9.345	9.187	3.203	34	9.130	3.335	36	8.555	2.568	28	8.536	2.566	30	8.771	2.614	30
28	SE	5.554	5.698	2.150	39	5.588	2.323	41	5.274	1.747	31	5.127	1.570	30	4.976	1.553	31
29	BA	40.920	43.195	17.415	43	43.560	18.407	43	42.598	14.317	33	42.466	13.068	31	43.928	13.591	31
31	MG	43.535	44.809	15.502	36	44.560	15.808	35	44.588	13.202	30	44.204	12.962	29	44.628	13.296	30
32	ES	8.630	8.810	3.091	36	8.987	3.156	36	9.445	3.219	36	9.867	3.212	34	10.912	3.848	35
33	RJ	30.580	30.334	8.818	29	29.878	9.223	30	29.708	7.734	26	30.525	9.013	30	31.802	9.690	30
35	SP	62.258	67.176	20.698	33	74.054	22.527	34	75.098	19.195	26	76.215	18.893	25	80.293	20.907	26
41	PR	31.062	31.781	8.406	27	31.887	9.109	29	31.869	7.545	24	32.240	7.747	24	32.904	8.510	26
42	SC	14.034	14.084	4.907	35	14.119	5.489	39	13.699	4.490	32	14.586	5.143	38	16.323	6.350	39
43	RS	23.212	22.626	6.080	26	22.416	6.970	31	22.339	6.188	28	22.551	6.411	29	23.441	6.859	29
50	MS	6.546	6.953	2.343	36	7.160	2.566	37	7.001	1.851	26	7.036	1.866	27	7.407	2.158	29
51	MT	7.281	7.352	3.127	43	7.400	3.067	42	7.129	2.608	35	7.098	2.771	39	7.306	2.951	40
52	GO	13.816	14.225	5.031	36	14.645	5.449	38	14.704	4.489	31	15.427	4.892	33	16.411	5.045	31
<b>Totais</b>		<b>435.640</b>	<b>450.443</b>	<b>154.961</b>	<b>36</b>	<b>458.902</b>	<b>167.834</b>	<b>37</b>	<b>456.996</b>	<b>140.208</b>	<b>31</b>	<b>462.487</b>	<b>139.892</b>	<b>31</b>	<b>481.676</b>	<b>148.865</b>	<b>31</b>

Fonte: MEC/INEP; dados fornecidos pela DEED/INEP, com elaboração do autor.



**Quadro B3 – Rotatividade de professoras dos anos iniciais do ensino fundamental na rede privada – 2007 a 2012**

Unidade da Federação		Docentes no Ensino Fundamental – Anos Iniciais – Rede Privada																
		2007		2008			2009			2010			2011			2012		
		Total	Total	Novos	Δ(%)	Total	Novos	Δ(%)	Total	Novos	Δ(%)	Total	Novos	Δ(%)	Total	Novos	Δ(%)	
11	RO	621	656	326	50	632	271	43	648	278	43	670	278	41	706	305	43	
12	AC	182	206	101	49	199	72	36	184	49	27	199	73	37	203	62	31	
13	AM	1.021	1.516	928	61	1.639	699	43	1.677	550	33	1.649	555	34	1.804	774	43	
14	RR	111	124	58	47	124	53	43	147	70	48	196	112	57	214	96	45	
15	PA	1.911	2.561	1.401	55	2.868	1.402	49	3.327	1.476	44	3.504	1.372	39	4.051	1.668	41	
16	AP	250	289	158	55	296	142	48	314	139	44	335	136	41	349	158	45	
17	TO	604	658	295	45	663	260	39	642	254	40	651	276	42	707	304	43	
21	MA	2.783	3.257	1.665	51	3.434	1.555	45	3.531	1.365	39	3.774	1.421	38	3.967	1.429	36	
22	PI	1.930	2.168	956	44	2.332	966	41	2.413	803	33	2.644	903	34	2.821	932	33	
23	CE	7.171	7.784	3.506	45	8.009	3.064	38	8.057	2.696	33	8.140	2.673	33	8.570	3.019	35	
24	RN	2.580	2.816	1.198	43	2.903	1.122	39	3.065	1.135	37	3.063	1.021	33	3.230	1.100	34	
25	PB	1.883	2.793	1.604	57	3.119	1.373	44	3.615	1.461	40	3.756	1.316	35	3.875	1.294	33	
26	PE	5.586	7.702	4.121	54	9.007	4.121	46	9.674	3.584	37	10.406	3.607	35	10.887	3.532	32	
27	AL	1.274	1.713	935	55	1.879	769	41	2.320	1.021	44	2.607	1.007	39	2.838	1.026	36	
28	SE	1.249	1.480	678	46	1.761	792	45	1.896	676	36	2.102	752	36	2.153	682	32	
29	BA	5.153	7.564	4.572	60	8.106	3.889	48	8.889	3.755	42	9.564	3.780	40	11.840	5.473	46	
31	MG	10.376	11.626	4.872	42	12.006	4.062	34	12.301	3.530	29	12.449	3.425	28	12.667	3.394	27	
32	ES	2.011	2.194	880	40	2.183	736	34	2.163	686	32	2.235	672	30	2.224	643	29	
33	RJ	12.361	18.401	9.888	54	18.274	6.157	34	19.268	6.308	33	21.281	7.299	34	22.275	6.884	31	
35	SP	32.808	34.478	11.447	33	36.225	10.777	30	37.171	10.013	27	38.146	10.228	27	39.222	10.549	27	
41	PR	5.386	6.508	3.118	48	6.933	2.599	37	7.106	2.334	33	7.367	2.451	33	7.822	2.718	35	
42	SC	2.681	3.223	1.409	44	3.337	1.065	32	3.450	936	27	3.578	1.033	29	3.876	1.204	31	
43	RS	4.497	4.923	1.633	33	4.950	1.280	26	4.977	1.175	24	4.931	1.245	25	5.767	1.922	33	
50	MS	1.620	1.737	799	46	1.742	694	40	1.742	634	36	1.768	678	38	1.849	690	37	
51	MT	1.371	1.702	862	51	1.793	750	42	1.805	640	35	1.862	692	37	1.887	701	37	
52	GO	4.029	5.007	2.551	51	4.930	2.043	41	5.324	2.156	40	5.401	2.039	38	5.764	2.278	40	
53	DF	2.070	3.210	1.938	60	3.166	1.140	36	3.204	1.131	35	3.199	1.061	33	3.344	1.135	34	
<b>Totais</b>		<b>113.519</b>	<b>136.296</b>	<b>61.899</b>	<b>55</b>	<b>142.510</b>	<b>51.853</b>	<b>36</b>	<b>148.910</b>	<b>48.855</b>	<b>33</b>	<b>155.477</b>	<b>50.105</b>	<b>32</b>	<b>164.912</b>	<b>53.972</b>	<b>33</b>	

Fonte: MEC/INEP; dados fornecidos pela DEED/INEP, com elaboração do autor.

**Quadro B4 – Rotatividade de professoras dos anos iniciais do ensino fundamental nas três redes – 2007 a 2012**

Unidade da Federação		Rede Estadual				Rede Municipal				Rede Privada			
		2007	2012		2007-2012	2007	2012		2007-2012	2007	2012		2007-2012
		Total	Total	Novos	Δ(%)	Total	Total	Novos	Δ(%)	Total	Total	Novos	Δ(%)
11	RO	2.429	2.481	1.469	60	3.264	3.602	2.332	71	621	706	578	93
12	AC	1.360	1.265	796	59	1.100	1.157	878	80	182	203	140	77
13	AM	3.472	3.013	1.890	54	8.896	9.985	6.533	73	1.021	1.804	1.517	149
14	RR	1.522	799	504	33	959	1.884	1.554	162	111	214	185	167
15	PA	3.201	2.683	1.778	56	20.590	26.210	17.365	84	1.911	4.051	3.421	179
16	AP	2.272	1.767	815	36	1.200	1.507	1.044	87	250	349	295	118
17	TO	2.319	2.012	1.320	57	3.868	4.124	2.791	72	604	707	552	91
21	MA	2.400	1.072	623	26	24.376	26.721	17.892	73	2.783	3.967	3.123	112
22	PI	3.295	1.842	1.263	38	11.202	12.634	8.268	74	1.930	2.821	2.093	108
23	CE	524	190	146	28	24.191	25.331	15.372	64	7.171	8.570	6.077	85
24	RN	2.798	2.071	1.328	47	7.551	7.592	4.233	56	2.580	3.230	2.363	92
25	PB	3.709	3.118	1.815	49	10.765	10.145	5.826	54	1.883	3.875	3.192	170
26	PE	3.724	1.407	930	25	20.905	21.682	13.642	65	5.586	10.887	8.686	155
27	AL	1.742	747	439	25	9.345	8.771	5.161	55	1.274	2.838	2.380	187
28	SE	2.125	1.737	1.127	53	5.554	4.976	2.730	49	1.249	2.153	1.699	136
29	BA	2.672	685	396	15	40.920	43.928	27.630	68	5.153	11.840	10.050	195
31	MG	28.683	22.402	11.163	39	43.535	44.628	24.717	57	10.376	12.667	8.530	82
32	ES	2.195	2.902	2.074	94	8.630	10.912	6.923	80	2.011	2.224	1.577	78
33	RJ	5.091	942	361	7	30.580	31.802	17.975	59	12.361	22.275	16.876	137
35	SP	37.558	32.461	14.570	39	62.258	80.293	46.033	74	32.808	39.222	24.095	73
41	PR	1.349	472	379	28	31.062	32.904	16.655	54	5.386	7.822	6.044	112
42	SC	8.526	7.527	4.961	58	14.034	16.323	10.520	75	2.681	3.876	2.796	104
43	RS	14.479	12.317	5.848	40	23.212	23.441	12.792	55	4.497	5.767	3.738	83
50	MS	3.365	3.374	2.021	60	6.546	7.407	4.402	67	1.620	1.849	1.409	87
51	MT	4.141	3.881	2.779	67	7.281	7.306	4.779	66	1.371	1.887	1.479	108
52	GO	3.268	1.628	1.071	33	13.816	16.411	10.351	75	4.029	5.764	4.522	112
53	DF	6.052	6.793	3.947	65	-	-	-	-	2.070	3.344	2.642	128
<b>Totais</b>		<b>154.271</b>	<b>121.588</b>	<b>65.813</b>	<b>43</b>	<b>435.640</b>	<b>481.676</b>	<b>288.398</b>	<b>66</b>	<b>113.519</b>	<b>155.477</b>	<b>105.874</b>	<b>93</b>

Fonte: MEC/INEP; dados fornecidos pela DEED/INEP, com elaboração do autor.

Para cada um dos quatro Quadros (B1, B2, B3 e B4), a coluna "Total" refere-se ao total de professoras contabilizadas pelo Censo Escolar no ano em questão. A coluna "Novos" refere-se à quantidade de professoras que estão presentes no Censo do ano indicado em determinada escola, mas que não estavam nessa escola no ano anterior. Diversas podem ser as explicações para esse fato, entre elas: mudança para escolas diferentes daquelas que se lecionou no ano anterior; professoras temporárias que atuam durante determinado período e depois são desligadas; diversos tipos de licença, com destaque para as licenças-gestantes, dado o grande contingente de professoras; cessões para outros órgãos; aposentadoria, morte ou demissão. Essa diferença de professoras, de um ano para outro, dentro da mesma escola, é assinalada na coluna "R" em termos percentuais e indica a rotatividade de professoras nas escolas. Assim, analisando-se o Quadro B1, para a rede estadual de Rondônia, a rotatividade de 33% entre 2008 e 2007 mostra que das docentes em efetivo exercício nas escolas em 2008, na época da coleta dos dados, 33% são novas e não estavam em exercício nas mesmas escolas em 2007. Quando, no Quadro B4, a porcentagem é maior que 100%, esse valor indica que, entre os dois períodos, houve um aumento no número de docentes em exercício, causado pelo aumento no número de turmas de alunos.

Analisando-se os Quadros B1, referente à rede estadual, e B2, referente à rede municipal, percebe-se que é na rede estadual que se observam as maiores e menores variações de professoras em cada escola, de um ano para outro. Entre as redes estaduais, o Piauí apresenta a maior alteração, com variações de 62%, 45%, 45%, 29% e 32%, para os períodos 2007-2008, 2008-2009, 2009-2010, 2010-2011 e 2011-2012, respectivamente, sendo que no período 2010-2011 só foi menor que nos anteriores provavelmente porque houve uma queda de quase 20% no número de docentes, talvez devido à transferência de escolas de anos iniciais dos Estados para os municípios. No entanto, a rotatividade volta a aumentar no período seguinte. A rede estadual do Rio Grande do Sul é a que apresenta a menor alteração, com variações de 17%, 20%, 21%, 22% e 22% para os mesmos períodos, respectivamente.

Ainda para as redes públicas, quando se analisam as regiões, em geral, a rotatividade é maior nas redes municipais do que nas redes estaduais, com destaque para a Região Norte, em que a rotatividade nas redes municipais, em todos os Estados, é maior que nas redes estaduais. Ocorre maior rotatividade nas redes estaduais, quando comparadas com as redes municipais, em nove Estados: Rio Grande do Norte, Paraíba, Sergipe, Espírito Santo, Paraná, Santa Catarina, Mato Grosso do Sul, Mato Grosso e Goiás. Portanto, como era de se esperar, o aumento da rotatividade das professoras é um fenômeno que se dá no interior de cada Estado,

e não é possível observar um padrão que possa explicá-lo. Em cada rede existem condições diferenciadas que devem favorecer, ou não, essa rotatividade.

Analisando-se as redes públicas dos Estados mais ricos, com relação aos mais pobres, também não é possível observar uma tendência clara. Por exemplo, a rotatividade de docentes na rede do Rio Grande do Sul, de São Paulo e do Rio de Janeiro estão entre as mais baixas, com média em torno de 20%, mas no Distrito Federal essa média é de 36%.

Quando se analisa o Quadro B3, relativo à rede privada, percebe-se que o problema da rotatividade não é privilégio apenas das duas redes públicas. O que basicamente diferencia as redes públicas das redes privadas é a possibilidade de haver demissão na rede privada, caso as professoras não tenham o desempenho profissional esperado pelas entidades mantenedoras das escolas. Talvez diminua a quantidade de solicitação de licenças, com exceção das licenças-maternidade, por conta de um controle mais rigoroso. Dependendo do Estado, a rede privada tem rotatividade ora maior que a rede estadual, ora maior que a rede municipal, mas em São Paulo a rede privada tem rotatividade superior às duas redes públicas. No caso da rede privada, é importante destacar que, nos últimos anos, com o aumento da renda da população, tem aumentado o número de matrículas nessa rede e, com isso, tem aumentado o número de novos postos de trabalho para as professoras.

Com uma variação tão expressiva no corpo docente das escolas de um ano para o outro, é muito difícil que o projeto pedagógico delas possa ser implementado com sucesso, pois, entre outras coisas, se dilui o espírito de corpo dentro daquele estabelecimento e, a cada ano, boa parte das relações deve ser refeita. Se fossemos tomar como exemplo o mundo empresarial, exemplo tão ao gosto dos reformadores empresariais, nenhuma empresa resistiria no mercado com uma rotatividade como a apresentada no conjunto das duas redes públicas, ultrapassando incríveis 30%! Assim, políticas de bonificação que levem em consideração o desempenho da equipe escolar não fazem nenhum sentido, pois a variação que sofre essa equipe é muito grande.

Quando se analisa o Quadro B4, no qual se compara o ano de 2007 com o ano de 2012, cinco anos depois, essa rotatividade fica ainda mais destacada, com variações médias de 43% e 66% para as redes estaduais e municipais, respectivamente. Com relação à rede privada, a variação média observada é de 93%. Nesse caso, além da rotatividade, houve também a abertura de novos postos de trabalho, mas quando se comparam os totais em um período de cinco anos, o aumento do número de postos de trabalho é inferior à rotatividade no período. Um exemplo disso é a rede privada do Distrito Federal, que passou de 2.070 para

3.344 professoras, um aumento de 1.274 postos de trabalho, inferior à quantidade de 2.642 novas professoras que não estavam nas mesmas escolas cinco anos antes.

Quando a rotatividade apresentada neste estudo é comparada com estudos produzidos para a realidade americana, a rotatividade de professoras nas escolas brasileiras impressiona. As pesquisas nos Estados Unidos mostram que 30% das professoras novas abandonam a profissão depois de cinco anos de exercício (INGERSOLL, 2001) e que em Estados como o Texas, anualmente, entre 11% e 18% das professoras deixam a profissão e entre 6% e 10% mudam de escola, sendo que a maior porcentagem refere-se a professoras com até dois anos de formadas (HANUSHECK; KAIN; RIVKIN, 1999).

Não foi o objetivo desta discussão detalhar as razões da rotatividade das professoras, mas apenas alertar para um problema sério, com grande impacto nas políticas educacionais. Por exemplo, muitos dos programas de formação continuada não levam em conta esse problema e estão programados para serem oferecidos durante um ciclo com duração entre um e dois anos. Terminado o ciclo, o programa encerra-se. Como a rotatividade das professoras é muito grande, se esses programas não forem repetidos continuamente, não será possível abarcar todo o contingente de professoras que estão em sala de aula.

Dessa discussão, é possível concluir, também, além dos pontos abordados, que se já é difícil atrair professoras para o ensino fundamental, tão, ou mais, difícil é mantê-las na profissão<sup>131</sup>. Portanto, não serão incentivos pelo maior desempenho dos alunos que farão com que melhore a atuação das professoras. Segundo os dados apresentados, é bem provável que o aumento da pressão sobre as professoras possa levá-las a procurar rapidamente escolas mais fáceis de ensinar, ou seja, escolas com menos pobres, ou a deixar a profissão, em busca de outra mais atrativa.

Portanto valorizar a profissão docente, sem conferir a ela um *status* e um reconhecimento social diferenciados, atrelados a salários e a planos de carreira atraentes, será muito difícil contar com um corpo de professoras estável, capaz de implementar programas educacionais de qualidade.

Por economia de espaço, não foram apresentadas aqui as tabelas relativas aos anos finais do ensino fundamental, mas verificou-se que o problema da rotatividade das professoras nessa etapa da educação básica se agrava ainda mais. Nessa outra etapa, além da rotatividade, existe falta de professoras em algumas áreas.

---

<sup>131</sup> A revista *Carta Capital*, n. 749, publicada em 16 de maio de 2013, traz em sua capa o título "Um dia na vida de um professor: entenda por que faltam interessados na profissão", referente a uma reportagem que acompanha o dia a dia de algumas professoras e que mostra as agruras da profissão.

## APÊNDICE C

### Os possíveis impactos do instrumental utilizado nas avaliações externas sobre o currículo e a aprendizagem

Neste texto o objetivo é analisar com mais profundidade a relação entre as Matrizes de Referência e os níveis de proficiência definidos pelas Secretarias de Educação e seus impactos sobre os currículos<sup>132</sup> e a aprendizagem em sala de aula. O estudo refere-se ao Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (Saresp), mas pode ser estendido a outras avaliações educacionais que estão sendo realizadas por diversos sistemas públicos de ensino.

#### *As Matrizes de Referência do Saresp e sua relação com os currículos*

Analisar a relação entre as Matrizes e Currículos é importante para verificar a coerência entre o que é testado pelo Saresp e aquilo que orienta o trabalho das escolas e das professoras.

Pela dificuldade de encontrar dados anteriores a 2007, para tratar das Matrizes de Referência, este trabalho se concentrará em analisar os documentos produzidos a partir desse ano.

Os currículos em São Paulo passaram por reformulação em dois momentos. No primeiro, durante a gestão de Maria Helena Guimarães Castro, foram produzidas as Orientações Curriculares para os anos iniciais do ensino fundamental<sup>133</sup>, para as áreas de Língua Portuguesa e Matemática (SÃO PAULO, 2008a). No segundo, durante a gestão de Paulo Renato Souza, foram produzidos os Currículos de Ciências Humanas e suas Tecnologias (SÃO PAULO, 2010a), Ciências da Natureza e suas Tecnologias (SÃO PAULO, 2010b), Linguagens, Códigos e suas Tecnologias (SÃO PAULO, 2010c) e Matemática e suas Tecnologias (SÃO PAULO, 2010d), destinados aos anos finais do ensino fundamental e ao ensino médio. Apesar das mudanças de gestão, esse trabalho foi coordenado, entre outros, por Maria Ines Fini e Lino de Macedo, que também fizeram parte da equipe que desenhou o

---

<sup>132</sup> Uma discussão sobre o tema pode ser encontrada em ARROYO (2011).

<sup>133</sup> A rede estadual de São Paulo divide o ensino fundamental em dois ciclos, I e II, correspondendo cada ciclo ao que é nacionalmente conhecido como anos iniciais e finais, denominações que serão usadas neste texto.

Enem, em 1998, quando essa estratégia de agrupar em grandes áreas alguns dos conhecimentos foi utilizada.

As Orientações Curriculares foram publicadas em 2008 e trazem como um de seus elementos importantes a definição de expectativas de aprendizagem, para cada ano escolar, que deveria orientar o projeto pedagógico da escola e os planos de aula das professoras. Os Currículos, publicados dois anos depois, abandonam a expressão *expectativas de aprendizagem* e tratam de conteúdos e habilidades bimestre a bimestre de cada ano escolar dos anos finais do ensino fundamental e do ensino médio.

Em razão da simplificação das análises que serão feitas a seguir, este trabalho vai ater-se apenas às Orientações Curriculares para o 5º ano do ensino fundamental,<sup>134</sup> apresentadas nos Quadros C1 e C2, respectivamente.

---

<sup>134</sup> No ano de publicação das Orientações Curriculares, 2008, a rede estadual paulista ainda não havia adotado o ensino fundamental de nove anos. Assim, apesar de no documento ser feita referência à 4ª série, neste texto será utilizado 5º ano.

### Quadro C1 – Expectativas de aprendizagem para a área de Língua Portuguesa, para o 5º ano do ensino fundamental

Expectativa – que os alunos sejam capazes de:	Atividade	Observar se o aluno...
I ó Participar de situações de intercâmbio oral que requeiram: ouvir com atenção, intervir sem sair do assunto tratado, formular e responder perguntas, justificando suas respostas, explicar e compreender explicações, manifestar e acolher opiniões, argumentar e contra-argumentar.	Roda de curiosidades. Roda de biblioteca. Conversas realizadas a partir de leituras compartilhadas ó coletivas ou em duplas. Discussões relacionadas aos projetos.	Expõe sua opinião sobre o que foi lido, complementa informações com conhecimentos que já possui e ouve os colegas com atenção, tanto nas situações coletivas como nos momentos de trabalho em duplas. Expõe oralmente conteúdos aprendidos durante os projetos, utilizando uma linguagem mais formal. Fundamenta suas ideias não apenas em opiniões pessoais, mas também em informações aprendidas. Refere-se às falas de seus colegas ou da professora, para associar às suas próprias ideias. Sabe contrapor suas ideias às de outros, retomando os argumentos utilizados e rebatendo-os com os seus próprios.
II ó Planejar e participar de situações de uso da linguagem oral, sabendo utilizar alguns procedimentos de escrita para organizar sua exposição.	Atividades de comunicação oral.	Comunica-se com uma linguagem formal, sem ter de, necessariamente, ler. Organiza slides ou cartazes relacionados à sua fala ó sem ser uma repetição dele, mas um complemento.
III ó Apreciar textos literários.	Leitura pela professora. Roda de biblioteca.	Escuta atentamente. Compara textos lidos ou ouvidos. Identifica seus autores e gêneros preferidos, buscando, por conta própria, na sala de leitura ou na própria classe, textos dos quais goste. Faz indicações literárias aos seus colegas, apoiando-se em características da trama, personagens, autor, gênero.
IV ó Selecionar os textos de acordo com os propósitos de sua leitura, sabendo antecipar a natureza de seu conteúdo e utilizando a modalidade de leitura mais adequada.	Leitura pelo aluno.	Utiliza títulos, subtítulos, sumários ou índices, para descartar textos que não interessam aos seus propósitos. Faz uma leitura global para separar o que pode lhe interessar. Sabe dizer por que escolhe ou descarta um texto/portador, apoiando-se em informações do conteúdo do texto, do seu portador ou do gênero.
V ó Utilizar recursos para compreender ou superar dificuldades de compreensão durante a leitura (pedir ajuda aos colegas e à professora, reler o trecho que provoca dificuldades, continuar a leitura com a atenção de que o próprio texto permita resolver as dúvidas ou consultar outras fontes).	ó	Pede ajuda a colegas e à professora, relê o trecho que provoca dificuldades, continua a leitura com a intenção de que o próprio texto permita resolver as dúvidas ou consulta outras fontes como dicionário ou glossário.
VI ó Reescrever e/ou produzir textos de autoria, utilizando procedimentos de escritor: planejar o que vai escrever, considerando a intencionalidade, o interlocutor, o portador e as características do gênero; fazer rascunhos; reler o que está escrevendo, tanto para controlar a progressão temática quanto para melhorar outros aspectos ó discursivos ou notacionais ó do texto.	Produção de texto pelo aluno.	Planeja o que vai escrever, escolhendo se os propósitos de seu texto serão atingidos e se a linguagem está adequada; faz rascunhos; relê o que escreve e altera quando não se dá por satisfeito.
VII ó Revisar textos (próprios e de outros), em parceria com os colegas, assumindo o ponto de vista do leitor, com a intenção de evitar repetições desnecessárias (por meio de substituição ou uso de recurso de pontuação); evitar ambiguidades, articular partes do texto, garantir concordância verbal e nominal.	ó	Participa das discussões em torno dos textos, propondo mudanças, e justifica suas propostas remetendo-se ao provável leitor. Propõe substituição de palavras repetidas; identifica problemas de concordância e procura solucioná-los.
VIII ó Revisar textos (próprios e de outros), do ponto de vista ortográfico.	Revisão de textos.	Fica atento aos aspectos ortográficos trabalhados em classe desde o 3º ano.

Fonte: SÃO PAULO, 2008a, com elaboração do autor.



**Quadro C2 – Expectativas de aprendizagem para a área de Matemática, para o 5º ano do ensino fundamental**

Conteúdos	Expectativas de aprendizagem
I ó Números	Compreender e utilizar regras do sistema de numeração decimal; reconhecer e representar números racionais; explorar diferentes significados das frações em situações-problema: parte-todo, quociente e razão; escrever e comparar números racionais de uso frequente, nas representações fracionária e decimal; identificar e produzir frações equivalentes.
II ó Operações	Compreender os diferentes significados das operações envolvendo números naturais; resolver adições e subtrações com números naturais, por meio de estratégias pessoais e do uso de técnicas operatórias convencionais; compreender diferentes significados da adição e da subtração, envolvendo números racionais escritos na forma decimal; resolver operações de adição e subtração de números racionais na forma decimal, por meio de estratégias pessoais e do uso de técnicas operatórias convencionais; resolver problemas que envolvem o uso da porcentagem no contexto diário, como 10%, 20%, 25%, 50%.
III ó Espaço e forma	Interpretar e representar a posição ou a movimentação de uma pessoa ou objeto no espaço e construir itinerários; reconhecer semelhanças e diferenças entre poliedros; identificar elementos como faces, vértices e arestas de poliedros; identificar semelhanças e diferenças entre polígonos, usando critérios como número de lados, número de ângulos, eixos de simetria, rigidez; compor e decompor figuras planas; ampliar e reduzir figuras planas.
IV ó Grandezas e medidas	Utilizar unidades usuais de tempo e temperatura em situações-problema; utilizar o sistema monetário em situações-problema; utilizar unidades usuais de comprimento, massa e capacidade em situações-problema; calcular perímetros de figuras; calcular área de retângulos ou quadrados; utilizar medidas como $\text{cm}^2$ , $\text{m}^2$ , $\text{km}^2$ , e alqueire.
V ó Tratamento da informação	Resolver problemas com dados apresentados de maneira organizada, por meio de tabelas simples, gráficos de colunas, tabelas de dupla entrada e gráficos de barras; ler informações apresentadas de maneira organizada por meio de gráficos de linha e de setor; construir gráficos e tabelas com base em informações contidas em textos jornalísticos, científicos ou outros; identificar as possíveis maneiras de combinar elementos de uma coleção e de contabilizá-las por meio de estratégias pessoais; utilizar a noção de probabilidade em situações-problema simples.

Fonte: SÃO PAULO, 2008a, com elaboração do autor.

É importante notar a diferença de formulação das expectativas de aprendizagem para as duas áreas. Para Língua Portuguesa, associam-se as expectativas com estratégias de avaliação, expressa no conteúdo da coluna "Observar se o aluno ...". O mesmo acontece com a área de Matemática. Além disso, enquanto em Língua Portuguesa apresentam-se expectativas mais gerais, as de Matemática estão ancoradas em conteúdos da área, e as habilidades aparecem de forma muito restrita como, por exemplo, "resolver problemas que envolvem o uso da porcentagem no contexto diário, como 10%, 20%, 25%, 50%", conforme o conteúdo II do Quadro C1. Apesar de não ser objetivo aprofundar essa análise, é possível prever alguma dificuldade para a professora que atua no 5º ano do ensino fundamental, trabalhando com as duas áreas de conhecimento e se deparando com formulações diferentes para elaborar seu plano de ensino e orientar suas atividades em sala de aula.

As atuais Matrizes de Referência do Saresp foram reformuladas em 2008, na mesma época da elaboração das Orientações Curriculares, trabalho também coordenado por Lino de Macedo e Maria Inês Fini. O esquema teórico indicado como base para a elaboração das Matrizes é complexo e parece gerar mais confusão que esclarecimento. Os documentos que as apresentam definem as Matrizes como balizadas por um conjunto de três grupos de Competências Cognitivas, que estariam ligadas às competências do sujeito: competências para observar, referindo-se a esquemas presentativos ou representativos; competências para realizar, que se caracterizam pela capacidade de resolver as tarefas propostas pelo teste; competências para compreender, que implicam o uso de esquemas operatórios (SÃO PAULO, 2008b, 2009f). Para cada um desses grupos, foi indicado um conjunto de habilidades para representá-las. Essas Competências Cognitivas e as habilidades a elas agregadas serviram, segundo o texto, de referência para a construção dos descritores das habilidades a serem testadas. O texto confunde o leitor, ao chamar esses descritores de habilidades simplesmente de habilidades e ao agrupar os conjuntos deles em outras novas competências chamadas de Competência de Área, aparentemente aquelas que seriam próprias a cada área do conhecimento, cujo número varia por área e ano escolar. Por serem diferentes das Competências do Sujeito e as habilidades a elas associadas, as novas Competências de Área e suas habilidades são expressas de outra forma e serão aquelas que orientarão a construção dos itens. Outra confusão ocorre quando as Competências de Área são agrupadas por Tema, que nas Orientações Curriculares são chamados de Conteúdos.

Um destaque deve ser dado ao trecho no qual se diz que

[...] muitas competências e habilidades indicadas na Proposta Curricular, embora importantes para o desenvolvimento dos alunos e para o trabalho em sala de aula, não foram incluídas nas Matrizes, pois não são passíveis de ser avaliadas em instrumentos formais de provas realizadas em larga escala (SÃO PAULO, 2009f, p. 20).

Afirmações como essas, apesar de evidentes, aparecem como um mantra não somente no Saresp, mas também em outros sistemas avaliativos estaduais, na Prova Brasil e no Saeb. Essas afirmações remetem ao óbvio de que outras coisas mais importantes acontecem na escola, e que os sistemas avaliativos não são capazes de medir todas elas. Apesar disso, praticamente todos os sistemas educacionais ignoram o fato no momento de divulgar os resultados dos testes avaliativos e nas análises realizadas. Esse é o caso dos testes da rede estadual paulista.

As Matrizes englobam as áreas de conhecimento de Língua Portuguesa e de Matemática, para os 5º, 7º e 9º anos do ensino fundamental e 3ª série do ensino médio; de Ciências, para os 7º e 9º anos do ensino fundamental; de Química, Física e Biologia, para a 3ª série do ensino médio; de História e Geografia, para os 7º e 9º anos do ensino fundamental e para a 3ª série do ensino médio.

Para se ter uma ideia geral de como está distribuído, por ano escolar, o conjunto de habilidades presentes nas Matrizes de Referência para as diferentes áreas de conhecimento, veja-se o Quadro C3.

**Quadro C3 – Síntese com o número de habilidades para as áreas de conhecimento testadas pelo Saresp, por ano escolar**

Área do conhecimento	Ano escolar/número de habilidades			
	5º ano EF	7º ano EF	9º ano EF	3ª série EM
Língua Portuguesa	40	40	40	50
Matemática	30	38	45	38
Ciências	ó	36	36	ó
Química	ó	ó	ó	57
Física	ó	ó	ó	47
Biologia	ó	ó	ó	50
Geografia	ó	38	38	49
História	ó	34	40	50

Fonte: SÃO PAULO, 2008b, 2009f, com elaboração do autor.

Analisando-se o Quadro C3, percebe-se que existe um equilíbrio, dentro de um mesmo ano escolar, do número de habilidades testadas por área do conhecimento. Além disso, parece que se buscou abranger grande parte do currículo. Destaca-se, também, que para o 5º ano do ensino fundamental, só foram desenvolvidas matrizes para Língua Portuguesa e Matemática.

Para a área de Língua Portuguesa, diferentemente do que ocorre com as Matrizes do Saeb, as do Saresp envolvem competências e habilidades relativas à produção de texto além das relativas à leitura. As Competências por Área são agrupadas em três grupos: leitura de gêneros não literários (variando de acordo com o ano escolar); leitura de gêneros literários (com recomendação de autores); produção de textos. Para a área de Matemática, cada Competência de Área é associada a um dos conteúdos referidos nas orientações curriculares, chamados nas Matrizes de *temas*: números, operações e funções; espaço e forma; grandezas e medidas; tratamento de informação.

Para ilustrar melhor a estrutura das Matrizes de Referência, foram elaborados os Quadros C4 e C5, para o 5º ano do ensino fundamental, nas áreas de Língua Portuguesa e Matemática, respectivamente. Como as Matrizes são extensas, escolheram-se, aleatoriamente, para cada área do conhecimento, duas competências de área e três habilidades.

**Quadro C4 – Exemplo de habilidades do Saresp para o 5º ano do ensino fundamental, agrupados por situação e Competências de Área para Língua Portuguesa**

Situação	Competência de Área	Habilidades
Leitura de gêneros não literários	1 ó Reconstrução das condições de produção e recepção de textos.	H1 ó Identificar a finalidade de um texto, mobilizando conhecimentos prévios sobre o formato do gênero, tema ou assunto principal. H2 ó Identificar os possíveis elementos constitutivos da organização interna dos gêneros não literários: histórias em quadrinhos, regulamentos, receitas, procedimentos, instruções para jogos, cardápios, indicações escritas em embalagens, verbetes de dicionário ou enciclopédia, textos informativos de interesse escolar, curiosidades (você sabia?), notícias, cartazes informativos, folhetos de informação, cartas pessoais ou bilhetes. H3 ó Identificar o público-alvo de um texto, considerando o uso de expressão coloquial, jargão, gíria ou falar regional.
	2 ó Reconstrução dos sentidos dos textos.	H4 ó Identificar o sentido denotado de vocábulo ou expressão utilizados em segmento de um texto, selecionando aquele que pode substituí-lo por sinônimo no contexto em que ele se insere. H5 ó Localizar item de informação explícita, posicionando em segmento inicial de um texto, considerando um único critério para recuperar a informação (o que, quem, quando, onde, como, por que). H6 ó Localizar item de informação explícita, com base na compreensão global de um texto. [...]

Fonte: SÃO PAULO, 2008b, 2009f, com elaboração do autor.

**Quadro C5 – Exemplo de habilidades do Saresp para o 5º ano do ensino fundamental, agrupados por temas e Competências de Área para Matemática**

<b>Tema</b>	<b>Competência de Área</b>	<b>Habilidades</b>
Números, operações, funções	1 ó Desenvolver o raciocínio quantitativo e o pensamento funcional, isto é, o pensamento em termos de relações e a variedade de suas representações, incluindo as simbólicas, as algébricas, as gráficas, as tabulares e as geométricas.	H1 ó Identificar a localização de números naturais na reta numérica. H2 ó Relacionar a escrita numérica às regras do sistema posicional de numeração. H3 ó Escrever um número natural pela sua decomposição em forma polinomial. [...]
Espaço e forma	2 ó Compreender as propriedades dos objetos e sua posição relativa e desenvolver o raciocínio espacial por meio de construções e de formas.	H17 ó Descrever a localização e a movimentação de pessoas ou objetos no espaço, em diversas representações gráficas, dando informações sobre pontos de referencia e utilizando o vocabulário de posição (direita/esquerda, acima/abaixo, entre, em frente/atrás). H18 ó Identificar formas geométricas tridimensionais, como esfera, cone, cilindro, cubo, pirâmide, paralelepípedo, ou formas bidimensionais, como quadrado, triângulo e círculo, sem o uso obrigatório da terminologia convencional. H19 ó Identificar semelhanças e diferenças entre polígonos, usando critérios como números de lados, número de ângulos, eixos de simetria e rigidez, sem o uso obrigatório da terminologia convencional. [...]

Fonte: SÃO PAULO, 2008b, 2009f, com elaboração do autor.

Comparando os Quadros C4 e C5, observa-se que as competências de área para Língua Portuguesa são mais gerais, e que as de Matemática são mais ligadas aos conteúdos, quase que como uma síntese das habilidades que agrupam.

Quando se comparam as Matrizes com as Orientações Curriculares, observa-se que para a área de Matemática existe uma proximidade muito grande entre as duas. Se isso, por um lado, mostra uma coerência entre os dois instrumentos, por outro, traz preocupações na medida em que o ensino dessa área poderá ficar restrito a seus conteúdos.

No caso de Leitura, a situação é distinta. As expectativas de aprendizagens definidas pelas Orientações são muito mais amplas e estabelecem uma relação íntima entre leitura e escrita, uma complementando a outra. Além disso, preveem a observação do desenvolvimento do aluno no bojo de um conjunto de atividades sociais que envolvem o próprio aluno, seus colegas e a professora. Quando se analisa a Matriz de Leitura, percebe-se que, apesar de estarem presentes competências bastante amplas, que poderiam realizar as ligações propostas pelas orientações, as habilidades se restringem a aspectos operacionais da Leitura. Para a professora, pode-se passar a impressão de que ela deve centrar seus esforços nos aspectos funcionais da Leitura e não em seu caráter social, como pregam as Orientações. Nesse sentido, a prática do uso das Matrizes para orientar o ensino, dada a ênfase crescente que o Saresp adquire, poderia significar um estreitamento do currículo.

Se o objetivo dos testes é verificar o desenvolvimento da aprendizagem dos alunos, e se essa aprendizagem está ligada diretamente às expectativas relacionadas nas Orientações, as

expectativas e as habilidades testadas deveriam ter alguma relação entre si. Para verificar como isso ocorre na rede estadual de São Paulo, será feita uma análise tomando como base o 5º ano do ensino fundamental.

Para Língua Portuguesa observa-se, analisando-se o Quadro C1, que, apesar das expectativas serem bastante genéricas, quando se compara a expectativa IV, é possível perceber que ela será atingida se as habilidades testadas apresentadas no Quadro C4 forem atingidas. Mas também é claro que a expectativa IV exige muitas outras habilidades que não necessariamente estão presentes nas Matrizes. Dessa forma, as Orientações Curriculares de Língua Portuguesa privilegiam um currículo amplo, com bastante espaço para a professora escolher as melhores estratégias de desenvolvimento para sua turma.

No caso de Matemática a correspondência entre as expectativas e as habilidades é muito grande. Por exemplo, tomem-se algumas das expectativas III no Quadro C2. Elas aparecem novamente como as habilidades H17 a H19 nas Matrizes de Referência, mostrando um currículo mais restrito, que trata do essencial e que permite maior aprofundamento dos fenômenos escolhidos, além de ser centrado nos conteúdos. Nesse caso, se as habilidades forem atingidas, as expectativas também o serão, numa correspondência quase biunívoca.

Portanto, para a mesma professora que atua no 5º ano do ensino fundamental, são apresentadas duas forma de tratar as expectativas de aprendizagem e sua relação com as Matrizes.

### ***Os níveis de proficiência do Saresp e seus impactos sobre a aprendizagem***

A Secretaria estabeleceu, baseada nas Orientações Curriculares e nos resultados do Saresp, diferentes níveis de proficiência e os relacionou com a escala de proficiência. O resultado dessa operação é apresentado no Quadro C6.

#### **Quadro C6 – Expectativas de aprendizagem definidas a partir dos níveis de proficiência do Saresp**

<b>Língua Portuguesa (Leitura)</b>			
Níveis	5º ano EF	9º ano EF	3ª série EM
Abaixo do Básico	< 150	< 200	< 250
Básico	Entre 150 e < 200	Entre 200 e < 275	Entre 250 e < 300
Adequado	Entre 200 e < 250	Entre 275 e < 325	Entre 300 e < 375
Avançado	> 250	> 325	> 375
<b>Matemática</b>			
Níveis	5º ano EF	9º ano EF	3ª série EM
Abaixo do Básico	< 175	< 225	< 275
Básico	Entre 175 e < 225	Entre 225 e < 300	Entre 275 e < 350
Adequado	Entre 225 e < 275	Entre 300 e < 350	Entre 350 e < 400
Avançado	> 275	> 350	> 400

Fonte: SÃO PAULO, 2009a, p. 2.

Para compreender melhor o significado do que é expresso no Quadro C6, é necessário relacionar suas informações com as descrições da escala de proficiência em Leitura, que será feita em primeiro lugar, e Matemática, feita a seguir. Busca-se, com isso, encontrar um significado pedagógico para os intervalos numéricos, de forma a verificar se é possível orientar as escolas e suas professoras com relação ao trabalho pedagógico a ser desenvolvido, para que o objetivo a ser alcançado não seja superar números, mas melhorar aprendizagens.

Em Língua Portuguesa (Leitura), a descrição da escala foi feita para intervalos de  $\frac{1}{2}$  desvio padrão, ou 25 pontos, a partir do ponto 125, obtendo-se 13 pontos distintos (SÃO PAULO, 2009b). Será feita uma análise, novamente, tomando por base o 5º ano do ensino fundamental.

Como a descrição da escala é bem longa, optou-se por analisar o ponto 225, que se encontra no meio do Nível Adequado, e o ponto 275, que se encontra no Nível Avançado conforme o Quadro C7. Algumas descrições foram negritadas e sublinhadas e serão discutidas mais à frente.

Se uma das utilidades da descrição da escala é poder oferecer referências para que as escolas possam rever seus projetos pedagógicos e para que a professora faça o mesmo com seu plano de curso, o conteúdo do Quadro C7 demonstra que essa descrição pode trazer informações equivocadas para essas revisões. Isso será analisado inicialmente para o ponto 225.

Tomando-se, por exemplo, a descrição da escala referente a **Comparar dois textos identificando o gênero e o assunto de cada um**, indicada para o Tema 4, pode dar a falsa impressão de que somente essa atividade é importante quando se está frente a dois textos. Portanto, pode ocorrer de as professoras acabarem treinando seus alunos para realizar essa tarefa, para com isso trazer sua escola para o nível adequado, mesmo que a descrição informe que os alunos também precisam dominar as habilidades que estão nos pontos mais baixos da escala. Além disso, em outras descrições da escala apresentadas no Quadro C7, as informações são bastante superficiais, como **Identificam o suporte de texto institucional**, indicada para o Tema 1, ou **Identificam os argumentos utilizados para convencer o interlocutor sobre determinado fato em carta familiar ou artigo de opinião**, indicada para o Tema 3. Essas e outras descrições podem trazer mais confusão do que ajuda de tão genérica que são. Por fim, muito se criticou sobre a ênfase conteudista dada ao ensino, mas parece que o que se está indicando para as professoras, a partir da descrição da escala, é uma lista de atividades a serem cumpridas para alcançar determinada pontuação considerada adequada. A ênfase deixa de ser a aprendizagem e passa a ser a realização de atividades.

Da mesma forma, as descrições associadas ao ponto 275 podem também trazer sinais confusos para as professoras, com o agravante de apresentar uma quantidade menor de descrições, conforme se observa no Quadro C7. Isso tem relação com a dificuldade para elaborar itens nos pontos mais altos ou mais baixos da escala, como já foi comentado quando se discutiu a escala de proficiência no Capítulo 1. Como o documento que traz as descrições da escala não explicita isso, podem surgir dúvidas nas professoras sobre o que pode ser feito para mudar de patamar.



**Quadro C7 – Descrição da escala de Língua Portuguesa (Leitura)  
para os pontos 225 e 275**

<i>Para cada ponto, os alunos do 5º ano do ensino fundamental, além das habilidades descritas nos níveis anteriores, também...</i>	
<b>Ponto 225 (Nível Adequado)</b>	<b>Ponto 275 (Nível Avançado)</b>
<i>Tema 1 – Reconstrução das condições de produção e recepção de textos</i>	
<ul style="list-style-type: none"> <li>Identificam o possível suporte de texto institucional.</li> <li><b>Identificam os elementos constituintes de história em quadrinhos ou de entrevista.</b></li> <li>Identificam: o interlocutor provável de propaganda institucional, pela análise do uso de pronome de tratamento ou por seu assunto, ou o sentido do uso de pergunta retórica em propaganda comercial.</li> </ul>	<ul style="list-style-type: none"> <li>Identificam os possíveis elementos constitutivos da <b>organização interna de um bilhete.</b></li> </ul>
<i>Tema 2 – Reconstrução dos sentidos do texto</i>	
<ul style="list-style-type: none"> <li>Identificam, em segmento de artigo de divulgação, o sentido denotado de vocábulo, selecionando aquele que pode substituí-lo no contexto em que se insere, ou de vocábulos de uso comum (termos técnicos da área científica);</li> <li>Identificam o sentido de palavra gramatical em segmento de notícia, considerando o assunto tratado no texto.</li> <li>Localizam itens explícitos de informação em parte de artigo de divulgação científica e de reportagem jornalística.</li> <li>Localizam informação explícita em verbete de enciclopédia, considerando o assunto tratado no texto.</li> <li>Localizam, em listas de instruções ou em artigos de divulgação científica, itens de informação relativos a diferentes tópicos, identificando sequências (antes/depois) e realizando inferências básicas.</li> <li>Organizam, em sequência, itens de informação explícita, distribuídos ao longo do cardápio ou de artigo de divulgação.</li> <li>Estabelecem relação entre imagem e texto escrito, em verbete de enciclopédia, para localizar uma informação.</li> <li><b>Inferem o tema de história em quadrinhos ou o assunto principal de carta.</b></li> </ul>	<ul style="list-style-type: none"> <li>Inferem o <b>sentido de expressão utilizada em artigo de divulgação científica.</b></li> <li>Inferem o <b>assunto principal de notícia, com base em informações contidas em títulos ou subtítulos ou corpo do texto.</b></li> </ul>
<i>Tema 3 – Reconstrução da textualidade</i>	
<ul style="list-style-type: none"> <li>Identificam os argumentos utilizados para convencer o interlocutor sobre determinado fato, em carta familiar ou artigo de opinião.</li> <li>Comparam os argumentos utilizados por diferentes interlocutores sobre um mesmo fato ou notícia.</li> <li>Identificam o sentido de uso de marca discursiva de temporalidade no encadeamento dos fatos em fábula.</li> <li>Identificam o efeito de sentido produzido em verbete histórico, pelo uso de marcas discursivas de temporalidade (coesão sequencial) no encadeamento dos fatos apresentados.</li> <li><b>Estabelecem relações coesivas entre segmentos de texto: propondo uma substituição de pronome pessoal por grupo nominal correspondente ou identificando o referente de um pronome de tratamento em carta familiar.</b></li> <li><b>Avaliam a paráfrase de parte do texto que apresenta relação de causalidade, em artigo de divulgação científica.</b></li> </ul>	<ul style="list-style-type: none"> <li>Estabelecem <b>relações de causa/consequência entre informações implícitas em texto de instruções.</b></li> <li><b>Avaliam o caráter opinativo de um trecho.</b></li> </ul>
<i>Tema 4 – Recuperação da intertextualidade e estabelecimento de relações entre textos</i>	
<ul style="list-style-type: none"> <li>Comparam dois textos identificando o gênero e o assunto de cada um.</li> </ul>	(Não apresentou-se descrição para o nível)
<i>Tema 5 – Reflexão sobre os usos da língua falada e escrita</i>	
<ul style="list-style-type: none"> <li>Identificam <b>padrões ortográficos na escrita, pela comparação de processos de prefixação, com base na correlação definição/exemplo.</b></li> <li>Identificam a frase do texto que apresenta um <b>substantivo no diminutivo.</b></li> </ul>	<ul style="list-style-type: none"> <li>Identificam o segmento de uma carta que é marcado por <b>expressão tipicamente familiar.</b></li> <li>Identificam o <b>sentido de formas verbais flexionadas no modo imperativo em texto instrucional.</b></li> </ul>
<i>Tema 6 – Compreensão de textos literários</i>	
<ul style="list-style-type: none"> <li>Identificam o sentido conotado de expressão utilizada em verso de poema, selecionando aquela que pode substituí-la no contexto em que se insere.</li> <li>Identificam informação subtendida em poema, considerando os recursos semânticos utilizados.</li> <li>Identificam o efeito de sentido produzido em poema, pela exploração de recursos morfossintáticos.</li> <li>Identificam o segmento de poema narrativo em que o enunciador determina o desfecho do enredo.</li> <li>Organizam em sequência os principais episódios de fábula ou de conto.</li> <li>Identificam o enunciador do discurso direto, em conto em que há vários enunciadores em diálogo simultâneo.</li> <li>Identificam uma interpretação adequada para poema, considerando o uso de determinada expressão.</li> <li>Identificam a moral da fábula.</li> <li>Inferem o sentido de humor do texto, considerando o uso intencional de ambiguidades (palavras, expressões, recursos iconográficos) em anedota ou conto.</li> </ul>	<ul style="list-style-type: none"> <li>Identificam o sentido conotado de expressão utilizada em poema, selecionando aquela que pode substituí-la no contexto em que ela se insere.</li> <li>Identificam o efeito de sentido produzido em poema pelo uso de repetição dos mesmos versos ao longo do texto.</li> <li>Identificam os diferentes lugares em que se passa o enredo de conto.</li> <li>Inferem o efeito de humor em trecho de romance, justificando o sentido de ambiguidade produzida pelo uso de expressão.</li> </ul>

Fonte: SÃO PAULO, 2009b, com elaboração do autor.

Outra perspectiva de análise é pinçar, para cada tema, descrições de habilidades próximas, com o intuito de verificar quais sinais são dados, que permitam algum tipo de análise do que a escola e suas professoras podem realizar em relação ao ensino, para permitir a passagem do Nível Adequado para o Avançado. Para isso, serão utilizadas apenas as descrições destacadas em negrito e sublinhadas no Quadro C7, pinçadas de cada tema.

Inicialmente, tome-se as descrições destacadas para o Tema 6. Elas são exatamente iguais, variando apenas sua extensão: para o ponto 225, a ênfase é um verso do poema, enquanto que para o ponto 275 a ênfase é todo o poema. Que orientação uma informação como essa pode dar à escola interessada em mudar de patamar? Deixar de trabalhar com alguns versos e passar a trabalhar com todo o poema? Qual o sentido pedagógico e prático disso?

Quanto ao Tema 5, que trata da reflexão sobre os usos da língua, as descrições destacadas remetem à identificação de substantivos no diminutivo, expressão tipicamente familiar (qual o significado dessa expressão?) e o sentido de formas verbais flexionadas no modo imperativo. Mas só isso deve ser enfatizado pela professora? Quanto a identificar padrões ortográficos na escrita pela comparação de processos de prefixação com base na correlação definição/exemplo, a descrição por si só não indica qual habilidade o item está procurando medir. Nesse caso, para melhorar a compreensão das professoras é necessário ver o item que deu origem a essa descrição, algo que nem sempre é possível dado as limitações utilizadas na metodologia dos testes em larga escala, conforme discutido no Capítulo 1.

Para o Tema 4, apenas uma descrição aparece no Ponto 225 e não existe nenhuma no Ponto 275 da escala. Assim, uma possível mensagem a ser transmitida às escolas e às professoras é que a prática de relacionar textos entre si poderia ser deixada de lado, pois sua influência para atingir os níveis Adequado e Avançado seria pequena, quando comparado com as habilidades dos outros temas.

As análises dos outros temas e as descrições destacadas no Quadro C7 seguem no mesmo caminho discutido aqui: ou dificultam sua compreensão ou apresentam aspectos muito limitados das várias possibilidades que envolvem a Leitura.

Assim como foi feito para Leitura, o mesmo será feito para Matemática. Para essa área, a descrição da escala foi feita para intervalos de  $\frac{1}{2}$  desvio-padrão, ou 25 pontos, a partir do ponto 150, obtendo-se 13 pontos distintos (SÃO PAULO, 2009c). É importante lembrar que a escala de Matemática é diferente da escala de Leitura, não existindo, então, nenhuma relação entre os valores numéricos das duas escalas. Dessa forma, o fato de 125 ser o ponto mais baixo da descrição da escala de Leitura não significa que a aprendizagem dos alunos

nessa área seja menor que em Matemática, em que o ponto mais baixo é 150, ou, então, que dois valores numéricos iguais signifiquem aprendizagens iguais para as duas áreas. Como foi feito para a área de Leitura, optou-se por analisar dois pontos: um que estivesse contido no Nível Adequado e outro no Nível Avançado, para o 5º ano do ensino fundamental. Os pontos escolhidos foram 225 e 300. Apresenta-se, a seguir, o Quadro C8 para os pontos 225 e 300. Para a construção deste Quadro, com o intuito de facilitar a análise, as descrições foram agrupadas por conteúdos, conforme as Orientações Curriculares de Matemática apresentadas no Quadro C2.

A primeira observação a ser feita é com relação ao número de descrições da área de Matemática, inferior ao da área de Leitura. Isso provavelmente deve ter relação com a qualidade dos itens da primeira área, pois, como foi destacado quando se discutiu a descrição das escalas, existem alguns critérios que devem ser observados para escolher os itens a serem descritos. E essa quantidade tem ligação estreita com sua capacidade de discriminação, ou seja, de distinguir diferentes proficiências. Assim, quanto menor essa capacidade, menor a quantidade de itens disponíveis para serem utilizados na descrição. Esse fato aliado à dificuldade de se elaborar itens de proficiências mais elevadas levou à existência de apenas seis descrições para o ponto 300, um dos pontos que caracterizaria o Nível Avançado. Portanto, a descrição da escala de Matemática oferece, ainda, menos informações para a escola e suas professoras, apesar de, diferentemente do que ocorreu com Leitura, algumas informações serem mais precisas.

As descrições agrupadas no conteúdo operações mostram claramente que no ponto 225 os alunos já desenvolveram maturidade para resolver problemas que envolvem as quatro operações, utilizando números racionais e demonstrando conhecer alguns de seus significados, o que está em acordo com as orientações curriculares da área de Matemática. Para esse ponto, os conteúdos operações e espaço e forma revelam, também, que existe uma relação estreita com as orientações, apesar de algumas de suas habilidades não aparecerem nas descrições, nesse caso, dificultando a relação entre as expectativas de aprendizagens estabelecidas nas Orientações e aquilo que foi estabelecido como Nível Adequado. Isso é mais crítico com relação aos conteúdos grandezas e medidas e tratamento de informações, em que apenas uma descrição de cada conteúdo foi realizada.

**Quadro C8 – Descrição da escala de Matemática para os pontos 225 e 300**

Conteúdos	Nível/Ponto da escala/Descrições	
	Nível adequado	Nível avançado
	225	300
Números	<ul style="list-style-type: none"> <li>• Identificam: o algarismo de dezena em número de quatro algarismos; o total de dezenas em um número de três algarismos; um número com sua decomposição pelas regras do sistema de numeração decimal; um número em que é dado o algarismo de centena; [reconhecem] o menor entre números de quatro algarismos com zeros intercalados.</li> <li>• Identificam a reta que melhor representa a posição 1,2 em uma reta numerada, a partir de 0 na escala 1 a 1.</li> <li>• Resolvem problema envolvendo: o cálculo de <math>\frac{2}{3}</math> de um número.</li> <li>• Reconhecem o menor entre os números de quatro algarismos com zeros intercalados.</li> </ul>	<ul style="list-style-type: none"> <li>• Identificam um número a partir de informações sobre os algarismos que o compõem e sobre sua posição, de acordo com as regras do sistema de numeração decimal.</li> </ul>
Operações	<ul style="list-style-type: none"> <li>• Identificam: 50% com <math>\frac{1}{2}</math>; a sequência de figuras que tem 25% coloridas; a fração correspondente à parte de um todo.</li> <li>• Resolvem problema envolvendo: a diferença entre dois números racionais apresentados na sua forma decimal; a subtração com significado de comparação, com números de até três algarismos.</li> <li>• Resolvem problema envolvendo a adição com significado de juntar, com números de até quatro algarismos, e identifica o algarismo da unidade.</li> <li>• Resolvem problema envolvendo: a multiplicação (saias e blusas); multiplicação com significado de proporcionalidade, por números de dois algarismos, e escrita decimal de cédulas e moedas; multiplicação com significado de adição de parcelas iguais, com escrita decimal de cédulas e moedas; multiplicação com significado de configuração retangular, com números de três e de dois algarismos.</li> <li>• Resolvem problema envolvendo: o quociente entre dois números naturais (com respostas do tipo <math>\frac{a}{b}</math> inteiros e sobram <math>y</math>); a divisão com significado de repartir, com dividendo de dois algarismos e divisor com um algarismo; a divisão com significado de proporcionalidade, com divisor de um algarismo.</li> <li>• Resolvem problema envolvendo o conceito de porcentagem</li> </ul>	<ul style="list-style-type: none"> <li>• Resolvem problema envolvendo cálculo de porcentagem (25% ou 75%).</li> </ul>
Espaço e forma	<ul style="list-style-type: none"> <li>• Identificam: a pessoa que está em frente a outra, em um desenho que mostra a disposição circular de um grupo de pessoas; a posição à direita diante do desenho que mostra a disposição de poltronas em uma sala.</li> <li>• Identificam as formas de losango, triângulo, hexágono e pentágono, como sendo as de pipas apresentadas por desenhos,</li> <li>• Relacionam planificação de um cilindro com seu nome,</li> </ul>	
Grandezas e medidas	<ul style="list-style-type: none"> <li>• Leem horas e minutos em relógios analógicos.</li> </ul>	<ul style="list-style-type: none"> <li>• Calculam: área de figura em malha quadriculada; perímetro de figura plana em malha quadriculada.</li> </ul>
Tratamento de informações	<ul style="list-style-type: none"> <li>• Resolvem problema envolvendo uma composição de relações, com informações obtidas em tabela e com transformações de cm para m.</li> </ul>	<ul style="list-style-type: none"> <li>• Identificam e usam regularidade apresentada em padrão geométrico.</li> </ul>

Fonte: SÃO PAULO, 2009c, com elaboração do autor.

Como o Nível Adequado engloba também os pontos 250 e 275, mesmo que as descrições correspondentes fossem aqui apresentadas, ainda assim haveria uma discrepância entre aquilo que as orientações mostram como expectativas e o que a descrição da escala classifica como Nível Adequado. É como se o Saresp indicasse para as professoras que o importante em Matemática é trabalhar os conteúdos de números, operações, espaço e forma, deixando de lado os conteúdos relativos a grandezas e medidas e ao tratamento de informações. As Matrizes acabam assumindo o lugar das orientações curriculares.

Apesar de a análise estar centrada em apenas dois pontos da descrição da escala e centrar-se apenas no 5º ano do ensino fundamental, os aspectos aqui abordados, discutindo a possibilidade de o currículo estar sendo estreitado e substituído pelas Matrizes de Referência, se repetem em outros pontos da escala e em outros anos escolares.

Depois de analisadas, a descrição das escalas e sua relação com as Orientações Curriculares, é possível retornar ao Quadro C6, que indica expectativas de aprendizagens e os níveis de proficiência. Buscou-se atribuir um valor numérico às expectativas, o que de todo não é algo ruim, desde que isso faça algum sentido e, principalmente, que se tenha clareza sobre o público a que essa informação se destina.

Para a Secretaria de Educação, informações sobre a proficiência dos alunos e outras informações sobre a escola podem ser úteis para planejar as ações que visem melhorar a qualidade da educação. Assim, tão importante quanto a proficiência é coletar informações sobre a rotatividade das professoras e seu perfil; se o quadro de professoras está ou não completo; se a região socioeconômica onde a escola está localizada pode favorecer a construção de um indicador que mostre o quanto determinado conjunto de escolas precisa ou não de apoio adicional para se desenvolver, fornecendo indícios sobre o que deve ser feito, ou identificando as escolas que poderiam ser motivo de estudo, para aprender um pouco mais sobre o funcionamento da rede, ou, ainda, verificando a necessidade ou não de capacitações específicas sobre determinados temas. No entanto, a única informação discutida são intervalos numérico e os adjetivos associados a eles. Essas são informações muito superficiais, com baixa possibilidade de fornecer apoio efetivo dentro das necessidades da escola.

Para os gestores das escolas, essa informação é de pouco auxílio. Eles podem, por exemplo, comparar sua escola com outras que atendam população semelhante à população dela e, eventualmente, buscar parcerias com essas escolas, caso identifiquem algum exemplo que tenha alguma perspectiva de possibilitar a melhoria da aprendizagem de seus alunos.

Como as informações pedagógicas são superficiais, talvez os gestores não tenham como mobilizar a comunidade escolar para fazer modificações importantes no projeto pedagógico da escola. Eventualmente, para um determinado grupo de escolas, o fato de elas serem apontadas como aquelas em que o nível de proficiência médio de seus alunos foi classificado com Abaixo do Básico, pode, por si só, ser o combustível para algumas transformações. No entanto, a profundidade dessas transformações depende do perfil da gestão escolar, do corpo docente, do apoio da comunidade escolar e da Secretaria de Educação. Esse processo pode não ser rápido e talvez não tenha fôlego suficiente. Em situações muito críticas, pode ser que essa pressão cause algum efeito, mas não terá muita sustentação, e logo se voltará ao estado inicial. Portanto, é muito difícil que a classificação por níveis de proficiência, sozinha, venha a contribuir para a melhoria da aprendizagem dos alunos, pois, como já comentado, a fragilidade das informações pedagógicas fornecidas é grande. Enfatiza-se aqui que, no caso de uma escola, mudanças não ocorrem somente por conta de melhorias da gestão ou de melhorias no planejamento, mas, principalmente, pela existência de um projeto pedagógico sólido e discutido com a comunidade escolar.

Quanto às professoras, da mesma forma que se afirmou com relação às escolas, que são poucas as informações pedagógicas oferecidas pelos resultados, são poucas as informações que possibilitem modificações nos planejamentos de curso deles. Ao mesmo tempo, podem ocorrer efeitos deletérios nos alunos, se as professoras optarem por atalhos para treiná-los com vistas à realização dos testes. Essa prática pode até surtir algum efeito para a melhoria dos indicadores, mas não favorece uma aprendizagem significativa, e pode acabar afastando a criatividade e a inovação da sala de aula.

Quanto às famílias, a informação sobre o Nível de Proficiência dos alunos da escola que suas crianças e jovens frequentam pouco impacto pode trazer. Se uma provável decisão fosse trocar de escola para uma em que o nível fosse mais elevado, nem sempre isso seria possível, pois dependeria da existência de vagas nessa outra escola e da possibilidade de transporte até ela. Outra decisão possível seria pressionar a escola e a Secretaria de Educação por mudanças, mas para que estas aconteçam são necessárias mobilização e organização. Mesmo que isso aconteça, quem garante que a decisão tomada pela escola não seja a mais fácil, a de ensinar para o teste?

Essa discussão torna-se ainda mais complicada para os sistemas municipais de ensino que participam do Saesp, pois as referências curriculares podem ser totalmente diferentes.

Portanto, apesar de tão propalados, os possíveis efeitos dos resultados dos testes avaliativos são bastante limitados.