

UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
PROGRAMA DE PÓS-GRADUAÇÃO EM EDUCAÇÃO

**LUÍS ANTÔNIO FAJARDO PONTES**

MEDIDAS DE EFICÁCIA ESCOLAR NO CONTEXTO DAS POLÍTICAS BRASILEIRAS  
DE RESPONSABILIZAÇÃO EDUCACIONAL: o caso do Índice de Desenvolvimento da  
Educação Básica, o Ideb, em Minas Gerais

**JUIZ DE FORA**

**2015**

**LUÍS ANTÔNIO FAJARDO PONTES**

MEDIDAS DE EFICÁCIA ESCOLAR NO CONTEXTO DAS POLÍTICAS BRASILEIRAS DE RESPONSABILIZAÇÃO EDUCACIONAL: o caso do Índice de Desenvolvimento da Educação Básica, o Ideb, em Minas Gerais

Tese de Doutorado apresentada ao Programa de Pós-graduação em Educação, área de concentração: Gestão, Políticas Públicas e Avaliação Educacional, da Universidade Federal de Juiz de Fora, como requisito parcial para a obtenção do grau de Doutor.

Orientador: Prof. Dr. Tufi Machado Soares

JUIZ DE FORA

2015

**LUÍS ANTÔNIO FAJARDO PONTES**

**MEDIDAS DE EFICÁCIA ESCOLAR NO CONTEXTO DAS POLÍTICAS  
BRASILEIRAS DE RESPONSABILIZAÇÃO EDUCACIONAL:  
O CASO DO ÍNDICE DE DESENVOLVIMENTO DA EDUCAÇÃO BÁSICA,  
O Ideb, EM MINAS GERAIS**

Tese aprovada como requisito parcial para  
obtenção do título de Doutor no Programa de Pós-  
Graduação em Educação da Faculdade de  
Educação da Universidade Federal de Juiz de Fora,  
pela seguinte banca examinadora:

---

Prof. Dr. Tufi Machado Soares (Orientador)  
Programa de Pós-Graduação em Educação, UFJF

---

Prof. Dr. Eduardo Magrone  
Programa de Pós-Graduação em Educação, UFJF

---

Prof. Dr. Beatriz de Basto Teixeira  
Programa de Pós-Graduação em Educação, UFJF

---

Prof. Dr. Marcelo Tadeu Baumann Burgos  
Programa de Pós-Graduação em Ciências Sociais, PUC-RJ

---

Prof. Dr. Nigel Pelham de Leighton Brooke  
Faculdade de Educação, UFMG

Juiz de Fora, 15 de Maio de 2015

## RESUMO

O presente estudo tem por objetivo analisar alguns mecanismos específicos de aferição da qualidade educacional que usualmente vêm sendo propostos por diversos sistemas educacionais no Brasil e, mais particularmente, nas escolas públicas de Minas Gerais, bem como propor alternativas de seu aperfeiçoamento. Neste sentido, discorre-se primeiramente sobre a questão da eficácia escolar e também sobre a maneira como este tópico vem ganhando uma importância cada vez maior na agenda das reformas educacionais tanto fora quanto dentro do Brasil. A seguir, apresenta-se o caso do Índice Brasileiro de Desenvolvimento da Educação Básica, o Ideb, discorrendo-se sobre a sua metodologia de elaboração e também sobre sua interpretação. Posteriormente, apresentam-se conclusões, específicas para as escolas públicas de Minas Gerais, acerca dos efeitos indesejáveis sobre as metas do Ideb causados, em parte, por uma certa instabilidade dos resultados educacionais, que fazem com que, a cada edição dessa avaliação transversal, ocorram desvios ou flutuações dos resultados em relação a uma tendência de crescimento prevista por um modelo longitudinal hierárquico. Utilizando-se então procedimentos associados a esta mesma metodologia, é proposto um modelo inédito de determinação das metas do Ideb, por meio de uma correção periódica que se faz sobre elas, com base nos resultados atualizados das avaliações, parte delas relacionadas, no caso deste exemplo de Minas Gerais, aos resultados anuais dos testes do Sistema Mineiro de Avaliação da Educação Básica, o SIMAVE, na 8ª série ou 9º ano do Ensino Fundamental. E, finalmente, estudam-se também casos específicos de melhoria significativa da qualidade das decisões sobre o desempenho educacional dos estabelecimentos assim avaliados, com base neste novo modelo.

**Palavras-chave:** Avaliação educacional; Ideb; Eficácia escolar; SIMAVE, Modelos longitudinais lineares hierárquicos aplicados à educação.

## ABSTRACT

This study aims to analyze some specific mechanisms of educational quality measurement which have been proposed by several education systems in Brazil and more specifically in Minas Gerais State public schools, as well as propose some alternatives for their improvement. In this sense, some questions about school effectiveness are proposed, along with some considerations about the way how this topic has acquired an increasing relevance in educational reform agendas both inside and outside Brazil. The specific case of the Brazilian Index of Basic Education Development – originally the *Índice Brasileiro de Desenvolvimento da Educação Básica*, or Ideb – is also presented, together with some considerations about its development methodology and interpretation. Later some specific conclusions about the Ideb and Minas Gerais State public schools are also discussed, related to the undesirable effect that has been detected on their respective Ideb targets, partly caused by a certain instability of educational results, by which it is possible to detect and gauge some deviations or fluctuations of the results in relation to a growth trend model at each year of the historical evaluation series, according to a longitudinal hierarchical model. The latter methodology is also employed in order to propose a new model of Ideb target setting, which is accomplished by their respective periodic correction based on updated results obtained by the same schools in other state standardized tests, such as the case, in Minas Gerais, of those related to the *Sistema Mineiro de Avaliação da Educação Básica* – or Minas Gerais Basic Education Evaluation System, the SIMAVE –, for the 9<sup>th</sup> year or 8<sup>th</sup> grade of Fundamental Level. Finally, the text presents and discusses some specific cases of occurrence of a significant quality improvement related to decisions that can be made about the educational performance of the same educational institutions evaluated according to this model.

**Keywords:** Educational evaluation; Ideb; School effectiveness; SIMAVE, Longitudinal hierarchical models applied to education.

## SUMÁRIO

INTRODUÇÃO.....	8
1. A AVALIAÇÃO DO DESEMPENHO EM ORGANIZAÇÕES PÚBLICAS.....	12
1.1. A necessidade e o funcionamento dos sistemas de avaliação de desempenho.....	12
1.2. Propostas e tentativas de mensuração do desempenho escolar.....	19
2. O MOVIMENTO NORTE-AMERICANO DE REFORMA EDUCACIONAL: SINOPSE DE SUA EVOLUÇÃO, DESAFIOS E ASSOCIAÇÃO COM OS SISTEMAS DE AVALIAÇÃO EM LARGA ESCALA.....	27
2.1. Introdução.....	26
2.2. Breve retrospectiva histórica do contexto educacional norte-americano.....	27
2.3. Críticas e reformas do sistema educacional americano.....	30
2.4. Reformas mais recentes nos EUA.....	33
2.5. Novos temas e tendências de reforma.....	41
2.6. Algumas objeções técnicas feitas ao NCLB.....	45
2.6.1. A imprecisão das medidas.....	46
2.6.1.1. Os problemas das estimativas de modelo agregado.....	48
2.7. A defesa das avaliações.....	53
3. AS AVALIAÇÕES EDUCACIONAIS EM GRANDE ESCALA NO BRASIL.....	57
3.1. Introdução.....	57
3.2. Uma visão preliminar: a expansão da educação pública no Brasil.....	58
3.3. Primeiro período (1990-1997): os alicerces.....	60
3.3.1. O Sistema Brasileiro de Avaliação da Educação Básica.....	61
3.3.2. Escalas de proficiência.....	62
3.3.3. A Teoria da Resposta ao Item (TRI).....	64
3.4. Segundo período (1997–2007): expansão.....	67
3.5. Terceiro período (2007–presente).....	69
3.5.1. O Ideb.....	70
3.5.2. Políticas de prestação de contas.....	72
3.5.3. Disseminação dos sistemas estaduais de avaliação.....	73
3.6. Possíveis progressos e desafios das avaliações no contexto brasileiro.....	74
4. A EFICÁCIA DAS ESCOLAS PÚBLICAS DO SIMAVE: UMA PROPOSTA DE MENSURAÇÃO.....	75

4.1. Introdução.....	75
4.2. Desempenho geral.....	76
4.3. Medidas de status.....	77
4.3.1. O status medido por N1.....	80
4.3.2. O status medido pelas médias de proficiência.....	82
4.3.3. O status medido por N3.....	83
4.4. Medidas de progresso.....	84
4.4.1. O progresso medido pelas variações de N1.....	86
4.4.2. O progresso medido pelas variações das médias de desempenho.....	87
4.4.3. O progresso medido pelas variações de N3.....	87
4.5. Um índice de desempenho para as escolas mineiras.....	88
4.5.1. A busca por um indicador de status.....	88
4.5.1.1. As escolas de maior status.....	92
4.5.1.2. As escolas de menor status.....	94
4.5.1.3. Interpretando os índices de status.....	95
4.5.1.4. Uma análise de cluster dos extremos de desempenho quanto ao status.....	98
4.5.2. A busca por um indicador de progresso.....	101
4.5.2.1. As escolas de maior progresso.....	103
4.5.2.2. As escolas de menor progresso.....	104
4.5.2.3. Interpretando os índices de progresso.....	106
4.5.2.4. Uma análise de cluster dos extremos de desempenho quanto ao progresso.....	108
4.6. Um indicador unificado de desempenho.....	110
4.6.1. As escolas nas extremidades do índice total de desenvolvimento.....	112
4.6.2. As escolas de menor índice total de desenvolvimento.....	114
4.7. Uma análise de cluster dos extremos de desempenho.....	115
4.8. Alguns passos adicionais a serem considerados na investigação da qualidade educacional.....	116
5. O ÍNDICE DE DESENVOLVIMENTO DA EDUCAÇÃO BÁSICA (IDEB): INTERPRETAÇÃO, METAS E RESULTADOS.....	118
5.1. Introdução.....	118
5.2. Considerações gerais sobre o Ideb.....	119
5.2.1. Cálculo e Interpretação do Ideb.....	120
5.3. O comportamento do Ideb no Brasil.....	125
5.3.1. A evolução do Ideb nacional entre 2005 e 2013.....	126

5.4. O Ideb em Minas Gerais.....	132
5.4.1. Perspectivas do cumprimento das metas do Ideb em Minas Gerais.....	135
6. A APLICAÇÃO DE MODELOS LINEARES HIERÁRQUICOS COMO UM MECANISMO DE APERFEIÇOAMENTO DAS METAS DO IDEB.....	138
6.1. Introdução.....	138
6.2. Revisão de conceitos e métodos clássicos de testes de hipóteses.....	139
6.2.1. Os testes t e a ANOVA.....	139
6.2.2. Análise de regressão.....	144
6.3. A modelagem linear hierárquica como uma alternativa para os modelos clássicos dos testes de hipóteses.....	148
6.3.1. O modelo nulo de dois níveis.....	149
6.3.1.1. Vantagens do modelo nulo.....	156
6.3.2. O modelo de nível 1.....	157
6.3.3. O modelo de nível 2.....	162
6.3.4. O caso longitudinal de dois níveis.....	164
7. VOLATILIDADE DOS RESULTADOS EDUCACIONAIS: MENSURAÇÃO E CONSEQUÊNCIAS NO CASO DAS ESCOLAS PÚBLICAS DE MINAS GERAIS.....	168
7.1. INTRODUÇÃO.....	168
7.2. Os resultados da 8ª série/9º ano do Ensino Fundamental.....	173
7.3. Análise dos resíduos: o efeito particular da escola.....	184
7.4. Regressão dos resíduos das médias em relação ao número médio de alunos avaliados.....	189
8. UMA PROPOSTA DE RETIFICAÇÃO DAS METAS DO IDEB.....	199
8.1. Introdução.....	199
8.2. Fórmulas para o cálculo das metas ajustadas do Ideb.....	202
8.3. A atualização do modelo de determinação das metas.....	207
8.3.1. Os resultados da regressão.....	210
8.3.1. A evolução temporal do parâmetro $P$ .....	214
8.4. As diferenças entre os dois métodos.....	218
8.5. Casos ilustrativos de escolas avaliadas diferentemente quanto ao cumprimento das metas.....	222
CONCLUSÕES.....	231
REFERÊNCIAS BIBLIOGRÁFICAS.....	238

## INTRODUÇÃO

O presente estudo tem por finalidade específica e principal discorrer sobre uma nova metodologia de cálculo das metas do Índice de Desenvolvimento da Educação Básica, o Ideb. Como se sabe, este índice, ao menos até presentemente, tem tido o seu uso generalizado nacionalmente, em diversas instâncias e para diversos propósitos. Entre estes últimos, destacam-se o gerenciamento das escolas e redes de ensino, o planejamento escolar, a prestação de contas, a informação para a opinião pública, etc.

Portanto, o grande uso do Ideb torna também necessário que as medidas com ele envolvidas tenham, de fato, uma elevada qualidade técnica. Caso contrário, aumenta-se a chance de ocorrência de vários e graves problemas. Por exemplo, uma decisão equivocada, por parte de uma gestão educacional de maior nível, sobre o sucesso ou o fracasso de uma dada escola em cumprir um determinado objetivo ou meta, pode gerar uma série de consequências indesejáveis, como, por exemplo, injustamente premiar o mérito ou punir os atores responsáveis pelos resultados.

Outro grave problema decorrente das imperfeições das medidas educacionais é a possibilidade de seus resultados confundirem as pesquisas de eficácia escolar que porventura se façam com base nesses dados. Afinal de contas, é natural que os pesquisadores escolares se interessem pelo desempenho das escolas, que muitas vezes é mensurado – de forma politicamente controversa ou não – por meio de testes em larga escala de disciplinas específicas, como Língua Portuguesa e Matemática. Portanto, num cenário como esse, é muito possível que pesquisadores venham a se interessar em estudar escolas, que, aparentemente, encontram-se com um desempenho muito bom, ou, inversamente, muito ruim, em busca de isolar fatores que determinam esse comportamento diferenciado. Entretanto, caso as medidas empregadas na obtenção dos resultados apresentem determinadas imperfeições, pode-se eventualmente constatar que o desempenho dessas escolas vistas como notáveis não seja tão excepcional assim, de modo que os pesquisadores estavam perseguindo e tentando explicar uma quimera...

A preocupação com problemas desse tipo passou, então, a guiar o desenvolvimento da metodologia estatística aqui apresentada. E esta será desenvolvida mais especificamente no final de uma série de capítulos, que estabelece um encadeamento de questões gerais e específicas sobre este tema.

O primeiro capítulo aborda a questão da avaliação do desempenho em organizações públicas, tanto do ponto de vista geral, como também do caso específico da educação. Neste

último aspecto, faz também uma rápida, porém abrangente, consideração metodológica sobre a mensuração da qualidade educacional, assunto central para o desenvolvimento dos sistemas de avaliação em larga escala que se vêm implementando no Brasil e no mundo em décadas mais recentes.

Seguindo por essa linha, o segundo capítulo ilustra muitas dessas discussões tanto políticas quanto metodológicas, que se têm produzido dentro de um cenário pioneiro desse movimento mundial: os Estados Unidos. Nesse sentido, trata-se de um país que há muito tempo vem implementando reformas educacionais de considerável impacto nacional e, em particular e ultimamente, também vem adotando iniciativas de políticas educacionais que colocam uma grande ênfase na testagem em larga escala e no estabelecimento de consequências para os resultados das avaliações.

Muito na esteira do que vem acontecendo no cenário norte-americano, começou a ocorrer algo semelhante também no caso brasileiro, assunto tratado pelo terceiro capítulo. Neste caso, discorre-se sobre as diferentes fases que o movimento de avaliação educacional tem experimentado em nosso país, desde quando ganhou força a partir do final do século 20. Algumas características básicas do movimento de avaliação educacional em larga escala no Brasil são o estabelecimento de sistemas nacionais de aferição da proficiência dos alunos em disciplinas acadêmicas específicas, como português e matemática nos testes do SAEB e da Prova Brasil; a difusão de sistemas estaduais de avaliação, em grande parte com seu respectivo desenvolvimento influenciado pelas avaliações nacionais, e também, com base nestas últimas, a criação e adoção, por parte do Ministério da Educação, do Índice de Desenvolvimento da Educação Básica, o Ideb. (MINISTÉRIO DA EDUCAÇÃO, 2008)

Entre os sistemas estaduais de avaliação educacional mencionados no capítulo 3, um deles interessa-nos particularmente: o Sistema Mineiro de Avaliação da Educação Básica, ou SIMAVE. Isto porque o propósito central deste trabalho, como já se disse, é desenvolver uma metodologia de aperfeiçoamento das metas do Ideb que leve em conta os resultados avaliativos não apenas da Prova Brasil ou SAEB, mas também os dos testes estaduais que costumam ser aplicados em diversas unidades da federação brasileira. Neste sentido, o SIMAVE é um dos mais antigos e consolidados desses sistemas, possuindo uma série histórica de testes de Língua Portuguesa e Matemática iniciada no ano 2000, e que se estende até o presente momento.

Assim sendo, a proposta apresentada neste texto tem como foco as escolas públicas mineiras, avaliadas tanto pelos testes nacionais como pelos estaduais. Em relação a esses últimos, o quarto capítulo propõe uma metodologia capaz de detectar, levando em conta uma

série histórica de resultados, tanto o comportamento médio da escola em termos de status de proficiência, como também de sua evolução ao longo do tempo. E, para fazer isso, são propostos índices de medida dessas características, que, entre outras possibilidades, permitem a identificação de estabelecimentos que, nos últimos anos, vêm se destacando da população de escolas, tanto em termos de maior quanto de menor desempenho.

No quinto capítulo, retoma-se, com um pouco mais de detalhamento, o tópico da medida nacional de qualidade educacional representada pelo Ideb. Abordam-se então algumas características próprias da sua elaboração, bem como do desenvolvimento de suas metas. E, na sequência, são também apresentados e analisados os resultados do Ideb, tanto do ponto de vista nacional quanto do estado de Minas Gerais, que, conforme se disse, possui um interesse especial para o presente trabalho.

A fim de desenvolver uma nova metodologia de determinação das metas do Ideb, foi necessário buscar e aplicar uma solução estatística para os problemas que seriam por nós identificados. Para isso, utilizou-se a método de modelagem linear hierárquica, que permitiu, neste caso, considerar como nível de análise as escolas e suas respectivas trajetórias de evolução temporal da proficiência. Em razão disso, o sexto capítulo discorre sobre os critérios que regem a necessidade do emprego dos modelos lineares hierárquicos, bem como sobre as principais características técnicas destes últimos.

Uma vez definida a técnica estatística a ser empregada neste estudo, apresenta-se, no sétimo capítulo, um exemplo de sua utilização. Trata-se de um modelo linear hierárquico de crescimento longitudinal, aplicado a uma grande parcela das escolas públicas de Minas Gerais, com o objetivo inicial de detectar o grau de volatilidade das suas respectivas medidas de proficiência. Com isso, foi possível mensurar o erro médio – no sentido estatístico – das médias de proficiência em Língua Portuguesa e Matemática de cada escola pública mineira avaliada pela Prova Brasil e pelo SIMAVE, no caso específico do 9º ano (antiga 8ª série) do Ensino Fundamental.

Um ponto de grande relevância, abordado no final do sétimo capítulo, é o impacto que essa volatilidade dos resultados exerce sobre a determinação das metas do Ideb. E essa questão novamente é colocada dentro do contexto das escolas públicas mineiras. Para tanto, realiza-se uma análise de regressão logística que relaciona de que modo as flutuações ocorridas no desempenho das escolas e mensuradas no ano de 2005, quando se calcularam as metas do Ideb, se correlacionam significativamente com os resultados – positivos ou negativos – do cumprimento das metas desse índice por parte das escolas mineiras.

Por fim, no oitavo e último capítulo se propõe uma nova metodologia para o cálculo das metas do Ideb, que leva em conta alguns pontos de grande relevância, os quais, entretanto, não foram utilizados no cálculo das metas oficiais do Ideb. Um desses elementos é o índice socioeconômico das escolas e outro, como se disse, corresponde à instabilidade das medidas educacionais. Dessa forma, é proposto um novo método de cálculo das metas do Ideb que leva em conta essas duas características. Este capítulo então formaliza o método, apresenta e discute os principais resultados das regressões multinível que possibilitaram o seu desenvolvimento e finalmente aborda as semelhanças e diferenças entre os resultados de cumprimentos de metas por parte das escolas, e vistos sob as duas possíveis óticas consideradas neste trabalho: a das metas oficiais e a por nós proposta.

Percebe-se então um grau considerável de semelhança entre os resultados empregados pelas duas metodologias. Sem embargo, também se observa uma quantidade significativa de situações de escolas cujo desempenho vem sendo avaliado de forma diferente por esses dois procedimentos. E, para exemplificar a conveniência da utilização do método aqui proposto, são apresentados e discutidos alguns casos específicos dessas disparidades, ressaltando-se os problemas reais experimentados pela classificação oficial do desempenho de algumas escolas específicas.

Por fim, apresenta-se uma conclusão que perpassa os principais pontos apresentados ao longo do trabalho, e que termina com uma breve discussão sobre a conveniência, ou mesmo a necessidade, de se empregar uma metodologia de aperfeiçoamento das metas do Ideb, preferencialmente nos moldes daquilo que aqui se propõe.

## 1. A AVALIAÇÃO DO DESEMPENHO EM ORGANIZAÇÕES PÚBLICAS

### 1.1. A NECESSIDADE E O FUNCIONAMENTO DOS SISTEMAS DE AVALIAÇÃO DE DESEMPENHO

Conforme já se comentou inicialmente, o presente estudo tem seu principal foco de interesse na investigação e proposição de mecanismos estatísticos voltados para o aperfeiçoamento da mensuração da qualidade educacional, neste caso, representados pelo exemplo prático e atual do Índice de Desenvolvimento da Educação Básica, o Ideb. Porém, antes de se considerar este tema propriamente dito, é preciso contextualizar melhor a situação-problema aqui tratada. Dessa forma, faz-se inicialmente necessário discutir a questão mais geral da avaliação do desempenho nas organizações públicas, bem como o caso específico da avaliação educacional. Em relação à tendência mundial que se vem observando nas últimas décadas e em vários países, de se proporem mecanismos oficiais de avaliação e prestação de contas nos mais diversos setores e aspectos da administração pública, conforme diz Arellano et al. (2012):

Praticamente em todo o mundo, avaliar o desempenho das organizações governamentais, de suas políticas e de seus programas, tem-se convertido em uma moda e uma obsessão para o setor público. Uma das razões disso é que, provavelmente, vem aumentando cada vez mais a distância causada pela diferença entre as grandes necessidades sociais que se precisam enfrentar e a quantidade de recursos disponíveis para fazê-lo. (...) Num contexto assim, os governos precisam demonstrar capacidade para atender e resolver problemas de natureza heterogênea, bem como conhecimentos e informações técnicas mais precisas, caso queiram manter sua capacidade de influência, ou seja, de afetar o comportamento das pessoas e das organizações, com o fim de alcançar seus objetivos sociais e econômicos.

Entretanto, se, por um lado, já tem se tornado quase um senso comum a importância e a necessidade de se avaliar o desempenho das organizações – especificamente neste caso, o das públicas –, por outro lado, existe um grande e variadíssimo debate acerca de como medir o desempenho, e também sobre quem avalia, quem ou o que é avaliado, e quais são ou devem ser as consequências das avaliações, entre uma enorme série de outras questões fundamentais para o efetivo delineamento e implementação de um sistema de avaliação de desempenho, doravante chamado, no presente trabalho, de SAD.

Um ponto que merece ser ressaltado é que a avaliação do desempenho de uma organização pública se dá num contexto eminentemente político, onde, via de regra, existe

uma grande quantidade e variedade de atores dos mais variados níveis hierárquicos e funções, e vinculados a diferentes matizes político-ideológicos e interesses de classe, corporativos ou pessoais, que fazem com que o processo avaliativo, ao menos em tese e desde o princípio, possa vir a ser objeto de negociação e também de conflitos. Nessa mesma linha, cabe ressaltar que a própria determinação dos problemas sociais que se querem prioritariamente resolver, e também daquilo que, em relação a tais políticas, se deseja avaliar, é resultado da adoção de uma determinada postura política que possivelmente pode variar, dependendo de quem são os principais agentes responsáveis pela condução das ações governamentais.

Mas se por um lado, todos esses fatores são capazes de tornar a avaliação das organizações públicas uma tarefa bastante complicada e vulnerável a erros e a críticas, por outro lado, também se deve observar que tudo isso acaba sendo inevitável e até mesmo saudável numa sociedade verdadeiramente democrática, onde o livre e constitucional confronto de ideias e posições político-ideológicas tem a vantagem de fazer com que as ações estatais sejam baseadas na multiplicidade de opiniões e na construção de um consenso mínimo em torno de como se deve promover o bem-estar e a evolução da sociedade.

Diante do que se acaba de dizer, observa-se que os sistemas de avaliação de desempenho preferencialmente se desenvolvem e se implementam como mecanismos formais, que, por meio da adoção de pressupostos e de técnicas de uso e formulação consagrados na literatura e na prática, sejam capazes, não somente de permitir o surgimento de instrumentos mais eficazes de avaliação, como também de, politicamente, possibilitar um maior consenso acerca das práticas avaliativas, o que vem a ser de suma importância para o sucesso, tanto de sua implementação, como também para o estabelecimento de ações de aperfeiçoamento derivadas das conclusões levantadas pelas avaliações (SHAW, 2006).

Nesse sentido, uma abordagem bastante utilizada pelos SADs na atualidade relaciona-se ao planejamento estratégico de uma avaliação. E, nessa mesma linha, logo de princípio impõe-se a necessidade de se considerar, para fins de se avaliar um programa específico executado por uma dada organização, quais são os marcos normativos que orientam a própria origem e propósito desta última. Afinal de contas, avaliar implica, naturalmente, fazer uma comparação entre aquilo que realmente se obtém e aquilo que se considera como bom ou aceitável, seja do ponto de vista de um determinado critério de desempenho – estabelecido pela literatura, por consenso, ou por outros meios –, seja por meio de um critério normativo, pelo qual os indivíduos ou unidades avaliadas são comparados com outros casos na mesma situação, de forma a se estabelecer uma gradação de desempenho mensurada numa dada dimensão de interesse.

E assim, um procedimento comumente utilizado no estabelecimento inicial de um sistema de avaliação de desempenho de um dado programa executado por uma organização, é verificar, primeiro, qual é a missão, e também a visão que esta organização possui acerca de suas respectivas ações. A missão de uma organização, via de regra, se formula como uma declaração de suas principais incumbências e atribuições, onde também se costuma evidenciar, ainda que em linhas gerais, os mecanismos ou prerrogativas pelos quais essa mesma organização conseguirá, em tese, atingir os fins aos quais se propõe. Por outro lado, a visão de uma dada organização corresponde ao cenário desejável que ela espera obter, caso seus respectivos mecanismos consigam adequadamente funcionar segundo aquilo que deles se espera.

Neste ponto, também é necessário considerar que nenhuma organização pública, por mais eficaz ou eficiente que possa vir a ser, consegue exercer um poder total sobre o seu próprio *output*. Ou, em outras palavras, é preciso também observar que, por maiores que sejam os poderes de uma dada organização sobre certos insumos ou *inputs* referentes a um programa ou ação específica, ainda assim é preciso considerar que a realidade social em geral é de uma complexidade tal que o alcance dos resultados de fato obtidos depende também, ao menos em parte, substancialmente de outras realidades ou atores sociais sobre os quais a organização em questão possui pouco ou nenhum controle.

O caso da educação, por exemplo, é uma situação clássica onde tais limitações acontecem. Neste sentido, um marco da literatura sobre avaliação educacional ocorreu nos Estados Unidos no ano de 1966, através do conhecido Relatório Coleman, pelo qual se procurou demonstrar que uma parcela extremamente elevada da variabilidade dos resultados educacionais é estatisticamente “explicada” pelo nível socioeconômico dos alunos que fazem os testes. Ou, em outras palavras, quanto mais elevada for a condição socioeconômica do aluno testado, maiores tendem a ser as suas estimativas de desempenho.

O relatório Coleman, apesar de sua grande importância no estudo das associações entre a proficiência e outros fatores – ou talvez até mesmo devido a essa própria relevância –, após um período inicial, em que tendeu a ser bastante aceito por autoridades educacionais de diversos níveis e naturezas, posteriormente tornou-se objeto de críticas. Algumas destas concentraram-se, por exemplo, no aspecto formal da metodologia empregada por Coleman e seus colegas, alegando, por exemplo, que não se prestou a devida atenção aos processos que se desenrolam dentro das escolas, e particularmente dentro das salas de aula, onde, afinal de contas, o aprendizado de fato ocorre. Assim, segundo esses críticos, a mera associação de uma variável “externa” à escola – como o nível socioeconômico dos alunos, que se encontra

estreita e diretamente vinculada às condições de vida de suas respectivas famílias – pode mascarar a grande riqueza de possibilidades relativas à qualidade educacional, determinada pelas ações e recursos das escolas. (BROOKE, 2008)

Nessa mesma linha, e retomando a questão envolvendo o contexto político de um sistema de avaliação de desempenho, também é preciso atentar-se para a necessidade de se tratar dos diferentes atores relacionados aos programas e processos educacionais a serem avaliados. Afinal de contas, as ações governamentais resultam da interação de um enorme e nem sempre muito bem delimitado conjunto de pessoas e organizações, de modo que, em geral, não é nada simples ou mesmo exequível, isolar determinado efeito de modo a vinculá-lo às ações de um ator específico.

A qualidade do aprendizado educacional, por exemplo, pode depender – e, de fato, depende – de fatores diretamente escolares, como o trabalho dos docentes, a gestão e coordenação dos estabelecimentos educacionais, a infraestrutura física das escolas, etc. –, porém também é influenciada por vários outros agentes e fatores, como os gestores de nível mais elevado nas instâncias educacionais – como os técnicos e funcionários das Secretarias municipais e estaduais de educação, autoridades federais, etc. – além, como se disse anteriormente, de outros fatores extraescolares, relacionados, por exemplo, às comunidades e às famílias dos estudantes (FRANCO, 2014).

Todos esses fatores têm o potencial de, ao menos em parte, afetar consideravelmente o desempenho educacional que se mensura num teste externo, como os que normalmente se empregam nas avaliações em grande escala que têm se tornado tão presentes e importantes atualmente. Porém, além de se mensurar o desempenho, normalmente é também preciso elaborar, teoricamente, mecanismos causais que tentem explicar de que modo as ações dos diversos atores e fatores levam a determinados resultados.

Este assunto é de particular interesse no campo da educação, principalmente em tempos mais recentes, quando a responsabilização pelos resultados educacionais vem se revestindo de uma importância cada vez maior nas políticas públicas de diversos países, entre os quais no próprio Brasil. Dessa forma, um ponto que tem sido alvo de uma acesa polêmica é a responsabilização dos docentes e outros profissionais escolares pelos resultados de seus respectivos alunos, conforme aferidos pelos atuais sistemas de avaliação educacional em grande escala. Até que ponto é politicamente justo, e tecnicamente defensável, estabelecer consequências positivas ou negativas sobre esses profissionais específicos, quando se considera o fato, sobejamente conhecido, de que não são apenas eles os responsáveis pelos resultados observados nas escolas?

Nessa linha, Reimers (2008) ressalta a necessidade de se estabelecerem diálogos iniciais entre as diversas partes envolvidas com os resultados, antes de se proceder à avaliação propriamente dita. E tal procedimento possui, pelo menos, duas grandes vantagens. Uma delas é que, assim, aumentam-se as chances de se atingir um consenso envolvendo os diversos atores da questão. E esse aspecto mais democrático e consensual do processo avaliativo permite torna-lo mais aceitável para os indivíduos e instituições avaliados, o que diminui a possibilidade da ocorrência de boicotes ou oposição ao processo avaliativo. Uma segunda vantagem desse consenso é que ele permite também que a própria instituição responsável pelo programa, e também os atores a ele pertinentes, possam dialogar entre si e, assim, compreender melhor o papel que cada um deles desempenha dentro dessa grande cadeia de causas e efeitos. E assim, o processo avaliativo passa a ser importante não apenas no que diz respeito à avaliação em si, mas começa a se constituir em algo igualmente – ou talvez até mais! – valioso, que é a capacidade que a instituição adquire de pensar, de forma organizada e racional, sobre si mesma, seus objetivos e as maneiras como tenta alcançá-los.

Nesse aspecto, um procedimento particularmente útil ao processo avaliativo é o estabelecimento de uma matriz de marco lógico (MML), que consiga, ainda que de modo tentativo, estabelecer ou descrever os mecanismos gerais que conectam os principais atores e ações organizacionais aos produtos e resultados que se desejam obter. É através do encadeamento desses diversos elos de ação e reação que, então, se cria uma concepção de como um programa pode e deve atuar, e como se dá a contribuição dos diversos atores e mecanismos responsáveis para a ocorrência de um efeito desejável.

Todo esse procedimento pode receber outras denominações, que não a determinação de matrizes de marco lógico. Por exemplo, Carol Weiss (1997) chama-o de teoria do programa, e a esta se refere nos seguintes termos:

Por teoria, não quero dizer qualquer coisa pseudo-intelectual ou polissilábica. Quero dizer o conjunto de crenças subjacentes à ação. A teoria não tem de ser uniformemente aceita. *Não tem de ser correta*. É um conjunto de hipóteses sobre as quais as pessoas constroem seus planos de programas. É uma explicação dos elos causais que unem os inputs do programa aos seus outputs esperados, ou, como afirmou Bickman (1987), “um modelo plausível e sensato do modo como se espera que um programa funcione” (p. 5). Wholey (1987) diz que a teoria de programas identifica “recursos do programa, suas atividades e resultados pretendidos, enquanto especifica uma cadeia de suposições causais que une os recursos, atividades, resultados imediatos e metas finais do programa”.

Uma grande vantagem desse procedimento é que, com ele, torna-se disponível uma espécie de “mapa rodoviário”, que permite ao avaliador deter-se em diversos pontos específicos de interesse da grande cadeia causal que assim se estabelece. Nesse sentido, e a título de ilustração, apresentamos a seguir um exemplo retirado do já mencionado livro *Evaluation*, de Carol Weiss.

Trata-se de uma política pública destinada a promover um aumento do aprendizado dos alunos de um determinado sistema educacional, e que utiliza como um possível mecanismo para isso o aumento salarial dos professores atuantes nessa mesma rede. Embora algumas pessoas talvez possam pensar que o simples fato de se aumentar a remuneração dos docentes seja capaz de resultar, por sua vez, num aumento da aprendizagem dos alunos, ainda assim, cabe dizer que um processo avaliativo que se detenha sobre essa questão, necessita elaborar melhor o problema, de modo a destrinchar, mais exatamente, de que modo ocorre essa associação causal entre o input (o aumento salarial dos professores) e o output (o aumento do nível de aprendizado de seus alunos).

A Figura 1, a seguir, retirada de uma tradução independente do livro de Carol Weiss, ilustra então essa abordagem:

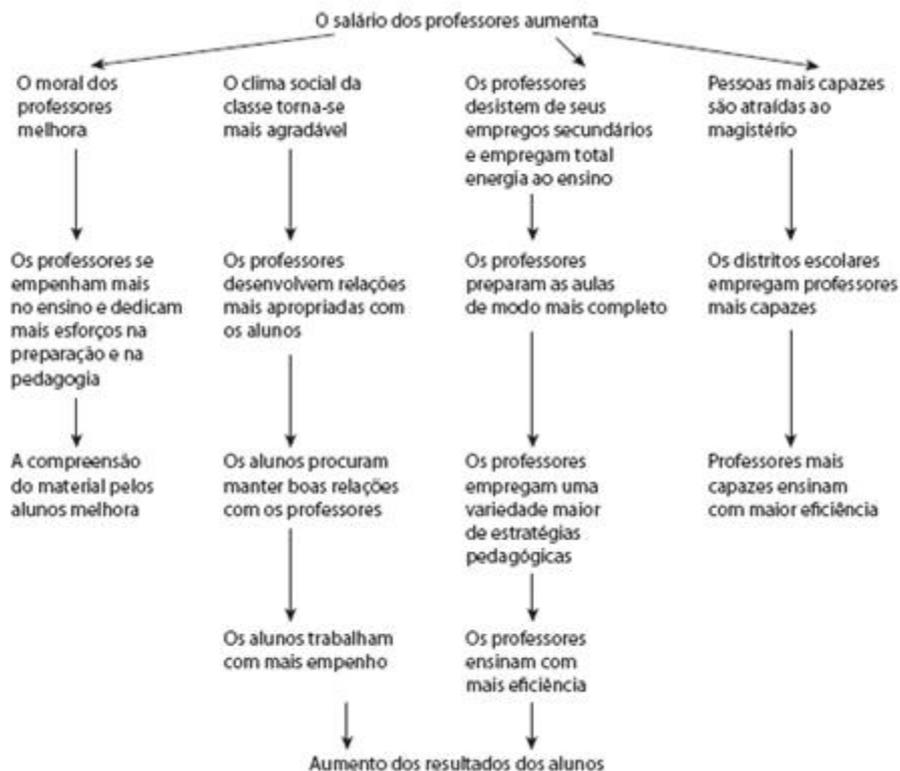


Figura 1: Uma ilustração do desenvolvimento de uma cadeia causal (WEISS, 1997).

Observa-se que, pela cadeia causal assim estabelecida, podem ocorrer vários mecanismos pelos quais o aumento salarial dos professores é capaz de resultar no efeito final desejado, qual seja, o aumento dos resultados acadêmicos dos alunos. Por exemplo, conforme mostra o “galho” descendente na extremidade direita da árvore, a elevação dos níveis de remuneração docente pode, a princípio, servir como um atrativo para que profissionais mais capazes ingressem no magistério, o que resulta num aumento da qualidade do ensino oferecido. Ou então, como mostra a terceira possibilidade da esquerda para a direita, a maior remuneração dos professores também pode fazer com que eles desistam de acumular diversos empregos. E assim, terão mais tempo e energia para se dedicarem a um número menor de aulas presenciais, o que, por sua vez, também resulta num aumento da qualidade educacional.

Essa ilustração é particularmente útil, entre outras coisas, porque ela permite que se atente para um procedimento de vital importância no funcionamento de um SAD: a determinação de indicadores ou índices, estrategicamente localizados em pontos-chave de uma cadeia causal, a fim de mensurar, com a maior precisão e clareza possíveis, o que, de fato, está acontecendo com determinados aspectos do programa.

Por exemplo, na mesma ilustração acima descrita, acerca da hipótese de que uma maior remuneração pode atrair pessoas mais qualificadas para o magistério, um indicador adequado talvez fosse um que mensurasse o nível de qualificação acadêmica dos docentes, antes e depois da implementação dessa nova política. Se, por exemplo, decorrido algum tempo após o estabelecimento do aumento salarial, aumentasse a proporção de professores com formação superior, ou com pós-graduação completa, ou ainda egressos de instituições superiores de maior prestígio acadêmico, então isto seria uma possível evidência de que, ao menos quanto a este aspecto específico, a política parece ter surtido o efeito desejado. Entretanto, ainda caberia discutir outros problemas, como se, de fato, professores com maior qualificação acadêmica são realmente capazes de aumentar significativamente o nível de aprendizado de seus alunos. Embora o senso comum talvez diga que sim, e embora também seja inegável que, para bem ensinar uma determinada disciplina, uma condição necessária seja conhece-la bem, ainda assim, restam dúvidas sobre o fato de que, digamos, professores do ensino fundamental e com pós-graduação *stricto sensu* – muitas vezes voltada para a pesquisa, e não para o ensino – necessariamente se saem melhor do que seus colegas que só possuem, digamos, a graduação superior para ministrar essa mesma disciplina.

Neste ponto, torna-se interessante discorrer com um pouco mais de detalhes sobre esses instrumentos extremamente importantes para a avaliação do desempenho educacional, representados pelos testes de proficiência aplicados em larga escala.

## 1.2. PROPOSTAS E TENTATIVAS DE MENSURAÇÃO DO DESEMPENHO ESCOLAR

Conforme já se comentou rapidamente no início deste texto, ao menos desde o final do século 20 tem havido, tanto no Brasil, como em outros lugares, um interesse crescente pela mensuração estatística da qualidade educacional. E esta tendência tem se traduzido no surgimento de sistemas de avaliação baseados em testes aplicados em grande escala, muitas vezes no nível nacional e estadual. Também se percebe que esses esforços se manifestam em associação com movimentos de reforma educacional que se vêm implementando nesses mesmos contextos, objetivando um aumento da qualidade e da equidade de tais sistemas, ao mesmo tempo em que estimulam a responsabilização das instituições e dos agentes responsáveis por tais melhorias educacionais (MELLO, 2013).

Normalmente, o desempenho nesses testes costuma ser analisado segundo procedimentos baseados em critérios (comparando o desempenho escolar com padrões específicos e usualmente pré-determinados) ou segundo procedimentos baseados na norma (comparando o desempenho das escolas ou alunos entre si) (HOGAN, 2006).

Entretanto, o que muitas vezes ocorre é que os resultados dessas avaliações são tomados como certos, isentos de erros, ou seja, não se questionam – ou, pelo menos, não muito intensamente – a validade e a fidedignidade dessas medidas. Esse tipo de falha é capaz de suscitar problemas consideráveis, por exemplo, de natureza política, quando ele potencialmente pode implicar em consequências injustas para as escolas e os sistemas. Isso, sem contar, obviamente, outro grave problema decorrente do fato de que, agindo desse jeito, tais sistemas podem também enviar sinais equivocados para os tomadores de decisão, algo que tem o potencial de comprometer seriamente o desempenho das ações públicas objetivando à melhoria da educação (KORETZ, 2008).

Portanto, é extremamente relevante que as discussões envolvendo o estabelecimento de sistemas de avaliação educacional se baseiem em procedimentos técnicos rigorosos, capazes de produzir medidas mais acuradas e confiáveis para as tomadas de decisão nos mais diversos níveis e aplicações, decorrentes de tais avaliações. Isso, no caso educacional, leva à necessidade de se inicialmente considerar quais são os possíveis efeitos que as escolas exercem sobre os alunos que as frequentam, e que, exatamente por esse motivo, correspondem ao chamado efeito-escola.

Entre os vários autores que abordam mais especificamente esta questão, Teddlie, Reynolds e Sammons produziram um estudo intitulado *The methodology and scientific*

*properties of school effectiveness research* (literalmente, *Metodologia e propriedades científicas da pesquisa em eficácia escolar*), em que tratam, com um nível razoável de detalhamento, das relevantes e múltiplas questões envolvendo o uso adequado de princípios científicos para obter resultados apropriados de avaliação educacional. Os autores começam seu estudo procurando definir o efeito-escola, que, segundo eles, “*é a capacidade das escolas (normalmente, públicas) de afetar os resultados (normalmente, a proficiência) dos estudantes por elas atendidos*” (TEDDLIE et al., 2000).

Naturalmente, os autores aqui se detêm em um tipo muito específico de resultados educacionais – a chamada “proficiência”, usualmente mensurada através de testes elaborados para cobrir o conteúdo das disciplinas acadêmicas. Um ponto que já aqui merece uma consideração é o fato de que, mesmo em casos de se considerar nos testes estritamente o conteúdo curricular, pode haver sérios problemas em se mensurar o desempenho acadêmico dos alunos em testes de múltipla escolha, como, por exemplo, em Matemática, a área de desenho geométrico.

A partir daí, surge uma crítica que, certamente, merece ser melhor apreciada, de que a expansão da testagem em larga escala supostamente está provocando um estreitamento curricular, no sentido em que determinadas habilidades que também deveriam ser tratadas pelas escolas, pelo fato de não serem testadas, vêm tendo seu respectivo ensino negligenciado. E assim, as matrizes de referência dos testes, mais limitadas em escopo e possuindo propósitos específicos (a testagem, usualmente em grande escala), vêm tomando o lugar das grades curriculares, mais amplas, ambiciosas e voltadas para a educação propriamente dita, e não especificamente para a mensuração de sua respectiva qualidade.

Mas, se dentro de uma mesma disciplina “dura”, como Matemática, já acontecem problemas de falta de cobertura das avaliações externas, naturalmente o problema se torna ainda maior quando se consideram outras habilidades que, ainda que possivelmente correlacionadas com o sucesso acadêmico, possuem uma natureza bastante diferente daquilo que os testes de proficiência normalmente avaliam. Indo por essa linha, existe toda uma vasta gama de trabalhos de teoria educacional apontando para o fato de que, além do domínio de conteúdos acadêmicos mensuráveis por testes em larga escala, os alunos também deveriam ser capazes, ao passar pelas diversas etapas da escolarização, de desenvolver muitas outras habilidades e potencialidades. E, dessa forma, a educação formal seria capaz de ajuda-los a, por exemplo, transformar-se em cidadãos mais conscientes, produtivos e participativos na

sociedade. Nesse sentido, existem vários outros critérios, que não a proficiência em disciplinas acadêmicas específicas, que poderiam ser incluídos num estudo de eficácia escolar. Exemplos disso, apenas para ficarmos em certos casos recentemente mais em voga, são as chamadas “habilidades do século 21”, que incluem a consciência crítica, adaptabilidade, a capacidade de trabalhar em grupo e de resolver conflitos, a criatividade, além do desenvolvimento da ética e da cidadania nos alunos, etc. (HILTON & PELLEGRINO (2012); BINKLEY *et. al.* (2010))

Porém, cabe ter em mente que nenhum sistema de avaliação – da educação ou de qualquer outra área – é capaz de abarcar sozinho tudo aquilo que ele supostamente possa ou deseje mensurar. Naturalmente, para funcionar, tais sistemas precisam estreitar, em menor ou maior grau, o seu foco de interesse, razão pela qual a avaliação também passa a ter um componente fortemente político em seu bojo, conforme já se comentou anteriormente. Afinal de contas, o que é prioritário, em termos de monitoramento e intervenção, pode depender grandemente do ponto de vista e também das funções dos diferentes atores envolvidos com o processo avaliativo.

Uma linha de ação governamental que, concretamente, se vem observando em diversos lugares, inclusive no Brasil, é a consideração prática de que é preciso mensurar algo relacionado ao aprendizado dos alunos nos sistemas educacionais públicos e privados. E, nesse mesmo sentido, também é preciso que essas medidas sejam exequíveis e que igualmente sejam, ao menos em tese, capazes de suscitar mudanças e possíveis acréscimos de desempenho dentro das próprias escolas e sistemas assim avaliados. Portanto, essa necessidade prática torna necessária a implementação de testes de disciplinas específicas, elaborados em formatos capazes de garantir uma correção automática que, por isso mesmo, também é relativamente rápida, barata e bastante fidedigna. Além disso, também é preciso considerar o fato de que, mesmo que não se teste um grande número ou variedade de disciplinas e de áreas curriculares, ainda assim normalmente existe uma elevada correlação entre todas elas, visto que o maior desempenho educacional numa disciplina, geralmente se associa significativa e positivamente ao maior desempenho em outras, independentemente do nível de análise, como aluno ou escola, por exemplo.

Por tudo isso, um dos propósitos do presente texto é trabalhar com a hipótese de aperfeiçoar estatisticamente esses sistemas de medição em larga escala do aprendizado escolar, nos moldes em que eles atualmente existem no Brasil e em vários outros lugares do

mundo, neste início do século 21. Não se deseja, aqui, negar o fato de que a discussão sobre o que deve ser medido pelos sistemas de avaliação educacional é algo extremamente rico e variado. No entanto, adotar-se-á, no presente trabalho, um foco mais específico, qual seja, a discussão de alguns dos principais tópicos da testagem educacional, como a mensuração do desempenho, porém já no contexto das avaliações que efetivamente se realizam atualmente.

Dessa forma, as investigações estatísticas empregadas no presente trabalho basearam-se nos testes de português e matemática de alguns grandes testes aplicados nas escolas públicas de Minas Gerais, sendo dois deles de caráter nacional – o SAEB e a Prova Brasil – e outro estadual, correspondente ao SIMAVE. Cabe dizer que todos esses testes têm passado por uma série de validações, sendo compostos de itens devidamente pré-testados e calibrados segundo a Teoria da Resposta ao Item. E o conjunto inteiro dos itens dessas avaliações tende a cobrir, com seus diversos descritores, a totalidade das matrizes de referência que lhes servem de base, ainda que com graus variados de sucesso, devido a problemas técnicos que limitam a introdução de certos itens nas provas. Entretanto, é preciso considerar que as disciplinas acadêmicas já “consolidadas” nos estudos de avaliação educacional – como é o caso, no Brasil, do português e da matemática – por si só já apresentam uma série de enormes desafios envolvendo a validade e a fidedignidade de sua respectiva mensuração.

Tal fato é, portanto, mais uma das razões pelas quais o presente estudo se propõe a deter-se nas duas disciplinas acadêmicas acima mencionadas. O raciocínio por trás disso é que, ainda que se reconheça a existência de uma série de habilidades relevantes que eventualmente possam ser desenvolvidas entre os alunos pelas escolas por eles frequentadas, uma *proxy* da qualidade educacional será dada por essa dupla de disciplinas acadêmicas – português e matemática – cujas proficiências vêm, pelo menos, sendo medidas com níveis razoáveis de frequência, abrangência e qualidade.

Pois bem, uma vez tendo delimitado consideravelmente nosso escopo do tipo de efeito-escola que pretendemos mensurar, cabe aqui mencionar os diversos estudos que se podem conduzir em relação a tal assunto. Nesse sentido, Teddlie e outros (2000) referem-se a vários tipos de estudos possíveis, entre os quais se podem citar:

1) *A existência do efeito-escola*: Tal tipo de estudo visa a detectar os possíveis efeitos sobre o aprendizado dos alunos decorrentes, por exemplo, da implementação de determinados programas ou de políticas públicas educacionais. Portanto, uma pergunta básica que se faz em

casos assim é: será que a proficiência dos alunos que passaram por esse programa é maior do que a daqueles que dele não participaram, mantidas iguais todas as demais condições?

2) *A magnitude do efeito-escola*: Um determinado componente ou fenômeno do efeito-escola pode existir, porém, ainda assim, ter uma importância relativamente pequena, devido ao fraco impacto que ele porventura venha a exercer sobre o seu respectivo contexto. Isso, naturalmente, torna o estudo da magnitude do efeito-escola um tópico de considerável interesse. Nessa linha, um ponto que deve ser devidamente observado nos estudos estatísticos acerca da qualidade educacional é o fato de que, notadamente nos grandes sistemas de avaliação, onde o número de estudantes investigados pode chegar às centenas de milhares, ou mesmo milhões, a significância das medidas tende a ser elevada, visto que ela cresce com o tamanho das amostras (BOLFARINE, 2005). No entanto, não raramente ocorrem casos em que, embora estatisticamente significantes, as possíveis diferenças entre grupos de examinados podem ser extremamente pequenas, em termos das escalas de desempenho adotadas. E tal fato pode levar a conclusões imprecisas acerca da real dimensão ou importância das diferenças de desempenho observadas entre diferentes grupos.

3) *Os efeitos do contexto*: Estudos que tratam desse tipo de efeito concentram-se em escolas de contextos distintos entre si, como, por exemplo, aquelas pertencentes a redes de ensino diferentes, ou, no caso de se estar considerando estabelecimentos dentro de uma mesma rede, analisar escolas atendendo a clientela com diferentes médias de índice socioeconômico. Um ponto relevante associado a esses estudos é a capacidade de um determinado efeito observado – por exemplo, uma dada melhoria educacional, decorrente de um programa específico –, poder ou não ser generalizada para cenários ou contextos diferentes daquele onde se observou a melhoria em questão. Tais estudos tratam, portanto, da possibilidade de se generalizarem os efeitos observados para diferentes escolas ou outros níveis de agregação.

4) *Consistência*: Em psicometria, esta é uma questão de *validade concorrente*, pela qual diversas medidas ou variáveis diferentes conseguem – ou não – fornecer leituras consistentes de uma determinada dimensão latente que está sendo, ao menos razoavelmente, mensurada por essas variáveis. Estudos dessa natureza procuram avaliar as escolas levando em conta a correlação existente entre o desempenho medido segundo dimensões e instrumentos diferentes. Um estudo típico desses casos, no contexto aqui considerado, é a associação entre Língua Portuguesa e Matemática, tanto do ponto de vista do status quanto do progresso de diferentes unidades de interesse, como as escolas.

5) *A estabilidade dos efeitos escolares*: Outro ponto de grande interesse nos estudos de eficácia das escolas é a capacidade destas manterem um determinado comportamento específico ao longo de um período considerável de tempo. Naturalmente, a estabilidade pode se dar de muitas maneiras, dependendo do objeto de interesse. Há, por exemplo, estudos acerca da estabilidade do status de proficiência, como também a estabilidade das taxas de sua respectiva variação temporal. E, às vezes, uma estabilidade pode apontar para uma realidade mais desejável, como a manutenção de elevados níveis de aprendizado, ou então o contrário, caso os níveis alcançados sejam baixos e não deem indícios de que estejam variando. Muitos dos estudos de estabilidade considerados nesta classificação têm as escolas como unidade de análise, ou seja, preocupam-se com a conservação – ou não – dos resultados dos alunos agregados para o nível do estabelecimento de ensino por eles frequentados.

6) *Os efeitos diferenciais*: Esse tipo de estudo considera os diferentes efeitos que podem existir entre grupos de alunos diferentes dentro de uma mesma escola. Por exemplo, será que a introdução de uma ação específica – digamos, um acréscimo no número de horas que os alunos passarão na escola – aumentará o desempenho tanto de alunos com rendimento baixo como dos de rendimento alto? E tal medida afetará igualmente os alunos de famílias mais ricas e os de famílias mais pobres? Esse tipo de estudo trata, portanto, da capacidade da escola de lidar com alunos provenientes de diferentes contextos e situações, e associa-se a uma importante característica que os sistemas educacionais devem possuir: a equanimidade, ou seja, a possibilidade de bem atender aos alunos, independentemente das possíveis diferenças existentes entre os estudantes, seja decorrentes de variáveis que não estão sob o controle direto da escola, como o índice socioeconômico, seja decorrentes de outras sobre as quais elas, as escolas, podem influenciar mais, como o desempenho ou a atitude dos alunos frente às atividades escolares.

7) *A continuidade dos efeitos escolares*: Esse tipo de estudo relaciona-se aos efeitos pelos quais os alunos que foram avaliados num determinado teste continuam apresentando resultados consistentes em outros exames feitos ao longo do tempo. Diferentemente da quarta categoria acima mencionada, entretanto – a da estabilidade dos efeitos escolares –, os estudos de continuidade têm o aluno, e não a escola, como unidade de análise. Nesse sentido, preocupam-se com o fato de que os sujeitos examinados sejam capazes de reproduzir padrões característicos de desempenho. Por isso, também são chamados de estudos de *validade preditiva*, uma vez que, com base nos resultados anteriores de um examinado, eles permitem

antecipar, dentro de uma dada margem de erro, como será o desempenho do aluno em avaliações posteriores.

### *Os efeitos-escola do Tipo A e do Tipo B*

Além das diversas categorias de estudos sobre o efeito-escola que se acabam de mencionar, e que se diferenciam umas das outras pela área ou tópico específico que abordam, há também a possibilidade de classificar esse efeito quanto ao interesse dos usuários de suas informações. Segundo Raudenbush e Willms (1995), isto é possível devido à existência de dois tipos de efeitos causais que podem ser estimados por um sistema de avaliação educacional, e que por eles são chamados de Tipo A e Tipo B.

Basicamente, os efeitos do Tipo A são aqueles detectados no sentido de se determinar o que as escolas mais agregam ao aprendizado dos alunos. E as “melhores” escolas assim identificadas são precisamente aquelas que, ao final de uma determinada etapa de escolarização, são capazes de produzir os maiores ganhos de proficiência, e/ou os maiores resultados brutos nas avaliações. Esse é um tipo de efeito que interessa diretamente, por exemplo, aos pais dos alunos, visto que estes naturalmente desejam para seus filhos os maiores resultados escolares possíveis.

Por outro lado, os efeitos do Tipo B referem-se às estimativas da capacidade das escolas de produzirem bons resultados levando-se em conta a sua própria realidade, numa espécie de análise de eficiência, ou de relação custo-benefício. Dessa forma, por exemplo, é possível haver uma determinada escola X, inserida em um contexto socioeconomicamente bastante desfavorável, mas que tenha um resultado nas avaliações bastante acima do que a sua realidade socioeconômica leva-la-ia a conseguir. E, ainda que os ganhos ou os resultados brutos dessa escola X sejam, digamos, inferiores aos de uma escola Y situada num contexto mais privilegiado, talvez a escola X ocupe, nos resultados de uma avaliação do Tipo B, uma posição superior à de Y, visto que a primeira conseguiu fazer mais com menos recursos. Em geral, esse efeito não interessa tanto aos pais, que simplesmente desejam que as escolas de seus filhos produzam os melhores resultados, não importando o lado “meritocrático” da questão. Mas os efeitos do Tipo B, entretanto, interessam bastante, ou deveriam interessar, aos gestores educacionais, visto que possibilitam avaliar a eficiência de diferentes instituições integrantes de um dado sistema educacional.

Uma pergunta que um pai, interessado no efeito do Tipo A, poderia fazer é acerca de como o seu filho se sairia, em termos de aprendizado, caso tivesse a chance de experimentar

várias escolas diferentes. E aquelas escolas que lhe possibilitassem aprender mais, naturalmente seriam as mais bem avaliadas. Por outro lado, um gestor interessado no efeito do Tipo B, poderia, ao contrário, estar interessado, não em várias, mas sim em uma escola ou rede específica, e poderia se perguntar como seria o desempenho global dessa escola ou rede, dependendo dela adotar ou não diferentes práticas de ensino.

No estudo que será apresentado posteriormente neste texto, serão feitas abordagens envolvendo, embora não a totalidade, uma grande quantidade dos tópicos de eficácia escolar acima mencionados. Por exemplo, estar-se-á interessado tanto na existência do efeito-escola, conforme mensurado pelas avaliações em larga escala, como também pela sua magnitude e consistência. Outro ponto relevante será a estabilidade temporal desses efeitos, algo que, precisamente, se associa ao emprego de séries temporais consideravelmente dilatadas, englobando o maior número de edições disponíveis, tanto da Prova Brasil quanto do SIMAVE.

Porém, antes de se entrar propriamente nas considerações acerca das medidas educacionais com as quais se trabalhará, também se faz necessário apresentar e discorrer sobre o contexto específico no qual essas medidas surgiram. Naturalmente, a Prova Brasil e o SIMAVE situam-se num cenário relativamente recente no Brasil, correspondente ao final do século 20 ou início do século 21, e chegando até o presente momento. No entanto, a difusão dos sistemas de avaliação no Brasil não se deu, naturalmente, a partir do nada. Muito pelo contrário, ele faz parte de uma tendência geral que se tem observado nesse mesmo sentido e aproximadamente ao mesmo tempo em diversas outras partes do mundo.

E, neste caso, um país em particular exerceu, e vem exercendo, uma considerável influência sobre o movimento de testagem educacional no mundo e no Brasil contemporâneos: os Estados Unidos, que devem essa influência não somente ao seu protagonismo internacional em diversos aspectos – como o político, ideológico, econômico e científico, por exemplo – como também ao fato de que, lá, os sistemas de avaliação educacional possuem uma história consideravelmente mais antiga do que outros países, como do próprio Brasil, que lá buscou inspiração para muitas das medidas que aqui se vêm implementando nessa mesma área.

Em função disso, apresentar-se-á, a seguir, uma síntese da evolução do sistema educacional norte-americano, tendo como fio condutor as grandes reformas pelas quais ele tem passado, particularmente no tocante ao desenvolvimento de sistemas de mensuração da qualidade educacional em grande escala.

## **2. O MOVIMENTO NORTE-AMERICANO DE REFORMA EDUCACIONAL: SINOPSE DE SUA EVOLUÇÃO, DESAFIOS E ASSOCIAÇÃO COM OS SISTEMAS DE AVALIAÇÃO EM LARGA ESCALA**

### **2.1. INTRODUÇÃO**

O propósito central deste capítulo é tecer algumas considerações sobre a maneira como o movimento de reforma educacional nos Estados Unidos vem evoluindo, bem como observar suas relações com o desenvolvimento de seus respectivos sistemas de avaliação educacional em grande escala. Uma análise desse tipo é, a nosso ver, bastante pertinente, dado o fato de que a experiência norte-americana tem influenciado sobremaneira diversos outros países, inclusive o Brasil, na busca pelo aprimoramento de seus respectivos sistemas educacionais. Portanto, discorrer sobre algumas características básicas das reformas e avaliações norte-americanas tem a vantagem de fazer o presente estudo deter-se inicialmente em um contexto onde um conjunto de significativas propostas de avaliação educacional foram implementadas de forma pioneira. E, dessa forma, torna-se possível colher algumas conclusões que podem ser valiosas para os sistemas de nosso país, ainda que, entre os casos brasileiro e o americano, naturalmente haja óbvias variações de contexto e de adaptações face às diferenças estruturais e históricas entre os dois países.

Além disso, conforme ser verá em capítulos posteriores do presente texto, um ponto que nos interessa particularmente neste estudo são os sistemas estaduais de avaliação no Brasil e, mais particularmente no presente caso, o Sistema Mineiro de Avaliação da Educação Básica, o SIMAVE, que começou a ser implementado no ano 2000 no estado de Minas Gerais. A concepção do SIMAVE, por sua vez, foi fortemente influenciada pela do Sistema Avaliação da Educação Básica, o SAEB, o qual, como se sabe, é uma pioneira avaliação nacional brasileira, planejada e implementada na década de 1990 pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, o INEP, do Ministério da Educação, ou MEC.

Entretanto, o SAEB, por sua vez, possui um desenho muito similar ao do *National Assessment of Educational Progress*, o NAEP, ou *Avaliação Nacional do Progresso Educacional*, que vem sendo aplicada nos Estados Unidos desde 1964. Tanto no caso nacional americano quanto no brasileiro, inspirado no primeiro, as avaliações têm sido amostrais, bianuais, enfatizam as disciplinas de leitura e matemática e também constituem-se de testes transversais aplicados às séries finais de cada ciclo de escolarização, que, no

contexto brasileiro, no “antigo” sistema de seriação então vigente, correspondiam às 4<sup>a</sup> e 8<sup>a</sup> séries do Ensino Fundamental e à 3<sup>a</sup> série do Ensino Médio.

Façamos, portanto, uma recapitulação do que vem acontecendo na América do Norte no tocante à evolução de seu sistema de educação básica, atrelada ao estabelecimento e desenvolvimento de seus sistemas de avaliação. Nesse sentido, mais particularmente, o objetivo específico do presente capítulo é descrever em linhas gerais, algumas características marcantes das reformas que se vêm discutindo e implementando nos Estados Unidos, com vistas a promover um aumento da qualidade e da equidade educacional naquele país.

Analisar-se-ão aqui, com um pouco mais de detalhe, dois mecanismos específicos de discussão e promoção dessas reformas. Um deles é a adoção de sistemas de prestação de contas baseados em resultados de testes em larga escala de habilidades acadêmicas específicas dos alunos. Outro refere-se à considerável atenção que vêm recebendo as chamadas habilidades do século 21. O texto então observa como o debate e a prática das reformas e dos sistemas de avaliação vêm evoluindo ao longo das últimas décadas nos EUA. E igualmente disserta sobre o modo como eles vêm influenciando o estudo das habilidades do século 21, assim como também vêm por este sendo influenciados. Algumas constatações do texto explicitam uma série de avanços e problemas das recentes reformas e sistemas norte-americanos de avaliação educacional. E outras relacionam essas características à maneira como o debate acerca das novas competências vem se incorporando a essa discussão, vista como de interesse estratégico para os Estados Unidos.

## 2.2. BREVE RETROSPECTIVA HISTÓRICA DO CONTEXTO EDUCACIONAL NORTE-AMERICANO

O desenvolvimento da educação pública norte-americana apresenta uma estreita relação com as características de sua colonização. Inicialmente levada a cabo por pioneiros britânicos estabelecidos na costa atlântica do nordeste do país ainda em formação, a ocupação da nova colônia se fez junto à tentativa de se fundar, em terras americanas, uma “Nova Inglaterra”, nome que passou a se referir àquela própria região. Isto também queria dizer que, diferentemente de outros casos mais comuns de exploração das terras americanas por potências europeias, nos Estados Unidos, houve, pelo menos desde o século 17, uma tentativa organizada dos colonos se estabelecerem com suas famílias, de modo a fazer do novo país

uma espécie de continuidade da pátria europeia original, e não apenas uma área de exploração temporária e extrativista.

Em termos educacionais, uma consequência disso foi o fato de que, por exemplo, no estado de Massachusetts, um dos mais proeminentes da Nova Inglaterra, já no ano de 1647 – portanto, muito mais de um século antes da independência norte-americana – uma legislação local já previa que toda cidade abrigando, pelo menos, 50 famílias, deveria ter uma escola para atender às crianças de sua comunidade. (COMER, 2004)

Houve também dois outros aspectos de relevo em que os norte-americanos se destacaram na educação pública. Um deles foi a questão da igualdade de oportunidades entre os sexos, também tratada com bastante antecedência, em termos comparativos internacionais. Através de uma lei de 1789 – aprovada, portanto, 13 anos após a independência dos Estados Unidos, e coincidentemente no mesmo ano da Revolução Francesa –, determinou-se que as escolas públicas devessem atender tanto a mulheres quanto homens. Um segundo aspecto de relevo foi a obrigatoriedade do ensino fundamental, cujas leis começaram a ser aprovadas em meados do século 19, como, por exemplo, no já mencionado caso de Massachusetts, em 1854.

Uma consequência de tudo isso foi o fato de que, já no século 19, os Estados Unidos, naturalmente amparados por circunstâncias favoráveis tanto econômica quanto politicamente, alcançaram um progresso invejável em sua educação básica quando comparados com outros países, ou ao menos no que dizia respeito à sua população branca. Entre outros indicadores, em 1865, o país orgulhava-se de ter alfabetizado 94% de sua população branca nos estados do norte e 86% nos do sul do país. (BRINKLEY, 2009)

Entretanto, esses últimos dados revelam também um ponto bastante contraditório e controverso da educação pública norte-americana: seu elevado grau de desigualdade racial. Com efeito, quando se observa o progresso das leis e práticas antirracismo na educação pública nos EUA, percebe-se que a adoção de práticas mais igualitárias na educação demorou ainda mais para se concretizar plenamente do que, por exemplo, em muitos países menos desenvolvidos educacionalmente, como os da América Latina. No caso americano, por exemplo, ainda em 1954, nada menos do que 17 dos 50 estados norte-americanos, ou um terço deles, ainda requeria por lei a segregação racial no ensino fundamental, além de outros quatro estados que a permitiam, e mais outros 11 para os quais não havia nenhuma legislação específica.

Não obstante, essa situação passou a alterar-se radicalmente com a eclosão do movimento de direitos civis na década de 1960. Uma consequência disso foi o

desaparecimento das últimas escolas segregadas nos Estados Unidos no final da década de 1970. (CHUNG & REIMERS, 2014)

Dessa forma, os EUA consolidaram, assim, uma posição já privilegiada do ponto de vista do alcance de habilidades educacionais básicas por parte de sua população em geral. E o vigor e a antecipação com que esses problemas foram tratados ao longo de sua história, anteriormente à sua própria independência, foram fatores preponderantes para o protagonismo mundial que esse país exerceria no cenário internacional a partir do século 19.

### 2.3. CRÍTICAS E REFORMAS DO SISTEMA EDUCACIONAL AMERICANO

Entretanto, apesar de todo esse progresso quanto ao desenvolvimento de oportunidades educacionais para a sua população, o sistema educacional dos Estados Unidos também experimentou, ao longo de sua história, vigorosas sugestões de aperfeiçoamento, embasadas em fortes críticas que se faziam à sua qualidade e equidade. Inicialmente, esse movimento se amparou em colaborações de alguns críticos que, já desde o século 19, preconizavam uma educação baseada numa formação humanística mais ampla, sem perder, entretanto, a aquisição dos chamados “três Rs”, que, em inglês, correspondiam às habilidades tidas como básicas: *Reading* (leitura), *wRiting* (escrita) e *aRithmetic* (aritmética). Dessa forma, por exemplo, preconizava-se também o ensino de ortografia, geografia, história, constituição americana, ciências naturais, educação física, arte e música, que também deveriam ser ministradas nas escolas. (BOHAN & NULL, 2007)

Não obstante todas essas áreas e disciplinas – ou talvez mesmo por causa delas... –, as críticas à eficiência do sistema educacional americano continuaram. Por exemplo, em 1906, um relatório do Estado de Massachusetts recomendou que, nas escolas públicas, se introduzissem elementos relacionados à educação técnica e industrial. A alegação para isso era o fato de que a educação de então, vista como “antiquada”, estimulava a evasão escolar, ao mesmo tempo em que não conseguia fornecer, aos alunos que continuavam no sistema, uma formação capaz de prepará-los para os grandes avanços industriais e técnicos que se viviam naquele início do século 20. (CHUNG & REIMERS, 2014)

Pouco após isso, outros movimentos de reforma educacional nos Estados Unidos se intensificaram, sendo que um de seus principais proponentes foi John Dewey. Com este, veio uma ênfase na ideia de que a democracia norte-americana dependia enormemente da capacidade do país de fornecer a toda a sua população uma educação de maior qualidade e equidade. Tudo isso resultou, por exemplo, no surgimento, em 1919, da *Progressive*

*Education Association*, ou Associação para a Educação Progressiva, que passou a clamar por reformas educacionais orientadas por essas novas críticas e análises.

Nessa mesma linha, alguns teóricos norte-americanos da educação começaram, no início do século 20, a defender a capacidade das escolas de estimular a aprendizagem experimental, o raciocínio crítico, a resolução de problemas, o trabalho colaborativo e em equipe, os projetos de aprendizagem, a educação com responsabilidade social, etc. Portanto, percebe-se que, curiosamente, durante esse período da virada do século 19 para o 20, começaram a ser propagadas ideias educacionais que, de modo um tanto quanto irônico, também são hoje em dia defendidas sob o nome de “habilidades do século 21”, mas que, na verdade, também eram importantes já no século 20, no 19 e mesmo nos anteriores...

Ao menos em termos de resultados práticos e aparentes – e, apesar do até então não resolvido problema da desigualdade racial de oportunidades –, o sistema educacional norte-americano dava mostras de um considerável vigor, em termos comparativos internacionais. E este foi certamente um fator de peso para a vitória do país nas duas Guerras Mundiais, bem como para o protagonismo por ele desempenhado na reorganização da ordem mundial, após a derrota definitiva do Eixo fascista em 1945.

Entretanto, o pós-guerra caracterizou-se também por uma ferrenha competição com a União Soviética e seus aliados comunistas, num conflito geopolítico e ideológico de proporções mundiais, conhecido como a Guerra Fria, e que dominaria grande parte do mundo durante quase que inteiramente a segunda metade do século 20. No ápice dessa disputa pela supremacia mundial, a União Soviética, no ano de 1957, tomou uma dramática dianteira ao colocar, pela primeira vez na história da humanidade, um satélite em órbita em torno da Terra. Isto também significava que, a partir daí, os russos eram capazes de, por exemplo, colocar um engenho espacial voando a 100 quilômetros de altura sobre Washington, DC, carregando qualquer tipo de carga – de câmeras fotográficas ou de TV a ogivas nucleares – sem que os americanos nada pudessem fazer para impedi-los...

O lançamento do Sputnik provocou um pânico, tanto na opinião pública quanto nos mais altos escalões da sociedade e do governo norte-americano, que levou também à realização de uma profunda autocrítica nacional. E, de forma natural, a educação pública do país passou a ser mais uma vez revista, em meio a um pensamento praticamente unânime de que a perigosíssima defasagem americana em relação aos soviéticos se devia, em boa ou em sua maior parte, aos problemas da educação pública do país (BROOKE, 2012).

Tudo isso, somado ao já mencionado Movimento de Direitos Civis, e também à chamada *Guerra à Pobreza*, uma ampla política federal e social, levada a cabo

aproximadamente naquela mesma ocasião, fez com que os Estados Unidos promovessem, a partir do final da década de 1950, um vigoroso movimento de reforma educacional, que se estenderia ao longo de alguns decênios seguintes.

Dessa época, por exemplo, datam as múltiplas tentativas de se introduzirem reformas curriculares que fossem capazes de pôr os alunos do ensino fundamental e médio em contato com as mais recentes descobertas e conceitos científicos. Tal movimento também se traduziu numa tentativa de se introduzirem reformas metodológicas nas escolas, que enfatizassem a experiência e a capacidade analítica dos discentes, preparando-os melhor para enfrentarem situações mais realistas de tomadas de decisão baseadas na análise de informações.

Entretanto, não obstante todas essas inovações, os resultados práticos dessas novas medidas, de um modo geral, foram bastante modestos, para dizer o mínimo. Possivelmente para grande decepção de muitos educadores nacionalistas americanos, os programas de avaliação então implementados para aferir a eficácia ou eficiência dessas novas reformas, não conseguiram detectar, na realidade, os grandes saltos de aprendizagem que inicialmente se esperavam alcançar.

Hoje em dia, parece ser um consenso que uma das causas desse fracasso foi a estratégia de certo modo autocrática que se adotou para se desenvolverem tais reformas. Durante a concepção e mesmo a implementação dessas últimas, deu-se muita atenção ao que tinham a dizer os chamados grandes especialistas ou autoridades científicas, o que levou até mesmo à inclusão de pesquisadores vencedores de Prêmios Nobel na confecção dos materiais didáticos utilizados nas escolas de nível fundamental e médio. Entretanto, faltaram em tudo isso o convencimento, engajamento e participação de certos elementos cruciais para o sucesso de qualquer reforma educacional: os professores que deveriam implementá-la no dia a dia da sala de aula, e também os gestores de nível local que deveriam zelar para que tais mudanças ocorressem. (BROOKE, 2012)

Mas, se por um lado, os Estados Unidos não conseguiram avançar tanto quanto queriam inicialmente, por outro lado, em termos de Guerra Fria, as coisas até que não foram tão ruins assim para eles. O país ainda era dono de uma formidável liderança mundial científica, tecnológica e econômica, e isso, somado ao fato de que o bloco socialista também tinha seus graves problemas – que resultariam no colapso do comunismo ainda no final do século 20 –, fez com que não se realizassem as previsões catastrofistas de alguns americanos, que temiam uma derrocada de sua sociedade frente à “ameaça” soviética.

Mas, apesar desse alívio parcial, e aproximadamente ao mesmo tempo, os Estados Unidos começaram a sofrer uma nova ameaça, não tão belicista, porém não menos

preocupante, à sua supremacia econômica, política e social no mundo: a concorrência de grandes economias emergentes – como a do Japão e, mais recentemente, da China e de outros países orientais, liderados pelos chamados “Tigres Asiáticos”, que passaram a disputar, muitas vezes, com vantagem, os mercados até então dominados pelos norte-americanos.

Outro ponto preocupante para os Estados Unidos é que, ao longo desse período mais recente de sua história, a ameaça de seus competidores não se tem restringido ao campo econômico, mas também tem se manifestado no preparo educacional das populações envolvidas. A partir das últimas décadas do século 20, começaram a surgir testes internacionais, que puderam servir de parâmetro de comparação entre diversos países, no tocante à capacidade de suas respectivas populações estudantis desenvolverem habilidades mínimas ou básicas em diversas áreas do conhecimento. Apesar das restrições que se fazem à capacidade de comparabilidade desses testes, seus resultados têm sido bastante divulgados pela grande imprensa, e também certamente têm servido de tópicos de discussão nas agendas políticas dos países envolvidos.

Para os Estados Unidos, especificamente, embora não sejam catastróficos, os resultados das avaliações não têm se mostrado nada animadores. O desempenho do país tem sistematicamente ficado bastante abaixo de diversas outras nações “menores”, que vão desde países do norte europeu, como a Finlândia, até países emergentes, como Singapura, no Extremo Oriente.

#### 2.4. REFORMAS MAIS RECENTES NOS EUA

Esse estado de coisas suscitou o surgimento de uma nova onda de alarmismo norte-americano, cujo marco fundamental foi um relatório do governo daquele país, publicado em 1983. O próprio título desse relatório já dava o tom dramático de suas constatações: “*A Nation at Risk*”, ou “Uma Nação em Perigo”. Esse informe foi produzido pela *National Commission in Excellence in Education*, ou Comissão Nacional de Avaliação em Educação, a pedido do então Presidente Ronald Reagan.

Algumas de suas constatações não tinham muita novidade. Elas apenas reforçavam a ideia, já aventada desde tempos anteriores, de que a educação era um aspecto de grande importância para o país, não apenas em termos de sua economia, mas também de sua própria segurança. E que, apesar dessa relevância, o assunto continuava sendo insuficientemente atacado nos Estados Unidos, conforme se percebia na acentuada defasagem dos resultados norte-americanos nas avaliações internacionais. Tanto que, em uma de suas constatações, o

relatório atesta que “o alicerce educacional de nossa sociedade presentemente está sendo erodido por uma onda crescente de mediocridade”. (GARDNER, 1983)

Essa preocupação nacional com o suposto descalabro da educação básica americana contribuiu para que, no ano de 2002, ocorresse uma insólita comunhão de esforços que abarcava de democratas a republicanos, ao longo de considerável parte do espectro político norte-americano. Isso resultou na aprovação, em 5 de janeiro de 2002, da Lei Federal 107-110, mais conhecida como *No Child Left Behind Act* (NCLB), ou *Lei Nenhuma Criança Deixada para Trás*, com o apoio de ambos os partidos do Congresso Federal americano. E, aumentando ainda mais a força e o simbolismo dessa comunhão de esforços, a lei foi apresentada ao Senado americano por Ted Kennedy – um histórico líder democrata, irmão do falecido Presidente John Fitzgerald Kennedy –, mas foi aprovada no governo de George Bush Filho, um expoente do conservadorismo republicano nos Estados Unidos.

Essa lei veio a exercer, instantaneamente, um enorme impacto na educação básica americana. E conseguiu isso por meio, primeiro da adoção de um gigantesco pacote de apoio financeiro aos estados, distritos e escolas do país. Para o ano fiscal de 2002, o primeiro a receber os fundos vinculados a essa lei, haveria, por parte do Governo Federal, a disponibilização de US\$13,5 bilhões. Posteriormente, esses valores sofreriam, por sua vez, aumentos anuais e progressivos de cerca de US\$2 bilhões em relação ao ano fiscal anterior, fazendo com que, em 2007, os recursos previstos chegassem a US\$25 bilhões. Este último valor, somado às quantias despendidas nos anos anteriores, desde o início dessa política, fez o governo federal norte-americano despender nada menos do que US\$116,25 bilhões nesse programa, ao longo de seus seis primeiros anos (US SECRETARY OF EDUCATION, 2002)

Com essa verba, esperava-se e instava-se que os estados, distritos e escolas americanas tivessem condições de promover melhorias radicais na aprendizagem dos seus respectivos alunos. E isso poderia ser feito através da contratação de melhores professores e outros profissionais, e também por meio de estudos e práticas de reforma curricular, em investimentos da infraestrutura e de outros importantes recursos escolares, entre outras medidas.

Só que, para se beneficiarem dessas verbas, era preciso que as unidades educacionais dessem mostras de que já estivessem no caminho do aperfeiçoamento. E a comprovação disso seria medida pelo chamado progresso anual adequado – *adequate yearly progress*, ou AYP, no original (DORAN, 2004). Neste ponto, entrava em cena o segundo grande componente dessa lei: a instituição de um sistema de prestação de contas de nível nacional, que trazia

também em seu bojo um forte movimento de testagem do aprendizado e de responsabilização pelos resultados educacionais assim obtidos.

Mas, além das possíveis compensações pelos bons resultados, havia também a instituição de sanções atreladas a um desempenho insuficiente nos testes. E, caso esse déficit de desempenho se acentuasse ou não fosse significativamente diminuído ao longo das diversas edições da testagem, as escolas e os distritos envolvidos poderiam sofrer uma ingerência externa. Dessa forma, poderia haver, por exemplo, demissões de professores, a transformação de um estabelecimento de ensino numa *charter school*, que funciona com recursos públicos porém é administrada de modo privado, ou até mesmo, em casos extremos, o fechamento de um estabelecimento particularmente mal sucedido.

Boa parte da base para essas decisões extremas seriam os testes de desempenho, os quais, por sua vez, deveriam avaliar, anualmente e ao menos uma vez, os alunos matriculados nas séries de 3 a 5, de 6 a 9 e de 10 a 12 do ensino básico. Além disso, um ponto extremamente relevante dessas avaliações é que elas deveriam incluir todos os alunos dos departamentos de educação e das escolas, daí resultando o próprio nome pelo qual essa lei ficou popularmente conhecida, de “nenhuma criança deixada para trás” (US SECRETARY OF EDUCATION, 2002).

A princípio, os argumentos por trás dessas determinações pareciam bastante legítimos aos olhos do grande público e mesmo das autoridades. Afinal de contas, quando não havia essa obrigatoriedade de testar todos os alunos, era perfeitamente possível, ou talvez até comum, haver fraudes mais ou menos disfarçadas nas testagens, pelas quais os alunos tradicionalmente com os piores resultados acadêmicos eram solicitados a não comparecer à escola no dia do exame... Isso, obviamente, provocava uma inflação dos resultados obtidos, o que não interessava a quase ninguém.

Portanto, agora, a nova lei exigia que se testassem todos os alunos, e tal medida seria capaz – assim se esperava – de promover nas escolas e nos profissionais nelas atuantes o despertar ou um aumento do senso de responsabilização pelos resultados de todo o conjunto de seus alunos, sem exceção (HANUSHEK, 2014).

Além disso, havia outro ponto bastante rico em todo esse movimento, vinculado à notória desigualdade de oportunidades educacionais que então existia e ainda existe, tanto nos Estados Unidos quanto em outros países. No caso americano, por exemplo, havia lacunas de aprendizagem significativas, por exemplo, entre estudantes pertencentes à maioria – brancos ou também, étnica e convencionalmente, do Extremo Oriente – ou à minoria, como os negros, os latinos e os índios. Também havia diferenças significativas de desempenho que

acompanhavam, *grosso modo*, o chamado gradiente socioeconômico, ou seja, refletiam uma dependência ou associação positiva entre os resultados educacionais e o nível socioeconômico dos alunos. Tudo isso sem contar outros tipos de lacunas, como o universal problema dos estudantes portadores de deficiência física ou mental, e também, no caso específico norte-americano, a grande quantidade de jovens e crianças filhos de imigrantes estrangeiros e que não tinham o inglês como primeira língua.

Tudo isso fez com que, no NCLB, houvesse a determinação de que o progresso anual adequado devesse ocorrer, por exemplo, não numa escola do ponto de vista médio ou geral, mas sim em cada um dos segmentos nos quais ela se dividia quanto a seu respectivo corpo discente. Ou seja, deveria haver, por exemplo, melhoras anuais substantivas entre os alunos das maiorias e igualmente das minorias, e também entre os mais ricos e os mais pobres, separadamente. Portanto, para que fossem de fato bem classificadas nas avaliações, as escolas deveriam conseguir bons resultados em todas as diversas faixas de alunos nelas matriculados.

Outro ponto de interesse neste tema é o modo específico de mensurar o progresso anual adequado. Neste caso, o que se fez foi reconhecer a notória e relativa independência legal e administrativa que os estados norte-americanos têm dentro de sua federação. No caso educacional, por exemplo, os EUA são um país altamente descentralizado, onde cada estado e, dentro de cada um deles, cada distrito escolar dispõe de uma ampla autonomia para regulamentar e deliberar sobre amplas questões educacionais, que vão desde o estabelecimento de um currículo até as políticas de contratação dos professores e do financiamento das escolas.

Ora, num contexto destes, ficaria extremamente desafiador criar, digamos, um teste único, baseado num currículo único, para medir o avanço da proficiência dos estudantes em todos os cantos do país. E assim, optou-se pela descentralização também neste caso. Ou seja, cada estado implementaria seus próprios critérios de desempenho, que apenas teriam que obedecer aos critérios federais básicos e, ao mesmo tempo, ser suficientemente razoáveis e coerentes. Uma das exigências da lei federal era que cada estado definisse e justificasse o que, dentro de sua legislação educacional, deveria ser considerado como um progresso anual adequado, levando para isso em conta o objetivo maior de diminuir satisfatoriamente as lacunas de resultados educacionais então observadas.

Entretanto, neste ponto já aparece um problema bastante considerável, que alimentou e segue alimentando um conjunto de pesadas críticas a esse sistema de avaliação. Ora, se o desempenho dos alunos tem que sofrer uma evolução, ao saltar de níveis baixos para níveis mais altos de proficiência, e se esses níveis, por sua vez, podem ser definidos pelos diferentes

estados com um grau, no mínimo razoável, de arbitrariedade, então, por conveniência, é perfeitamente possível adotar padrões artificiais de desempenho satisfatório, de modo que neles possam ser incluídos mesmo aqueles alunos que, por outras medidas mais válidas de desempenho, estariam com um considerável déficit de aprendizagem. Em outras palavras, os critérios para se classificar um desempenho como “adequado”, por exemplo, tornaram-se consideravelmente baixos, mascarando, assim, os resultados educacionais. E as consequências disso foram, ou têm sido, extremamente ruins. Com tal estratégia, não só se premia a ausência ou insuficiência de esforços e resultados, como também se transmite uma mensagem enganosa, para os alunos, suas famílias, os próprios sistemas educacionais e a sociedade como um todo, de que as coisas estão indo bem, quando na verdade este talvez não seja o caso...

Uma constatação real deste problema se verifica quando, por exemplo, se comparam os resultados dos estados americanos conforme medidos por suas próprias avaliações, moldadas pelo NCLB, com os resultados desses mesmos lugares obtidos pelo NAEP.

Esta última, conforme há pouco se comentou, é uma avaliação nacional e amostral das séries finais de três diferentes ciclos da educação básica norte-americana, muito semelhante ao modelo original do Sistema Nacional de Avaliação Básica – o SAEB brasileiro. Pois bem, os dados do NAEP têm demonstrado que, na grande maioria dos estados norte-americanos, os progressos educacionais ao longo dos anos têm sido muito mais modestos do que aqueles apontados pelos novos testes derivados do NCLB, havendo até casos de uma estagnação ou mesmo decréscimo de desempenho segundo o NAEP, acompanhado da constatação de um significativo progresso, no mesmo momento e lugar, segundo o teste estadual baseado nos novos critérios!

Outro ponto que vem sendo alvo de crítica ao NCLB é a irrealidade da meta proposta, das escolas conseguirem tirar todos os seus alunos, sem exceção, dos níveis mais baixos de proficiência, no espaço de poucos anos contados a partir da aprovação da lei. Embora, em termos de discurso político, isso possa soar muito agradável, parece óbvio que, não somente nos Estados Unidos, mas em qualquer outro lugar do mundo, é impossível que um estado garanta 100% de bons resultados educacionais para sua população como um todo, pelo simples motivo de que tais resultados dependem também de outros fatores, que não estão sob o controle direto das instituições educacionais. Exemplos destes, conforme já se comentou, são o nível socioeconômico dos alunos, a existência de um ambiente culturalmente propício em casa para a sua aprendizagem, a saúde física, mental e emocional dos discentes, etc.

Entretanto, ao menos no momento de aprovação do NCLB, essa quimera de zerar o déficit de desempenho da totalidade dos alunos americanos a curto prazo pareceu algo

extremamente razoável para a maioria dos congressistas e da opinião pública americana... Com o passar do tempo, entretanto, ficou claro que, mesmo com a mencionada inflação dos resultados ocorrida em boa parte dos estados, ainda assim, tais metas estavam longe de ser cumpridas. Isso desencadeou uma série de “ajustes”, pelos quais os estados reconheciam que não teriam como alcançar determinadas metas, e, em decorrência disso, revisavam essas previsões para patamares mais realistas, submetendo tais reformulações ao juízo de instâncias federais. Porém medidas desse tipo bem pouco contribuíram para aumentar a eficiência e a credibilidade desse sistema, que passou a ser atacado por um número crescente de críticos saídos de vários setores, como o político e o científico.

No meio dessas críticas, também se objetou que a colocação de consequências tão elevadas à testagem – tendência conhecida no original inglês como *high stakes testing* – estimulava as escolas e os professores a privilegiar principalmente as disciplinas e conteúdos mais avaliados segundo os critérios do NCLB. Outro ponto preocupante era o fato de que as novas medidas desestimularam as escolas e os distritos a aceitar alunos com um histórico escolar ruim, por qualquer que fosse o motivo. Afinal de contas, todos os alunos teriam que ser testados, de forma que não era nada interessante para essas instituições educacionais receber discentes cuja aprendizagem fosse mais lenta, incerta e demandasse mais esforços.

E, mesmo entre os alunos já admitidos pelas escolas e distritos, também passou a haver uma concentração de interesse por aqueles de proficiência aproximadamente no meio da escala de desempenho específica para a série ou ano escolar em questão. Isto porque os resultados ficavam melhores quando se conseguia mover esses alunos de um nível inferior de proficiência para o que lhe ficava um pouco acima. Por outro lado, reduziu-se o interesse pelos alunos situados em níveis bastante baixos, ou, inversamente, muito altos de desempenho. No primeiro caso, ainda que se conseguisse um progresso, este provavelmente não seria elevado o suficiente para elevar tais alunos a um patamar aceitável de desempenho. E, no outro extremo, os alunos já tinham um alto nível de proficiência, de modo que não valeria tanto a pena aumentar esse nível ainda mais, visto que tal elevação muito pouco ou mesmo nada acrescentaria ao desempenho educacional da instituição, segundo os critérios de mensuração então adotados.

Críticas adicionais de cunho fortemente social também se faziam. Por exemplo, a de que os professores e gestores eram profissionais e, portanto, deveriam ter sido mais ouvidos nas deliberações dessas leis, o que não parece ter acontecido. E assim repetiu-se, já no início do século 21, o que tinha ocorrido cerca de meio século antes, de uma reforma educacional sofrer graves reveses em decorrência de sua incapacidade de atuar de baixo para cima,

falhando, dessa forma, em arrebanhar os esforços dos profissionais mais relacionados com o dia a dia da prática escolar...

Também na linha das críticas de ordem política e social, destacavam-se aquelas dizendo acerca da suposta injustiça de se responsabilizar, por exemplo, quase que inteiramente os professores e gestores escolares pelo desempenho de seus respectivos alunos. Afinal de contas, o desempenho discente associa-se, como já se mencionou, a uma série de outros fatores sobre os quais a escola tem pouco ou nenhum controle. Portanto, esse novo movimento de testagem e responsabilização gerou um considerável grau de estresse, pressão e consequências injustas para os profissionais educacionais por ele avaliados.

Outra crítica era que, do ponto de vista pragmático e social, em muitos casos era melhor, ou menos pior, ter uma escola ruim numa determinada comunidade do que não ter escola nenhuma. Ainda que seus resultados não fossem porventura suficientemente bons, havia que se reconhecer que não poucas escolas – talvez a sua maioria – eram instituições naturalmente nascidas do seio de suas comunidades. Para virem a existir, contaram com o esforço e a participação de líderes, cidadãos e instituições locais, e bem ou mal, encerravam em si próprias uma parte da história e da identidade cultural, social e mesmo política da comunidade na qual se situavam e à qual serviam. Portanto, muitos viam como um ato violento e insensível de arbitrariedade, fechar um estabelecimento desse tipo sob a alegação de que ele não estava alcançando metas cujas próprias mensurações e determinações, conforme se disse, em si só já eram bastante questionáveis.

Uma expoente em todas essas críticas foi a educadora Diane Ravitch, que vem desempenhando um papel especial em todo esse revisionismo, devido à sua própria postura inicial em relação a tais medidas. De orientação conservadora, D. Ravitch havia se impressionado sobremaneira – como, de resto, praticamente todo o país – com as conclusões do relatório *A Nation at Risk*. Só que, além disso, ela teve uma bem-sucedida carreira na política educacional americana, o que a levou a ocupar o cargo de Secretária Assistente da Educação – uma espécie de Vice-Ministra daquela pasta – no governo conservador de George Bush Filho, precisamente o mesmo que instituiu o *NCLB* em 2002.

Naquele momento, Ravitch defendeu ferrenhamente o sistema de testagem e de responsabilização com altas consequências para as instituições educacionais, tendo que, para isso, bater-se contundentemente com os críticos a esse mesmo sistema, tanto na academia como, por exemplo, nos sindicatos de professores nos Estados Unidos. No começo, como se viu, o ponto de vista de Ravitch prevaleceu, e as reformas do *NCLB* ditaram as regras do cenário educacional norte-americano ao longo de praticamente todo o início do século 21.

Posteriormente, entretanto, Ravitch saiu do governo federal, que também trocou de mãos, passando dos republicanos para os democratas de Barack Obama. Entretanto, à semelhança do que já ocorrera desde a implementação do NCLB, houve também, com o novo governo democrata norte-americano, uma adesão ou continuação da nova política de testagem e responsabilização, nos mesmos moldes do que já vinha ocorrendo antes.

Foi quando a massa de críticas reiteradas a esse sistema, possivelmente junto com uma visão gerada de uma perspectiva histórica um pouco mais ampla, que Ravitch se deu conta de que, segundo sua impressão, o sistema educacional americano não estava, de fato, se beneficiando dessas novas medidas, pelos mesmos motivos esboçados há pouco neste texto. Segundo Ravitch:

Os fundamentos de uma boa educação serão encontrados na sala de aula, em casa, na comunidade e na cultura, mas os reformadores do nosso tempo continuam a procurar por atalhos e respostas rápidas. Desvinculados de qualquer filosofia genuína da educação, as atuais reformas irão nos desapontar, assim como fizeram outras no passado. Nós iremos, com o tempo, vê-las como distrações, desvios equivocados e oportunidades perdidas. É tempo de reconsiderar não apenas os detalhes das atuais reformas, mas também a nossa própria definição de reforma. (...)

As nossas escolas não irão melhorar se os políticos que nos governam se meterem no território pedagógico e tomarem decisões que deveriam ser feitas por educadores profissionais. O Congresso e os legisladores estaduais não deveriam dizer aos professores como ensinar, tanto quanto eles não devem dizer aos cirurgiões como realizar operações. (RAVITCH, 2011)

Com palavras assim, a ex-defensora das políticas agressivas de testagem e responsabilização acabou fazendo um *mea culpa*, ao mesmo tempo em que ecoou e resumiu uma ampla crítica que já se esparramara por boa parte da sociedade norte-americana. Vem se constatando nos EUA, dessa forma, uma espécie de fim da ilusão de que seria possível para a sociedade norte-americana, a curto prazo, dispor de um sistema capaz de atingir os níveis sonhados de qualidade e equidade no seu sistema educacional.

## 2.5. NOVOS TEMAS E TENDÊNCIAS DE REFORMA

Hoje em dia, nos Estados Unidos, tem tomado um novo impulso o movimento em prol de uma educação mais ampla e de qualidade. Nesse sentido, as discussões não chegam a trazer grandes novidades em relação a muita coisa que já havia sido dita antes.

Na verdade, uma boa parte do novo discurso em prol do desenvolvimento de novas habilidades dos alunos do sistema escolar americano já haviam surgido poucas décadas antes, mais precisamente em 1991, ano em que o Ministério do Trabalho dos Estados Unidos – lá chamado de *US Department of Labor* – publicou o relatório SCANS, sigla que deixa explícita a ideia de um exame cuidadoso (*scan* em inglês) e que é a abreviatura de *Secretary's Commission on Achieving Necessary Skills*, ou seja, uma comissão da secretaria de governo sobre a aquisição de habilidades necessárias, por parte dos alunos do sistema educacional norte-americano.

Logo no início, os autores do relatório, em uma declaração, afirmaram que, em seu entendimento, “as escolas fazem mais do que simplesmente preparar as pessoas para ganhar a vida. Elas também as preparam para viver plenamente, para participar da vida comunitária, para criar suas famílias e desfrutar o lazer que é o fruto de seu trabalho. E uma educação sólida é a sua recompensa” (US DEPARTMENT OF LABOR, 2015).

Após 12 meses reunindo-se, estudando e entrevistando empresários, empregadores públicos, trabalhadores, etc. a comissão elaborou uma lista de habilidades e atitudes que, segundo sua visão, poderiam ser importantes no ambiente de trabalho do mundo contemporâneo. Com base nisso, chegou-se a três conclusões básicas, que foram as seguintes: (1) a necessidade de todos os alunos do Ensino Médio nos EUA se formarem após terem desenvolvido plenamente um conjunto de competências de elevada importância para a sua vida futura; (2) a urgência das escolas também passarem por uma evolução semelhante, no trato com seus alunos; e (3) a necessidade de se pensar o desempenho educacional em termos estratégicos da segurança nacional americana.

Por sua vez, essa “nova” abordagem educacional preconizava a utilização produtiva de cinco tipos de insumos, que são os seguintes: (1) recursos físicos, como verbas, trabalho, tempo, espaço e materiais; (2) desenvolvimento de habilidades interpessoais, como a liderança, o trabalho em equipe, a negociação e o respeito à diversidade cultural; (3) o uso de informações para fins de análise de dados, concomitante com a capacidade de comunicação; (4) o domínio de sistemas capazes de, quando aplicados, lidar eficientemente com, por exemplo, o monitoramento e a correção do desempenho, e, por fim, (5) o uso de tecnologias,

adequadas e devidamente aperfeiçoadas e atualizadas para as mais diversas tarefas no mundo de competição global dos dias de hoje.

E, para fazer tudo isso, era preciso também promover entre os estudantes, a consolidação de um alicerce fundamentado em outros três requerimentos: o desenvolvimento das habilidades básicas (como ler, escrever, falar, ouvir, calcular), das habilidades de pensamento (como a criatividade, a aprendizagem e a argumentação), bem como também das qualidades ditas pessoais, do que são exemplos a responsabilidade individual, a sociabilidade e a responsabilidade. (US DEPARTMENT OF LABOR, 2000)

À publicação do SCANS, outros esforços se seguiram aproximadamente nesta mesma linha. Um fator adicional e, de certa forma, complicador dessa nova fase da discussão sobre os objetivos e missões da educação nos Estados Unidos, veio a ser a questão de como mensurar essas habilidades. Afinal de contas, o sistema educacional americano vinha experimentando uma vigorosa tendência de incorporação dos métodos de testagem em grande escala. A partir daí, a necessidade de avaliar os conteúdos trabalhados em sala de aula virou quase um lugar comum na política educacional americana.

Entretanto, agora se estava falando, não mais, digamos, de leitura, matemática e ciências naturais ou humanas, habilidades essas que, por sua vez, já suscitavam e suscitam desafios complicadíssimos de se resolver quanto, por exemplo, à questão da validade e da fidedignidade dos testes. Só que, como se isso não fosse suficiente, enfatizava-se agora a necessidade de testar outras habilidades ainda bem menos “concretas” do que aquelas que até então vinham sendo tratadas, como, por exemplo, a liderança, a tolerância ou apreço pela diversidade, a criatividade, a comunicação, etc.

No caso das chamadas disciplinas convencionais, o desafio de mensurar a proficiência já se resvalara em muitos casos para a adoção de medidas mais “simples” ou pontuais daquilo que o aluno deveria ter aprendido, o que, muitas vezes, acabava sendo feito por meio de testes de múltipla escolha. Porém, agora, esse tipo de testes – que possui muitas vantagens na avaliação convencional de disciplinas também convencionais – certamente passava a ser de bem menos valia ou aplicação. Portanto, uma boa parte dos esforços norte-americanos mais recentes tem sido gasta na tentativa de se criarem instrumentos e sistemas capazes de testar essas habilidades, em sua grande maioria, não-cognitivas.

Porém, antes de testar a significância ou importância das habilidades não-cognitivas, é preciso definir, com um grau razoável de precisão, o que elas, de fato, vêm a ser. Neste sentido, um esforço significativo iniciou-se em 1996, através da criação do *Common Core Movement* ou *Movimento para um Núcleo Comum*, uma iniciativa da Associação Nacional de

Governadores dos Estados Unidos (*National Governors Association*, ou NGA), juntamente com o Conselho de Secretários Estaduais de Educação daquele mesmo país (*Council of Chief State School Officers*, ou CCSSO).

Não obstante, cabe lembrar que as habilidades descritas pelo relatório final dessa associação, na verdade, diziam respeito àquilo que os estudantes deveriam saber ou serem capazes de fazer, porém, *não* eram definidas formalmente como “habilidades do século 21”, ou algo do gênero. E assim, as organizações e indivíduos que tratam especificamente dessas novas competências, veem-se na necessidade de “garimpar” referências a elas, existentes nos currículos formais dos diversos sistemas, na ausência de uma legislação educacional específica para isso. Tal fato mostra que, nos Estados Unidos, e possivelmente em muitos outros lugares, como no Brasil, a referência às chamadas habilidades ou competências do século 21, na verdade, não chega a representar propriamente uma novidade em relação ao que já se vinha falando antes, mas encontra-se diluída em um conjunto de habilidades semelhantes, descritas em currículos pré-existentes ao surgimento formal dessa nova discussão.

Ainda em relação ao estudo conjunto do *Common Core Movement*, cabe dizer que um grande desafio por ele enfrentado, foi, ou tem sido, a notória pulverização curricular em um país tão descentralizado, em termos educacionais, como os Estados Unidos. Em função disso, os esforços da associação supracitada concentraram-se em inglês e matemática – duas disciplinas tradicionalmente consideradas como “fundamentais”, inclusive em testagem –, deixando, assim, de lado, as ciências sociais e as naturais. Em geral, o *Common Core Movement* salienta, entre outros pontos, a necessidade, no estudo do inglês e matemática, de uma ênfase maior no pensamento crítico, colaboração, comunicação e criatividade, que são todas habilidades tipicamente identificadas como “do século 21”.

Nesse aspecto, aliás, surge uma tensão entre duas correntes opostas. De um lado, existe uma tendência de se pensar a aprendizagem por meio de uma descrição e encadeamento das características que devem possuir os alunos, para se adaptarem melhor ao mundo contemporâneo e futuro. Mas, por outro lado, nos Estados Unidos, como se viu, existe também um movimento, de certa forma, no sentido oposto, que estimula um certo reducionismo curricular, pois se vincula à testagem em larga escala de habilidades acadêmicas muito específicas, e atrela consequências práticas – às vezes, particularmente altas – aos resultados dessas medidas.

Com base no que se apresentou e discutiu no presente capítulo, constata-se que os Estados Unidos atualmente passam por um período de reforma educacional, pelo qual vêm

tentando promover um aumento da qualidade e equidade de seu sistema, amparados por um significativo conjunto de estratégias e políticas federais. Uma delas é a prestação de contas, apoiada por um generalizado sistema de testagem de habilidades básicas. Outra é a preconização de uma educação que vá muito além de fornecer as habilidades básicas aos alunos, sendo também capaz de lhes proporcionar uma educação que os forme como cidadãos conscienciosos e produtivos na sociedade norte-americana.

Naturalmente, essas duas tendências não são necessariamente opostas. Muito pelo contrário, é até de se esperar que haja uma correlação positiva e significativa entre o domínio de habilidades mais complexas e o de habilidades mais básicas por parte dos alunos. Entretanto, em termos de políticas educacionais, as coisas podem se passar de forma bem diferente. Um exemplo que parece existir particularmente nos Estados Unidos – muito embora isso talvez se passe também em outros lugares –, é o fato de que, quando se põe uma grande ênfase num sistema *high-stakes* que cobra muito rigidamente avanços mensuráveis de desempenho rumo a determinadas metas, o efeito então pode ser exatamente o contrário do que se queria. Tal estratégia pode, ao menos em princípio, provocar um estímulo ao ensino de habilidades básicas para estudantes com um nível de proficiência específico, em detrimento de outras habilidades e de outros alunos, conforme já se disse ao longo deste texto.

Por outro lado, apenas essa crítica não é suficiente para se desconsiderar totalmente os sistemas de prestação de contas e de avaliação educacional. Afinal, não se pode negar que, graças a eles, houve uma ênfase no caráter social e inclusivo da educação, para dizer o mínimo. Dessa forma, a avaliação educacional, bem como o atrelamento de consequências a ela, ainda que em graus muito variados, são características que parecem ter se enraizado no sistema educacional americano, e também em vários outros.

O passo a seguir talvez seja, de certa forma, conciliar ambas essas tendências: o estímulo ao desenvolvimento das habilidades mais elevadas, paralelamente a uma “vigilância” eficiente do sistema, exercida com base, em geral, na avaliação das habilidades mais fundamentais. Nesse sentido, um caminho que parece estar tomando certo impulso é adotar estratégias que, de certo modo, evitem alguns “equivocos” historicamente cometidos. E talvez o mais notório deles seja o caso das reformas “top-bottom”, ou de cima para baixo, realizados nos anos pós-Sputnik, e também posteriormente. Ou seja, seria o caso agora de, num movimento contrário, tentar aumentar a participação dos atores educacionais mais diretos, como os diretores escolares, pedagogos, professores e até alunos e a comunidade. E assim, talvez seja possível intervir mais eficientemente no processo escolar, tornando-o capaz de fornecer uma educação de maior qualidade e equidade para seus alunos.

## 2.6. ALGUMAS OBJEÇÕES TÉCNICAS FEITAS AO NCLB

Olhando o processo de implementação do NCLB, é possível perceber que ele foi, de certo modo, bastante bem sucedido, pelo menos no que diz respeito à adoção prática da nova lei por parte dos diversos estados norte-americanos. Um dos motivos dessa grande aceitação, naturalmente, foi o já mencionado aceno de gigantescas verbas federais para os estados e os sistemas educacionais que se pautassem pela nova lei. Outra parte desse sucesso derivou do sentido de urgência de se promover um aumento significativo da qualidade educacional nos Estados Unidos, conforme também se disse há pouco neste texto. Tal percepção era, então, quase um lugar-comum, embora as formas de analisar e tratar desse problema certamente variassem, segundo o lado político-ideológico adotado pelos diversos agentes envolvidos.

Também um ponto que contribuiu para a grande aceitação do NCLB foi o seu forte caráter de inclusão social, ao exigir que as escolas prestassem contas por todos os alunos por elas atendidos, e não somente por uma amostra deles. Isso certamente veio ao encontro da opinião e das reivindicações de diversos grupos minoritários e outros nos Estados Unidos.

Além de tudo, outra característica defendida pelos proponentes do NCLB era o fato dele se basear em padrões e medidas, algo que atenuava bastante – chegando quase a eliminar – a crítica que alguns setores faziam, e ainda fazem, às metodologias de caráter mais qualitativo das avaliações educacionais, que não permitem, segundo essa visão, uma correta mensuração do nível de qualidade das escolas. Portanto, tinha-se, agora, uma “métrica”, que serviria para fornecer resultados supostamente “inquestionáveis”, relacionados ao cumprimento ou não de metas numéricas específicas.

Não obstante, críticas de caráter mais técnico ao NCLB também começaram a surgir, sendo que algumas das mais contundentes vieram de alguns estudiosos de psicometria. Nesse sentido, por exemplo, em um artigo sobre os possíveis modelos estatísticos a serem utilizados pelos sistemas estatais de responsabilização educacional nos Estados Unidos, Hill e DePascale (2003) alertam para o perigo de se fazerem escolhas técnicas específicas, ao mesmo tempo em que, numa nota meio otimista, sugerem a possibilidade desses sistemas terem um grau minimamente aceitável de qualidade, frente às necessidades da sociedade que eles pretendem satisfazer. Esses autores então dizem que:

O NCLB requer que os estados [norte-americanos] estabeleçam sistemas de prestação de contas que sejam tanto válidos quanto confiáveis. [Entretanto,] caso se siga a palavra da lei ao pé da letra, não há nenhum modelo que satisfaça a esses requerimentos. Porém, se a lei for interpretada de modo mais flexível, então é possível criar um modelo desse jeito. Os estados necessitam tratar desse problema com muito cuidado, caso consigam

equilibrar adequadamente as múltiplas probabilidades de tomar decisões incorretas acerca das escolas.

Estes mesmos autores também chamam a atenção para uma relevante distinção entre sistemas de avaliação e sistemas de responsabilização, visto que os primeiros se dedicam a coletar e devidamente processar um conjunto de informações com o propósito de aferir a qualidade educacional, ao passo que os segundos operam com o objetivo de atrelar consequências a esses resultados. E um problema considerável surge quando os sistemas avaliativos apresentam imperfeições de medida, tornando potencialmente complexas, e até perigosas, as práticas de distribuir consequências, positivas ou negativas, derivadas das conclusões das aferições que se fazem.

De fato, nos Estados Unidos, segundo observa Rogosa (2002), a partir do ano 2000, nada menos que 40 dos 50 estados norte-americanos começaram a avaliar suas escolas com base em testes padronizados de proficiência. E a metade deles também passou a atrelar compensações financeiras aos resultados dessas avaliações. Somente a Califórnia destinou 677 milhões de dólares em incentivos docentes no ano de 2001, quando algumas escolas de resultados mais expressivos chegaram a receber 25.000 dólares em bônus, para serem distribuídos entre seus profissionais.

#### 2.6.1. A imprecisão das medidas

Outra crítica de caráter mais técnico que costuma ser feita aos sistemas de incentivo observados recentemente nos Estados Unidos, segundo Kane e Steiger, (2002) é o fato de que tais sistemas, via de regra, destinam uma enorme quantidade de recursos a programas que, por sua vez, entregam bônus e outras recompensas com base em avaliações muito pouco válidas e fidedignas. Ou, em outras palavras, está-se aqui gastando muito dinheiro sem critério. E as consequências disso são nefastas sob mais de um ponto de vista: não somente porque se trata de uma injustiça em si – afinal de contas, pode haver casos de escolas estarem sendo premiadas ou punidas sem motivo justo –, como também pode ocorrer o problema da sinalização errada para os estudiosos, gestores e demais profissionais da educação. Isto porque o que talvez seja identificado por essas avaliações como uma atitude ou comportamento profissional muito bom, capaz de produzir grandes resultados, talvez não o seja na realidade, entre outros problemas.

Uma preocupação estatística diretamente associada a essas imprecisões, segundo esses autores, decorre da chamada volatilidade dos resultados das avaliações. Eles então a ela se referem como:

Essa volatilidade pode prejudicar seriamente os sistemas educacionais de responsabilização. A partir do momento em que as notas obtidas nos testes são capazes de ocasionar recompensas ou sanções, os profissionais nas escolas passam a correr um grande risco de serem punidos ou recompensados por resultados que, na verdade, estão fora de seu controle. Além do mais, a partir do momento em que essas classificações de desempenho são utilizadas para identificar as melhores práticas em educação, virtualmente qualquer filosofia [pedagógica] poderá um dia vir a ser endossada, o que só serve para aumentar a confusão acerca dos méritos de diferentes estratégias de reforma educacional.

Ainda segundo estes autores, as fontes de volatilidade costumam provir de duas principais origens possíveis. Uma delas deriva do fato de que, nas avaliações transversais, que são as mais comumente utilizadas nos Estados Unidos – e também no Brasil –, costuma-se avaliar uma mesma série escolar durante vários anos seguidos, fazendo com que, a cada edição do teste, tenha-se uma nova amostra de estudantes, que é diferente tanto das que vieram antes dela quanto das que vêm depois. E, ainda que certas características escolares, como o índice socioeconômico médio do seu respectivo alunado – não costumem variar muito de um ano para o outro, há, por outro lado, variações consideráveis que podem ocorrer numa escola em um curto espaço de tempo. Um simples exemplo disso é a possível deterioração do clima de aprendizado dentro de uma sala de aula decorrente do surgimento de uma liderança “negativa” entre os alunos, algo que, por sua vez, pode diminuir o resultado médio de proficiência de uma escola, particularmente nos casos em que a turma onde eventualmente apareça esse problema é uma das poucas, ou mesmo a única, existente naquela instituição docente, como é, de fato, muitas vezes o caso.

A tudo isso se junta uma segunda fonte de variação, esta de caráter essencialmente aleatório e praticamente incontrolável: um leque virtualmente infinito de possibilidades da ocorrência de eventos capazes de introduzir distúrbios ou alterações que podem impactar consideravelmente os resultados de um teste. Exemplos disso são, digamos, um barulho incontrolável e inadvertido no dia da prova, como o de um cão latindo ou de uma britadeira funcionando na rua, um surto de gripe, o estabelecimento de uma relação extremamente cordial – ou, ao contrário, bastante inamistosa – entre os examinados e os aplicadores do seu teste, etc. (KANE, 2002)

Passemos, então, a discorrer um pouco sobre as possíveis formas de atacar esses dois problemas.

### 2.6.1.1. Os problemas das estimativas de modelo agregado

Uma proposta para o problema acima mencionado, da utilização dos dados transversais, seria considerar a alternativa de usar dados longitudinais, pelos quais é possível traçar uma curva de evolução da proficiência de cada estudante individualmente. Para isso, faz-se necessário que os sistemas de avaliação criem mecanismos de rastreamento dos alunos, mesmo no caso daqueles que, por exemplo, eventualmente mudem de escola durante o intervalo de tempo considerado. Esse conjunto de dados pode se constituir de notas obtidas em duas ou mais edições diferentes de um teste. Dessa forma, é possível calcular a variação da nota do aluno à medida que aumenta seu tempo de escolaridade. Portanto, a qualidade da escola, neste caso, não mais se associa simplesmente ao status da proficiência, mas sim ao valor que a escola agrega àquilo que já existe, decorrente do desenvolvimento da proficiência anterior, que é toda descontada nos cálculos desses novos escores.

Dessa forma, segundo vários autores, como Ballou (2002), a superioridade conceitual dos modelos de valor agregado (ou *value added models*, VAM, em inglês) nas avaliações educacionais reside no fato de que, segundo esta metodologia, o rendimento escolar é mensurado em termos de ganhos de aprendizagem medidos no nível do aluno – possivelmente agregado para outros níveis, como o de uma turma ou de uma escola. Assim, é possível, por exemplo, observar a variação de proficiência sofrida num determinado período – digamos, durante um ano letivo, ou ao longo de um curso ministrado por um dado professor –, tendo como base o nível de aprendizado anterior, ou uma série histórica destes últimos. Por outro lado, conforme já se falou, as avaliações transversais examinam alunos diferentes em momentos diferentes – porém, comumente em séries ou anos específicos de escolarização –, de modo a se perder totalmente o controle da proficiência anterior dos alunos examinados numa determinada ocasião. E assim, os resultados transversais ficam mais suscetíveis a variações possivelmente aleatórias associadas ao nível médio de proficiência anterior dos estudantes avaliados, quando comparados com os resultados de VAM (RAUDENBUSH, 2004).

Entretanto, a despeito dessa suposta superioridade dos modelos de valor agregado em relação aos métodos transversais, Ballou observa que, mesmo os modelos de VAM costumam se ressentir de graves problemas, associados a, pelo menos, três fatores: (a) a mensuração dos ganhos de aprendizado costuma ser pouco acurada; (b) ao menos uma parte dos ganhos pode ser atribuída a outros fatores, que não a qualidade da educação oferecida e (c) não há bases

muito seguras para se compararem os ganhos de proficiência de alunos situados em níveis diferentes de habilidade. Falemos um pouco sobre estes problemas.

### *A instabilidade dos modelos de VAM*

Uma das críticas que se fazem aos modelos de VAM refere-se ao fato de que os escores produzidos por esta metodologia tendem a ser muito mais instáveis do que os escores de status apenas. Braun (2004), por exemplo, argumenta que esse problema deriva do fato de que a equação da proficiência  $Y_{ij}$  de um aluno  $i$  na escola  $j$  pode ser descrita como uma soma de fatores, que, para os propósitos do presente texto, reescrever-se-á como:

$$Y_{ijt} = \alpha_{ijt} + \beta_{ijt} + \varepsilon_{ijt}$$

Observa-se, nesta equação, que seu lado esquerdo é a proficiência do aluno. Já no lado direito, há uma soma de três valores. Uma primeira parte –  $\alpha$  – corresponde a um vetor ou um conjunto de características dos alunos; a segunda –  $\beta$  –, a características das escolas por eles frequentadas e a terceira –  $\varepsilon$  –, a um erro aleatório. Pois bem, caso esse aluno seja testado de novo, digamos, no ano seguinte, basicamente todas essas características dele próprio e também de sua respectiva escola incluídas na primeira equação, repetir-se-ão na segunda – a do tempo posterior –, como, por exemplo, o gênero do aluno, seu respectivo índice socioeconômico, o ISE médio da escola, a rede de ensino à qual esta última pertence, etc. Haverá, então, somente um valor adicional –  $\delta$  –, correspondente ao possível incremento da proficiência de um ano para o outro, segundo a equação:

$$Y_{ij(t+1)} = \alpha_{ij(t+1)} + \beta_{ij(t+1)} + \delta_{ij(t+1)} + \varepsilon_{ij(t+1)}$$

Como se vê, esta última equação é quase igual à penúltima, desta diferindo-se somente pelo salto temporal da proficiência,  $\delta$ , que, comparativamente representa uma parte muito pequena do lado direito da mesma equação. Isto porque as características coletivas e individuais das escolas e dos alunos conseguem explicar muito mais a variância dos resultados do que “somente” um ganho anual, por maior que este venha a ser. Tudo isto faz com que a determinação de resultados baseada nos próprios escores (os valores do lado esquerdo dessas últimas equações) apresente uma grande estabilidade, visto que as características das escolas e dos alunos em geral não variam muito, digamos, entre duas edições anuais de um dado teste.

Por outro lado, caso optemos por trabalhar com uma metodologia de VAM, necessariamente seremos levados a um procedimento de subtrair o primeiro resultado do segundo, calculando a diferença:  $[Y_{ij(t+1)} - Y_{ijt}]$ . Mas, ao fazermos isto, no lado direito da

equação, varrem-se do mapa todas as variáveis referentes aos alunos, turmas, escolas, municípios, redes, entre outras coisas mais, visto que os diferentes vetores  $\alpha$  e  $\beta$  mudam pouco ou quase nada de um momento para outro... Sobra apenas a variação anual,  $\delta$ , invariavelmente acompanhada de seu respectivo erro aleatório.

Mas cabe, neste ponto, observar que esse erro aleatório, quando antes se considerava o status da medida, representava tão somente uma pequena fração da nota final do aluno, pois tendia a ser ofuscado pelo grande poder explicativo das variáveis preditivas das escolas e dos alunos. Entretanto, agora, uma vez que estas últimas desapareceram, o erro aleatório passou a desempenhar um papel proporcionalmente muito maior do que antes, a ponto de comprometer seriamente a estabilidade do escore final, que vem a ser, via de regra, simplesmente o resultado da soma desse erro com o acréscimo temporal da proficiência.

Nesse sentido, um possível caminho, se não para corrigir esse problema totalmente, mas pelo menos para mensurá-lo e assim, tentar administrá-lo de um modo minimamente satisfatório – seria implementar procedimentos para se mensurar estatisticamente o erro aleatório. E, com esse propósito, os métodos de valor agregado costumam empregar procedimentos consideravelmente sofisticados para levar em conta tais erros.

Porém, ao fazerem isso, se, por um lado, cumprem o seu papel de aumentar a correção dos métodos de avaliação, por outro incorrem no problema nada trivial de deixarem esses testes ainda mais difíceis de compreender e interpretar, principalmente para uma boa parcela do público que absorve, ou deveria absorver, tais informações: os alunos e seus pais, e até mesmo os professores e diretores das escolas de ensino básico, que via de regra não têm um treinamento estatístico suficiente para compreender os complexos cálculos que se incluem nos resultados desses testes. E, nesse ponto, a avaliação passa a falhar em um ponto de grande importância, inclusive prática: a de ser algo facilmente inteligível para os atores a ela associados, o que dificulta, inclusive, a adequada aceitação e utilização dos resultados por parte dos diversos atores nela envolvidos.

Dessa forma, segundo Ballou, as metas devem ser tratadas com base em curvas teóricas obtidas a partir dos resultados reais, porém não devem se basear especificamente nos resultados em si. Entretanto, em relação a tudo isso, um problema prático bastante comum acontece devido ao fato de haver muito mais dados disponíveis para determinados alunos ou escolas do que para outros. E assim, no caso de haver poucas informações sobre determinadas escolas, os níveis de fidedignidade de suas respectivas medidas costumam ser muito mais baixos, o que complica e muito, por exemplo, possíveis decisões acerca da qualidade do ensino e que levem em conta todos os alunos e escolas examinadas.

Além disso, as discrepâncias de resultados educacionais costumam variar dependendo das disciplinas. Por exemplo, é sobejamente conhecido o fato de que o “capital cultural” do aluno – derivado em boa parte do nível socioeconômico e também cultural de sua família – encontra-se mais associado ao desempenho em leitura. Por outro lado, essa influência é menor no que se refere à matemática, disciplina cuja formalização depende mais da intervenção escolar. E, em decorrência disso, torna-se mais difícil detectar diferenças entre níveis da aprendizagem de leitura do que diferenças em matemática, do ponto de vista das escolas que as proporcionam.

Por sua vez, Doran (2004) aponta para outro sério problema psicométrico que surge em decorrência da equalização vertical dos testes, a qual naturalmente é necessária para os estudos de VAM. Segundo este autor, tal problema ocorre porque a variância dos parâmetros estimados dos itens utilizados nas equalizações tende a diminuir com o tamanho amostral. Entretanto, este último *não* influencia, e muito menos diminui, os erros de “linkagem”, os quais são, por outro lado, dependentes da quantidade e das características dos itens incluídos no processo de equalização. Dessa forma, uma simples – e, talvez, inevitável – variação de procedimentos e parâmetros utilizada entre duas equalizações é capaz de produzir ruídos que nada têm a ver com os eventuais ganhos de proficiência dos examinados. E, entre as recomendações finais deste autor, incluem-se as de “compreender melhor a natureza do erro de linkagem [e] incorporar este erro às estimativas dos escores de VAM.”

Outras críticas à metodologia do valor agregado são feitas por Braun, que observa que tais métodos, apesar de todo o entusiasmo que têm despertado em administradores educacionais, têm revelado a ocorrência de uma série de pontos preocupantes. Braun discorre sobre os modelos de VAM conforme desenvolvidos por Sanders e aplicado nos estados norte-americanos do Tennessee, Ohio e Pensilvânia, muito embora a discussão que ele faça possa se estender, em boa medida, a todos os modelos de VAM.

Em primeiro lugar, Braun observa que aquilo que se faz usualmente é tratar os indicadores de valor agregado como uma medida acurada da contribuição do professor ou da escola para o desenvolvimento da proficiência do aluno, a partir do que se faz uma interpretação causal dessas estimativas estatísticas. Entretanto, um problema que pode comprometer a validade desse procedimento é o fato de que, neste caso, não existe uma randomização de escolas e alunos, como teria que haver para que se se tratasse de um experimento aleatório. Obviamente, não há como destinar alunos a escolas específicas por simples sorteio, visto que os primeiros costumam se matricular em um ou outro estabelecimento de ensino por causa de diversos fatores independentes da vontade das

autoridades educacionais, como a escolha dos pais ou dos próprios alunos, o seu local de residência, seu nível socioeconômico, etc. E essa falta de aleatoriedade no processo de destinação dos alunos às escolas torna praticamente impossível a realização de inferências causais do tipo que se fazem nos testes controlados de laboratório.

Por outro lado, e falando em controles, também é possível controlar estatisticamente os resultados desses testes, muito embora apenas isso não seja suficiente para eliminar os problemas decorrentes da falta de randomização. E assim, a utilização dos resultados dessas avaliações para propósitos de consequências mais elevadas – como a compensação financeira dos profissionais envolvidos ou possíveis intervenções nas escolas – corre o risco de se tornar inadequada, caso se baseiem apenas nos resultados dessas avaliações.

Além disso, segundo Braun, os alunos podem diferir bastante uns dos outros quanto a diversas características relevantes que impactam a proficiência – como sua própria “taxa de progresso acadêmico, o apoio da família, a motivação, os hábitos de estudo, a dinâmica interpessoal e outras características relevantes”. Dessa forma, fica difícil separar os efeitos supostamente exercidos pelo professor ou pela escola de todo esse emaranhado de possíveis efeitos concorrentes.

Pois bem, diante de tantas dificuldades, alguns autores adotam a crença de que, mesmo nas situações cujo objetivo é identificar e premiar o desempenho, talvez o uso de sofisticados métodos de valor agregado não valha a pena. Não obstante essas críticas, alguns sistemas educacionais norte-americanos, como o do Tennessee e o da cidade de Dallas, no Texas, vêm utilizando métodos de valor agregado como uma opção de avaliação educacional para se adequar às exigências do *No Child Left Behind Act*. Cabe dizer que o NCLB não requer dos sistemas educacionais que realizem avaliações de valor agregado. Ao invés disso, a lei americana concede uma liberdade considerável para que os estados estabeleçam seus próprios parâmetros de desempenho e métodos de mensurá-los, muito embora, independentemente de quais métodos sejam esses, eles têm que ser coerentes e também estar de acordo com as imposições mais gerais, previstas na lei, do progresso anual adequado, conforme já se comentou anteriormente.

No caso do Tennessee, emprega-se uma metodologia de VAM para se tentar detectar diferenças de desempenho entre professores do ensino básico e médio. Nesse sentido, coletam-se dados censitários anuais de todas as séries dessas etapas de escolarização, atentando-se para rastrear longitudinalmente os alunos de uma série para a outra, e também, naturalmente, identificando a turma à qual o aluno pertence a cada ano. Com base nesses dados, o Estado do Tennessee tem conseguido isolar cerca de um quinto dos professores de

seu sistema de ensino, e que se distribui entre os dois extremos – maior e menor – de desempenho. No entanto, todos os demais 80% dos professores permanecem numa região intermediária, para a qual é difícil tomar decisões devido à ambiguidade dos dados a longo prazo.

## 2.7. A DEFESA DAS AVALIAÇÕES

Em um relatório feito por Walter Way em 2006 para o Educational Testing Service (ETS), este autor coteja as posições antagônicas tomadas, de um lado, por Kane e Staiger e por Linn e Haugh (2002) – segundo os quais os resultados das avaliações nos Estados Unidos, ao menos naquela ocasião, estavam sendo altamente comprometidos pelas imprecisões estatísticas das medidas de desempenho – com as opiniões contrárias defendidas por Rogosa (2012), de que, apesar das inevitáveis flutuações estatísticas das medidas, ainda assim era possível obter medidas confiáveis de desempenho das escolas. Nesse sentido, Way salienta o fato de que as possíveis deficiências estatísticas dos atuais sistemas avaliativos norte-americanos ainda são objeto de uma acesa controvérsia.

E, de fato, uma defesa vigorosa dos atuais sistemas de avaliação e de prestação de contas nos Estados Unidos foi feita por Rogosa, em uma série de artigos. Em um deles (2014a), este autor realiza uma análise de 304 escolas que receberam um prêmio – *AB1114 Awards* – pelo elevado salto de proficiência que apresentaram durante dois anos consecutivos (entre 1999 e 2000). A seguir, ele examina o que aconteceu com essas mesmas escolas na variação seguinte. Observou então que, em geral, essas escolas tenderam a apresentar novos ganhos na avaliação seguinte. Naturalmente, essa tendência de melhorar pela segunda vez consecutiva tende a declinar na segunda variação quando o primeiro salto for muito elevado. Ainda assim, mesmo nesses casos de elevadas taxas iniciais de progresso, na maioria das vezes essas escolas continuam aumentando suas médias, ainda que a uma taxa menor. A mensagem passada pelo autor é que, ainda que existam flutuações nos resultados de um ano para o outro, há também um considerável componente de estabilidade nos processos de desenvolvimento educacional conforme mensurado pelos testes. Dessa forma, segundo ele, a metáfora do “dente de serra”, segundo a qual as escolas têm suas médias sucessivamente aumentando e diminuindo, porém sem praticamente sair do lugar a longo prazo, não é muito adequada para a maioria dos casos observados. O mais correto seria usar uma metáfora de uma “subida com degraus”, pela qual as escolas, ora avançam menos, ora avançam mais,

porém, a longo prazo, acabam tendo seus desempenhos médios passando por significativos aumentos.

Em outro estudo, Rogosa se detém sobre as propriedades estatísticas dos programas de premiação das escolas, debruçando-se sobre a questão dos chamados “falsos positivos” e “falsos negativos”. Este último tópico, naturalmente, não é exclusividade das situações educacionais, mas se refere também, por exemplo, à literatura estatística médica, onde, por razões óbvias, há uma grande preocupação profissional em evitar esses dois tipos de erro, principalmente no que se refere à detecção de enfermidades graves.

Entretanto, no caso especificamente tratado por Rogosa, há complicações adicionais que, por exemplo, não tendem a serem observadas no Brasil. Isto porque, pela atual legislação norte-americana, baseada no NCLB, é preciso que as escolas apresentem progressos anuais adequados, não somente como um todo, mas também para cada um dos grupos socioculturais ou étnicos relevantes que compõem o seu respectivo corpo discente, conforme já se comentou anteriormente. Nesse sentido, por exemplo, os alunos são divididos entre pertencentes ou não a minorias, como as étnicas, e também separados quanto ao status socioeconômico – baixo ou elevado –, além de também o serem por razões linguísticas – nativos ou não do inglês. Tudo isso deu origem a uma metáfora, conhecida na recente literatura estatística educacional americana, de “conduzir gatos” (*herding cats*), ou seja, de fazer com que os diferentes e variados grupos de alunos, que, geralmente, possuem trajetórias bastante idiossincráticas, passem todos a caminhar na mesma direção – a do progresso –, o que nem sempre parece ser uma tarefa muito fácil de se realizar...

Neste contexto, portanto, um agravante para as dificuldades de se mensurar o desempenho reside no fato de que, quanto mais os alunos de uma escola se dividem em grupos menores, mais aumenta a instabilidade de suas respectivas medidas de proficiência. Um modo adotado por Rogosa para tratar desse problema é lançar mão de um procedimento de amostragem *bootstrap*, pelo qual, por meio de simulações computacionais, aumenta-se enormemente a quantidade de casos válidos para a avaliação, por meio da introdução artificial de “novos” casos, obtidos a partir da distribuição real de escores de fato obtida nos testes (MOONEY, 1984). Com essa metodologia, Rogosa (2014b) de fato, consegue reduzir a ocorrência de falsos positivos e de falsos negativos no sistema de premiação das escolas, o que, na prática, significa fazer com que haja critérios mais justos pelos quais uma escola recebe ou deixa de receber um prêmio pelo suposto aumento da proficiência de seus respectivos alunos. Porém, ainda mais importante do que isso, esse estudo de Rogosa também

serve para demonstrar que as imprecisões estatísticas não devem ser vistas simplesmente como uma barreira intransponível para os sistemas de responsabilização. Antes, ainda que se reconheça a existência de tais problemas, é preciso também levar em conta a possibilidade de, ao menos, reduzi-los a níveis aceitáveis, e seguir em frente, mais ou menos da mesma forma como os exames médicos também o fazem...

Outros autores – como Hill e DePascale – também salientam o fato de que é possível que os sistemas de prestação de contas se apresentem bastante fidedignos mesmo em situações em que os escores individuais dos examinados não o sejam. Isto é o que acontece, por exemplo, quando se aplicam testes relativamente curtos a um grande número de alunos, pertencentes, por exemplo, a uma determinada rede estadual de ensino. De fato, em casos assim, o pequeno tamanho dos testes é uma vantagem caso o objetivo principal seja o de avaliar o sistema como um todo, pois tal procedimento aumenta o tamanho amostral de alunos. Mas, por outro lado, isto faz com que os testes, do ponto de vista individual, tenham baixa fidedignidade devido ao seu tamanho. E assim, não seria recomendável utilizar esses testes, por exemplo, para tomar decisões com consequências muito altas para os alunos, ou mesmo para outros níveis pouco superiores de agregação, como, por exemplo, turmas ou mesmo escolas, principalmente as pequenas.

Nesse sentido, esses autores fazem algumas simulações para provar que, por exemplo, um aumento de 10% nos valores brutos dos escores de um ano para o outro, sob dadas condições, apresentam uma probabilidade de 12,6% de não conseguir ser captado após a segunda testagem – resultando, assim, num “falso negativo” – quando a escola avaliada possui 50 alunos testados. No entanto, se o número de alunos subir para 200, a probabilidade de ocorrência de um falso negativo cai drasticamente para 0,6%. Os autores também recomendam a utilização de pelo menos dois métodos diferentes para se estimar a fidedignidade, devido ao fato de que pode haver perturbações de diferente natureza em cada um deles.

Quanto ao fato de se incluírem os resultados de uma ou mais disciplinas, embora isso tenha uma consequência benéfica de aumentar o tamanho amostral, o que atenua um problema de medida, por outro lado, chama a atenção para o nada trivial problema da responsabilização. Isto porque, embora a correlação entre os resultados de diferentes disciplinas, tanto no nível dos alunos, das escolas e em outros, seja positiva e via de regra bastante acentuada, ainda assim ela está longe de ser perfeita. Assim, existe também, e talvez

não seja tão insignificante, a possibilidade de uma escola que cresceu bastante, digamos, em leitura, porém não em matemática, ser destacada ou premiada como se tivesse progredido em ambas as disciplinas, ou mesmo em várias outras não testadas, sobre as quais nada se sabe!

Diante de tudo isso, algumas recomendações que estes últimos autores ainda fazem são: (1) de que os dados devem apresentar uma alta correlação com os objetivos curriculares, (2) que a escala de proficiência deve ser abrangente, de modo a se estender até ambos os extremos da possível distribuição de resultados e (3) que deve haver 40 ou mais itens nos testes aplicados, para fins de fidedignidade das medidas.

Feitas essas considerações acerca da evolução dos sistemas de avaliação educacional nos Estados Unidos, passaremos, na sequência deste texto, a ver como se tem dado a evolução desse mesmo tema aqui no Brasil, que, como se disse, tem sua história de avaliação educacional bastante influenciada pela experiência norte-americana. Portanto, a história recente da avaliação educacional em larga escala no Brasil, e as implicações mais diretas disso sobre o tópico principal de nosso estudo, serão o assunto do qual se tratará um pouco mais detidamente no capítulo seguinte.

### 3. AS AVALIAÇÕES EDUCACIONAIS EM GRANDE ESCALA NO BRASIL

#### 3.1. INTRODUÇÃO

Conforme se comentou anteriormente, a implementação dos sistemas de avaliação educacional no Brasil teve sua concepção e metodologia fortemente inspiradas nos modelos norte-americanos discutidos no capítulo precedente. Entretanto, uma diferença em relação ao caso norte-americano refere-se ao tempo de funcionamento desses sistemas. Enquanto que, nos Estados Unidos, o NAEP, por exemplo, já vem existindo desde a década de 1960, no caso brasileiro, as avaliações educacionais em larga escala são um fenômeno que vêm sendo observado particularmente a partir da última década do século 20.

Contudo, apesar de relativamente nova em nosso país, tal tendência avaliativa vem ganhando um espaço cada vez maior, não somente em termos de disseminação pelos diferentes estados da Federação, como também quanto às implicações diretas para as escolas avaliadas por esses sistemas, notadamente as públicas. Essa tendência nacional de avaliar e, em alguns casos, estabelecer consequências para o desempenho educacional no contexto especificamente brasileiro, é o objeto de estudo do presente capítulo, que, para fins de exposição e análise do tema, divide a trajetória nacional de avaliação em larga escala em três períodos que, *grosso modo*, correspondem: (1) às suas origens; (2) à sua expansão e (3) à adoção de políticas de responsabilização, atreladas às medidas e metas de desenvolvimento educacional.

Por sua vez, a adoção de metas educacionais no Brasil encontra-se fortemente vinculada ao surgimento do Índice de Desenvolvimento da Educação Básica – Ideb –, proposto pelo Ministério da Educação em 2005. Dessa forma, no sentido de compreender melhor esse índice, aqui também se aborda, com um certo grau de detalhamento, a questão da concepção do Ideb e de seu uso para fins de monitoramento da qualidade da educação.

### 3.2. UMA VISÃO PRELIMINAR: A EXPANSÃO DA EDUCAÇÃO PÚBLICA NO BRASIL

No ano de 1950, ou seja, precisamente na entrada da segunda metade do século 20, o Brasil ainda convivia com uma situação educacional extremamente inquietante, visto que, de cada dois de seus cidadãos em idade escolar ou superior, um não sabia ler nem escrever. Em outras palavras, o Brasil tinha, nessa ocasião, uma taxa de analfabetismo de nada menos do que 50%, segundo os dados do Instituto Brasileiro de Geografia e Estatística, o IBGE. Portanto, os elevados índices nacionais de analfabetismo representavam um desafio, ao mesmo tempo, imediato e formidável, ao propósito governamental de fornecer a todos os brasileiros o acesso à educação básica.

Nas décadas que se seguiram imediatamente a esse período, observou-se uma contínua e acentuada redução do analfabetismo no Brasil, como resultado de uma considerável ampliação das redes educacionais no país. Dessa forma, em 1970, a taxa de analfabetismo reduzir-se-ia para cerca de um terço da população nacional e, em 1980, para um quarto dela. E essa tendência de expansão do ensino básico continuou a ser observada nas décadas mais recentes, fazendo com que, no ano de 2009, o analfabetismo se reduzisse para cerca de um décimo da população nacional, conforme se pode observar pelo gráfico a seguir.

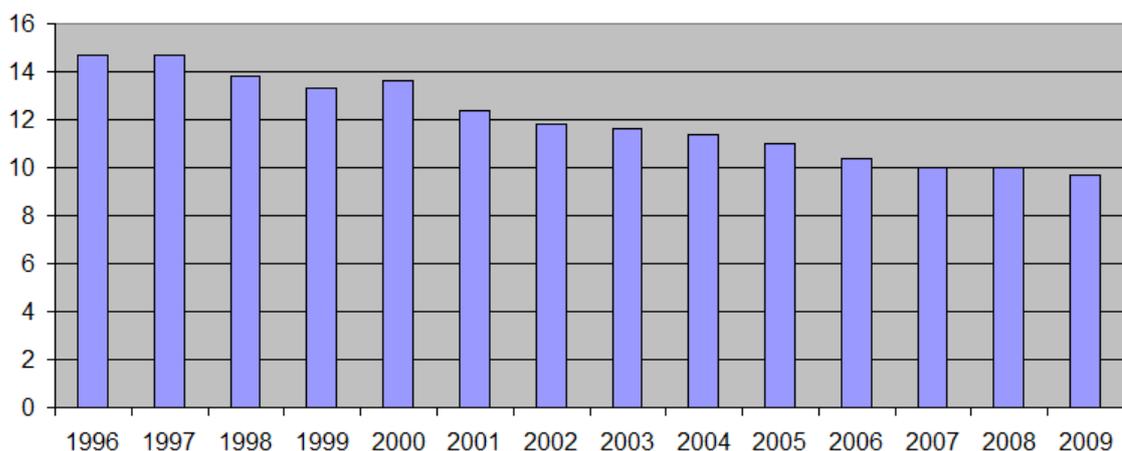


Gráfico 3.1: Percentual de analfabetismo no Brasil no período 1996-2009.  
Fonte: MEC (2012).

Naturalmente, a expansão das taxas de alfabetização não tem sido um processo uniforme, nem entre as diferentes regiões brasileiras, nem entre as diversas faixas etárias em

que se divide a população nacional. De fato, é possível constatar, por exemplo, que as taxas gerais de analfabetismo são consideravelmente mais elevadas nas regiões Nordeste, Norte e Centro-Oeste do país, comparativamente às do Sudeste e ao Sul. E também, naturalmente, elas diferem muito no que diz respeito à idade da população, sendo maiores para a parcela mais idosa dos brasileiros. Por outro lado, os mais jovens já vêm se beneficiando de um acesso muito mais amplo à educação, principalmente à de nível fundamental, conforme se pode observar na tabela abaixo:

Tabela 3.1: Taxas de alfabetização líquida e bruta no Brasil em anos recentes

Ano	Brasil			
	Ensino Fundamental <sup>1</sup>		Ensino Médio <sup>2</sup>	
	Líquida	Bruta	Líquida	Bruta
2001	93,4%	125,5%	37,5%	82,0%
2002	93,9%	124,9%	40,4%	83,2%
2003	94,0%	123,4%	43,5%	88,5%
2004	94,0%	121,2%	44,9%	89,8%
2005	94,6%	121,0%	45,8%	89,1%
2006	95,0%	119,5%	47,5%	90,2%
2007	94,6%	120,3%	48,3%	91,1%
2008	94,9%	119,2%	50,6%	92,5%
2009	95,3%	120,5%	51,1%	89,7%

1) Pessoas de 7 a 14 anos  
 2) Pessoas de 15 a 17 anos  
 Fonte: IETS -  
<http://www.iets.org.br>

Conforme se vê por esses dados, estamos atualmente caminhando para uma universalização do ensino fundamental no Brasil. Em 2009, nesta etapa de ensino, para as pessoas entre 7 e 14 anos de idade, era de 95,3% a taxa de escolarização líquida, e de 120,5% a taxa bruta no país. Este último número, o da taxa bruta, que se encontra consideravelmente acima do valor ideal de 100%, pode ser parcialmente explicado pelo fato de que muitos alunos que haviam evadido do sistema educacional foram por ele reabsorvidos, especialmente como uma consequência de programas nacionais como o Bolsa Escola, que, desde 2004, vêm concedendo às famílias pobres brasileiras um auxílio financeiro condicionado à frequência de seus filhos aos bancos escolares (SCHWARTZMANN, 2009).

Entretanto, se a frequência às escolas apresentou um grande progresso no Brasil, a qualidade e a equidade educacionais permanecem como um formidável desafio que os responsáveis pelas políticas públicas e os profissionais da educação ainda têm diante de si. Dessa forma, a discussão da qualidade educacional, e não somente da quantidade, vem

ganhando impulso nas últimas décadas e ocupando um espaço muito relevante em diversas agências governamentais, como o Ministério da Educação e os governos estaduais.

Um ponto que mereceu – e continua merecendo – destaque nessa agenda de reforma educacional no país é a questão da avaliação externa da aprendizagem, como um mecanismo de, ao mesmo tempo, fornecer um diagnóstico do nível de conhecimento atingido pelos alunos após cumprirem determinadas etapas da escolarização e de também permitir aos gestores planejar e implementar ações com vistas a aperfeiçoar os sistemas de educação no país. Nesse sentido, particularmente a partir da década de 1990, surgiu e floresceu no Brasil uma cultura de avaliação educacional externa e em grande escala, da qual inicialmente participou o governo federal, no que foi, entretanto, seguido por diversos estados da federação.

De uma certa maneira, portanto, o Brasil vem, nas duas últimas décadas, ou um pouco mais, e à semelhança do que se vem passando nos EUA, vivenciando um processo rumo à avaliação e à prestação de contas em educação. E, nesse sentido, nosso país tem experimentado uma trajetória que, *grosso modo*, pode-se dividir em três períodos, como se segue:

- 1) 1990 a 1997: Alicerces
- 2) 1997 a 2007: Expansão
- 3) 2007 ao presente: Metas e prestação de contas

É sobre cada um desses períodos que o presente texto tratará um pouco mais detalhadamente a seguir.

### 3.3. PRIMEIRO PERÍODO (1990-1997): OS ALICERCES

O primeiro período da avaliação educacional em grande escala no Brasil corresponde, aproximadamente, ao intervalo entre 1990 e 1997, anos durante os quais o país assistiu à implementação de uma série de políticas e metodologias de testagem de proficiência que viriam a exercer um impacto relevante e duradouro sobre os períodos que a ele se seguiriam. Há, pelo menos, três importantes aspectos que se devem ressaltar nessa primeira fase. Um deles foi a criação, pelo Ministério da Educação e Cultura (MEC), através do seu Instituto

Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), do Sistema Brasileiro de Avaliação da Educação Básica, o SAEB, que veio a servir de modelo a praticamente todos os sistemas de avaliação educacional em grande escala que continuariam a ser implementados no país. Uma segunda característica desse período foi a adoção das chamadas escalas de proficiência, com as quais se tornou possível averiguar, por exemplo, se os alunos avaliados estão ou não alcançando determinados níveis de conhecimento ou competência numa determinada disciplina avaliada. E, por fim, um terceiro passo importante nessa fase inicial, associado particularmente ao aspecto metodológico da avaliação, foi a adoção, pelo SAEB, dos procedimentos estatísticos da Teoria da Resposta ao Item no tratamento dos itens e dos testes aplicados, bem como na construção e na interpretação das escalas, em substituição aos modelos estatísticos clássicos até então utilizados. Passaremos, a seguir, a discorrer um pouco mais sobre cada um desses três importantes marcos da primeira fase da avaliação educacional em grande escala no Brasil.

### 3.3.1. O Sistema Brasileiro de Avaliação da Educação Básica

O SAEB foi aplicado pelo Ministério da Educação do Brasil pela primeira vez em 1990, tendo sido o resultado de alguns anos de estudos preliminares conduzidos no final da década de 1980 por técnicos e outros especialistas educacionais, inspirados principalmente, como já se disse, no NAEP norte-americano. Segundo informações do próprio INEP,

(...) o SAEB é a primeira iniciativa brasileira, em âmbito nacional, no sentido de conhecer mais profundamente o nosso sistema educacional. Além de coletar dados sobre a educação no país, procura conhecer as condições internas e externas que interferem no processo de ensino e aprendizagem, por meio da aplicação de questionários de contexto respondido por alunos, professores e diretores, e por meio da coleta de informações sobre as condições físicas da escola e dos recursos de que ela dispõe. (INEP, 2012a)

Portanto, além dos testes de proficiência aplicados aos alunos, o SAEB procura também averiguar as condições em que o ensino é oferecido e em que a aprendizagem ocorre. Essa diversidade de dados coletados permite, assim, a constituição de bancos com informações suficientes para a realização de pesquisas sobre as condições de ensino nas escolas e a sua associação com a aprendizagem dos alunos. (FRANCO, 2014)

A segunda edição do SAEB ocorreu em 1993 e, desde então, ele vem sendo aplicado ininterruptamente num intervalo de dois em dois anos. Na aplicação de 1995, houve, por parte da equipe responsável pela sua concepção e implementação, a adoção de uma série de procedimentos técnicos que visavam a alcançar uma maior comparabilidade de resultados ao longo dos anos. Entre outras medidas então tomadas, decidiu-se que o SAEB se concentraria em três séries correspondentes ao final de cada um dos três ciclos da escolarização básica, e que correspondem ao final do 5º e 9º anos (ou 4ª e 8ª séries, segundo a nomenclatura então utilizada) do ensino fundamental e à 3ª série do ensino médio. A partir daí, portanto, a concentração nessas três séries permitiria traçar a evolução dos níveis de desempenho na educação básica do país.

Além disso, outra característica relevante do SAEB é o fato de que ele trabalha segundo uma base amostral, ou seja, os testes e questionários não são aplicados a todos os alunos dessas séries no Brasil, mas sim a amostras que permitem obter uma representatividade por estado da federação (e também regiões do país), assim como por redes de ensino: federal, pública (estaduais e municipais) e particular.

Os instrumentos do SAEB consistem, basicamente, de provas de Língua Portuguesa e Matemática<sup>1</sup>, que são aplicadas aos alunos na forma de testes compostos de itens de múltipla escolha, formato este que tem o objetivo de facilitar a confiabilidade e a velocidade de correção. Por sua vez, os questionários contextuais, aplicados aos alunos, pais e diretores de escolas, tratam de uma grande variedade de dimensões que são consideradas relevantes ao processo de aprendizagem, como o nível socioeconômico dos principais atores escolares, as atitudes frente à aprendizagem, o clima escolar, as relações interpessoais e a infra-estrutura escolar. Como os testes, os questionários também se compõem de itens de múltipla escolha.

### 3.3.2. Escalas de proficiência

Em 1997, com a participação de especialistas em Língua Portuguesa e Matemática, o SAEB elaborou as especificações de cada uma das disciplinas de seus testes, que desde então foram utilizadas como padrões para a avaliação educacional no Brasil. Há documentos distintos para cada disciplina, que também são conhecidos como matrizes de referência da

---

<sup>1</sup> Embora as primeiras edições do SAEB também contivessem instrumentos para outras disciplinas, como Ciências.

avaliação. Cada uma dessas matrizes apresenta uma lista e descreve detalhadamente o que os alunos devem ser capazes de fazer com vistas a demonstrar que adquiriram um nível desejado de domínio na respectiva disciplina. Mais especificamente, as matrizes se compõem de um conjunto de descritores que se relacionam ao conteúdo daquilo que está sendo avaliado em cada série, sendo que cada descritor também se relaciona a um certo nível de operação mental cujo domínio é necessário para que os alunos demonstrem ter adquirido um certo nível de habilidade, conforme mensurado pelos itens de múltipla escolha. A ilustração abaixo apresenta um excerto de uma matriz de referência do Sistema Permanente de Avaliação da Educação Básica do Ceará, que se baseia na metodologia do SAEB e exemplifica a ideia dos descritores.

Tópicos	Descritor	Detalhamento
4 - Decodificação e compreensão de palavras.	D 10 - Decodificar palavras no padrão consoante/vogal.	Avalia se o aluno decodifica palavras formadas por sílabas no padrão canônico: consoante/vogal (ex: sí-la-ba).
	D 11 - Decodificar palavras nos padrões: vogal, consoante/vogal/consoante, consoante/consoante/vogal.	Avalia se o aluno decodifica palavras formadas por sílabas nos padrões não canônicos: vogal (ex: a-ba-ca-te); consoante/vogal/consoante (ex: tex-to, ve-ri-fi-car); consoante/consoante/vogal (ex: pa-la-vra).
	D 12 - Compreender palavras no padrão consoante/vogal.	Avalia se o aluno lê com compreensão palavras formadas por sílabas no padrão canônico: consoante/vogal (ex: sí-la-ba).
	D 13 - Compreender palavras nos padrões: vogal, consoante/vogal/consoante, consoante/consoante/vogal.	Avalia se o aluno lê com compreensão palavras formadas por sílabas nos padrões não canônicos: vogal (ex: a-ba-ca-te); consoante/vogal/consoante (ex: tex-to, ve-ri-fi-car); consoante/consoante/vogal (ex: pa-la-vra).

Figura 3.1: Excerto de uma matriz de referência de Língua Portuguesa. (CAEd, 2008)

#### As escalas do SAEB

Para cada disciplina (Língua Portuguesa e Matemática) avaliada no SAEB, foram elaboradas escalas que associam as pontuações obtidas pelos alunos nos testes à sua respectiva capacidade de resolver problemas de diferentes níveis de complexidade. Para facilitar a utilização e a interpretação dessas escalas, convencionalmente se estabeleceu que cada uma delas se compõe de valores que, *grosso modo*, variam de 125 a 500 pontos, e que os 250 pontos correspondem à média nacional da 8ª série do EF, conforme estabelecido arbitrariamente para a avaliação de 1997.

A fim de facilitar a interpretação pedagógica e a utilização dessas escalas, elas costumam ser divididas num pequeno conjunto de níveis progressivos de desempenho, como três ou quatro. Por exemplo, nos casos em que quatro níveis são definidos, em algumas situações eles costumam ser chamados de “abaixo do básico”, “básico”, “adequado” e “avançado”. A Figura 2, a seguir, ilustra a delimitação desses níveis para diferentes séries em Matemática, feita a partir da escala do SAEB, e utilizada pelo Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP).

Níveis de Proficiência	5º EF	7º EF	9º EF	3º EM
Abaixo do Básico	< 175	< 200	< 225	< 275
Básico	175 a < 225	200 a < 250	225 a < 300	275 a < 350
Adequado	225 a < 275	250 a < 300	300 a < 350	350 a < 400
Avançado	≥ 275	≥ 300	≥ 350	≥ 400

Figura 3.2: Determinação de níveis de proficiência em Matemática para diferentes séries no Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP), com base na escala nacional do SAEB. (SÃO PAULO, 2010)

### 3.3.3. A Teoria da Resposta ao Item (TRI)

O esforço sistemático do SAEB para garantir uma maior comparabilidade de seus resultados também levou esse sistema, a partir de 1997, a adotar a Teoria da Resposta ao Item, ou TRI, como método estatístico utilizado na elaboração e análise dos itens, bem como na própria elaboração dos testes. Antes dessa mudança, o tratamento dos itens e testes no SAEB se pautava pela chamada Teoria Clássica dos Testes – TCT –, que é o método comumente empregado, por exemplo, nas avaliações nas salas de aula, nas quais, digamos, a nota de um aluno se determina pelo número ou pelo percentual de acertos de itens num determinado teste. Um problema da TCT é que medidas fundamentais, como o parâmetro de dificuldade de um item ou questão, dependem dos alunos avaliados: quanto mais gente acerta um item, mais fácil ele é considerado, de modo que esse processo está sujeito a grandes instabilidades estatísticas. Por outro lado, na TRI, através de procedimentos probabilísticos cujo

detalhamento matemático foge ao escopo deste texto, é possível calcularem-se parâmetros para os itens que não dependem dos sujeitos que estão sendo avaliados, garantindo, assim, estimativas mais úteis e confiáveis desses parâmetros. Embora a TRI permita a concepção e uso de diversos modelos estatísticos para tratar dessas informações, há particularmente três que são bastante relevantes para os propósitos práticos dos testes de proficiência, e que se relacionam à/ao:

- 1) nível de dificuldade de um item ou questão;
- 2) capacidade de um item de discriminar os alunos segundo sua respectiva proficiência na disciplina avaliada (ou seja, trata-se da capacidade de distinguir quem conhece mais ou menos uma disciplina, com base no acerto ou erro do avaliado numa determinada questão);
- 3) acerto casual, ou seja, possibilidade de um aluno acertar um item mesmo tendo uma proficiência geral considerada como teoricamente baixa ou insuficiente para acertá-lo, devido à sorte de marcar inconscientemente a resposta correta num item de múltipla escolha (HAMBLETON, 1991).

A figura abaixo ilustra a chamada curva característica do item – CCI – segundo um modelo de três parâmetros da Teoria da Resposta ao Item. Nela, no eixo horizontal, é representada, numa escala arbitrária, a proficiência (*theta*) de um aluno, que é tanto maior quanto mais à direita avançamos na figura. No eixo vertical, representa-se a probabilidade de acerto –  $P(T)$  – de um dado item ou questão, que varia de zero (sem chance de acerto) até 1 (ou 100% de chance de acerto).

Observa-se que, à medida que se avança para o lado direito da figura, a curva sobe, indicando que indivíduos com maior proficiência têm uma maior probabilidade de acertar o item representado por esta curva. Aproximadamente no meio da curva, observa-se que ela sofre uma inclinação mais acentuada. Nesse trecho, diz-se que o item ou questão tem uma maior possibilidade de discriminação, visto que um pequeno deslocamento para a direita no eixo horizontal (ou seja, um pequeno aumento da proficiência do aluno), faz-se acompanhar de um aumento mais acentuado da curva no eixo vertical (ou seja, faz aumentar consideravelmente a chance de o aluno acertar a questão). Por fim, também se observa que, na extremidade esquerda da figura, a curva não se aproxima de zero, mas sim tende a assumir um valor positivo não-nulo. Isto quer dizer que, por menor que seja a proficiência do indivíduo, suas chances de acertar o item continuam sendo consideravelmente superiores a zero, visto

que é possível que ele chegue na resposta correta através de um “chute”, ou acerto casual, como frequentemente ocorre nos casos de questões de múltipla escolha.

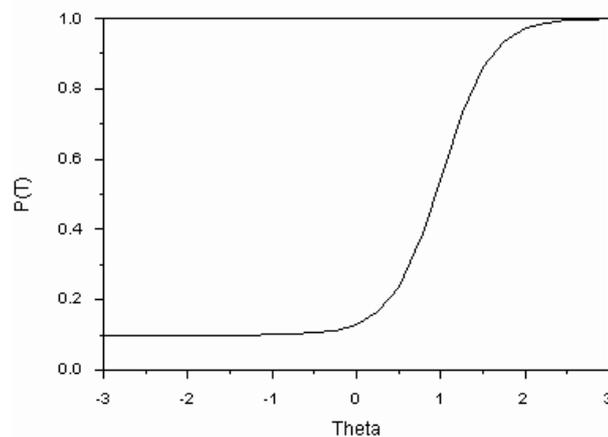


Figura 3.3: Curva característica de um item de três parâmetros segundo a Teoria da Resposta ao Item. (BRANNICK, 2012)

Uma forma comum de garantir a comparabilidade entre, por exemplo, testes feitos numa mesma região em momentos diferentes, ou entre diversas regiões num mesmo momento, é inserir num teste um número pequeno, porém significativo, de itens já utilizados na prova que está servindo como uma referência para a comparação. Esses itens em comum têm suas propriedades psicométricas todas já bem conhecidas – como seu nível de dificuldade e sua capacidade de discriminação –, de tal forma que, conhecendo os parâmetros dessas questões, é possível não só calibrar os itens não-comuns (ou seja, encontrar os seus respectivos parâmetros), como também obter a proficiência de todos os alunos avaliados nos dois testes diferentes. Esse procedimento, chamado de *equalização*<sup>2</sup>, é bastante empregado nos testes trabalhados com a TRI e, por ele, é possível, por exemplo, comparar os resultados do SAEB de diferentes edições – como as de 2009 e 2011 –, bem como comparar os testes do SAEB com os das diversas avaliações implementadas pelos estados brasileiros, como em São Paulo, Minas Gerais, Goiás, Pernambuco, etc., desde que esses testes apresentem entre si itens em comum, para serem utilizados na equalização.

<sup>2</sup> Também se costuma utilizar a expressão original em inglês para esse termo técnico, que é *equating*.

### 3.4. SEGUNDO PERÍODO (1997–2007): EXPANSÃO

O segundo período vivido pelos sistemas de avaliação educacional no Brasil tem seu início associado a um marco de grande relevância no âmbito dos esforços avaliativos no país: o surgimento da Prova Brasil, também realizada pelo MEC, e que compartilha com o SAEB uma série de características relevantes, embora em relação a ele também apresente algumas diferenças não menos importantes.

Dentre as características que a Prova Brasil tem em comum com o SAEB, destacam-se a periodicidade de dois anos, a avaliação de Língua Portuguesa e Matemática, o uso de escalas de proficiência do SAEB, o emprego de itens ou questões de múltipla escolha e a adoção da TRI nos procedimentos de correção e análise dos resultados.

Entretanto, diferentemente do SAEB, a Prova Brasil não inclui o 3º ano do ensino médio, mas somente o 5º e 9º anos (4ª e 8ª séries) do ensino fundamental, possui um desenho censitário (e não amostral como é o caso do SAEB) e somente se aplica às escolas públicas. Devido ao seu caráter censitário, a Prova Brasil produz resultados também no nível das escolas, que são utilizados não somente para propósitos de diagnóstico e planejamento, mas também para o monitoramento do desempenho individual de todas as escolas públicas brasileiras de nível fundamental.

Um segundo aspecto dessa fase vivida pela avaliação educacional em grande escala no Brasil é a expansão que se verificou, especialmente a partir dos últimos anos da década de 1990, dos sistemas estaduais de avaliação, impulsionada pelo desejo das diversas secretarias de educação de conceberem e implementarem seus sistemas de avaliação mais adequados às diversas realidades estaduais. As secretarias estaduais tinham, com isso, o interesse de realizar avaliações mais detalhadas, bem como de fazer com que esse processo se aproximasse mais das características e necessidades específicas de suas respectivas redes de ensino.

Nesse sentido, observa-se, por exemplo, o fato de que as avaliações estaduais tendem a se realizar anualmente, diferentemente do SAEB e da Prova Brasil que, conforme se viu, são aplicadas a cada dois anos. Como os governos estaduais são, a princípio, trocados a cada quatro anos em função das eleições para governador, a periodicidade de dois anos das avaliações federais, acrescida do fato de que há um tempo extra, que chega a cerca de um ano, para o devido processamento e análise dos dados dessas avaliações, faz com que as suas respectivas informações cheguem muito tarde para que as autoridades estaduais delas se

apropriem adequadamente e utilizem-nas para fins de aperfeiçoamento de seus respectivos sistemas. Portanto, a periodicidade anual, aliada ao processamento e análise muito mais rápidos que se conseguem obter com as avaliações feitas pelos próprios estados, faz com que os resultados tenham condições de fornecer subsídios para intervenções que podem ser concebidas, executadas e monitoradas várias vezes por uma mesma administração governamental ao longo de seu respectivo mandato.

Outro ponto que favorece a adoção dos sistemas próprios de avaliação dos estados decorre do fato de que, até recentemente, as escolas rurais não eram devidamente contempladas nas avaliações federais, seja porque eram tratadas apenas do ponto de vista amostral (como é o caso do SAEB), o que impossibilitava o monitoramento de todas elas, seja porque eram simples e totalmente excluídas das avaliações (como foi o caso da Prova Brasil em várias das suas primeiras edições). Portanto, os sistemas estaduais vieram a preencher essa lacuna, sendo que esse aspecto salienta também duas outras importantes características suas: a tendência em fazer avaliações censitárias, ou seja, que incluem todos os alunos e escolas das redes estaduais, e o propósito, encontrado em vários estados, de utilizar os resultados desses testes para criar mecanismos de monitoramento do desempenho educacional nas mais diversas instâncias da educação do estado, chegando mesmo ao nível da escola.

Via de regra, esses novos sistemas de avaliação foram primeiramente implementados nos estados economicamente mais desenvolvidos e densamente povoados do país, como alguns da Região Sudeste: São Paulo (que, na verdade, iniciara seu sistema próprio de avaliação um pouco antes disso, em 1996); Minas Gerais (desde 2000) e o Rio de Janeiro em 2004. Alguns estados de outras regiões também se situam entre os primeiros a aderir a essa tendência, como o Rio Grande do Sul, Pernambuco e o Ceará.

Do ponto de vista das características e da metodologia das avaliações estaduais, um ponto que, em geral, se observa nesses sistemas é a semelhança com as avaliações nacionais do SAEB e da Prova Brasil. Dessa forma, a grande maioria deles privilegia a avaliação das últimas séries dos principais ciclos do ensino básico, concentra-se em testes de Língua Portuguesa e Matemática, emprega itens de múltipla escolha e utiliza a metodologia da Teoria da Resposta ao Item no tratamento estatístico das questões e das provas. Além disso, outra característica extremamente relevante dessas avaliações é o fato de que elas tiveram – e têm tido até os dias de hoje – a preocupação de tornar seus dados nacionalmente comparáveis, o que se tem conseguido por meio da adoção das mesmas escalas de proficiência do SAEB. Na

prática, isso se consegue mediante a adoção de itens comuns com as avaliações nacionais do MEC, que permitem a equalização de resultados entre os diversos sistemas, conforme já comentado anteriormente.

Nesse mesmo sentido, um caso à parte é o estado de São Paulo, justamente o mais economicamente desenvolvido do país e também o pioneiro nas avaliações estaduais, que, em várias de suas edições iniciais, utilizava ainda escalas próprias de proficiência, que não tinham comparabilidade com as escalas nacionais do SAEB. Essa situação perduraria até 1997, ano em que o SARESP – Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – resolveu aderir à métrica nacional, passando, daí em diante, também a equalizar seus resultados com os do SAEB (SÃO PAULO, 2010). Tal fato marcou, tanto do ponto de vista simbólico quanto prático, uma consolidação das avaliações nacionais do SAEB e da Prova Brasil como referências incontornáveis na avaliação educacional em grande escala no país, situação esta que vem se verificando até os dias de hoje.

### 3.5. TERCEIRO PERÍODO (2007–PRESENTE)

O terceiro período vivido pelas avaliações educacionais em grande escala no Brasil, segundo a convenção esquemática aqui considerada, tem seu início no ano de 2007, com a determinação, pelo Ministério da Educação, das metas do Índice de Desenvolvimento da Educação Básica, o Ideb, um instrumento nacional de mensuração e monitoramento do desempenho do ensino básico no país. Outras duas importantes características se salientam nesse período, ambas significativamente associadas ao surgimento do Ideb: a adoção, cada vez mais difundida, de medidas de prestação de contas das escolas em diversos estados da federação e a disseminação, praticamente universal, dos sistemas estaduais de avaliação educacional. Passaremos agora a discorrer sobre cada uma dessas três características.

### 3.5.1. O Ideb

A partir do ano de 2005, o Ministério da Educação, através do INEP, começou a divulgar e utilizar um índice para aferir a situação geral do ensino básico no Brasil, medida esta que recebeu o nome de Ideb: Índice de Desenvolvimento da Educação Básica. O Ideb, para um ano específico e uma determinada unidade educacional (como uma escola, município, estado, região, ou mesmo o país como um todo) é calculado de acordo com a seguinte fórmula:

$$\text{Ideb} = N \times P$$

onde  $N$  é uma medida de desempenho combinado em Língua Portuguesa e Matemática na Prova Brasil ou SAEB, e  $P$  é uma taxa combinada de aprovação para todas as séries ou anos que compõem um ciclo escolar (por exemplo, as séries de 1 a 3 para o ensino médio).

Um aspecto relevante do Ideb é que, para aumentá-lo, é necessário que a unidade educacional em questão promova o aumento de um dos fatores do produto FLUXO X DESEMPENHO, sem fazer com que o outro fator diminua na mesma proporção. O ideal, nessa mesma linha, seria, na verdade, promover um aumento simultâneo de ambos os fatores, o que, via de regra, é uma tarefa difícil. Isto porque o aumento do desempenho escolar comumente se associa a um maior rigor nos critérios de avaliação interna das escolas, o que, por sua vez, resultaria, em tese, numa diminuição da aprovação ou fluxo dos alunos através das séries. E o contrário também tende a se verificar, ou seja, um relaxamento do nível de exigência para a aprovação acarretaria, a princípio, um aumento do fluxo, porém em prejuízo da aprendizagem e, conseqüentemente, do desempenho nas provas.

Outro ponto importante a respeito do Ideb é o fato de que o Ministério da Educação passou a projetar, para cada unidade educacional de interesse, como escolas e estados brasileiros, o valor desse índice que cada unidade deverá apresentar para fazer com que, no futuro, o Brasil, como um todo, experimente um vigoroso aumento da qualidade da sua educação básica. Para que o país atinja essa melhoria, tomou-se como referência o ano de 2022, que tem uma considerável importância simbólica para o país, visto que nele se comemorará o bicentenário da independência do Brasil<sup>3</sup>. Nesse ano, o MEC estabeleceu como

---

<sup>3</sup> Na verdade, o ano de referência para que se atinja um Ideb médio nacional de 6 pontos é 2021, visto que o SAEB e a Prova Brasil, necessários para o cálculo do Ideb, somente têm sido aplicados em anos ímpares. Entretanto, como o índice só é calculado de dois em dois anos, os valores de 2021 é que serão divulgados em 2022.

meta que o Brasil deverá atingir um Ideb médio igual a 6, numa escala que varia de 0 a 10. Segundo estudos comparativos internacionais, estima-se que esse Ideb de 6 corresponderia, *grossa modo*, a um “Ideb” médio que teriam os países da Organização para a Cooperação e o Desenvolvimento Econômicos – OCDE –, um organismo internacional que congrega as economias mais avançadas do mundo, como os Estados Unidos, Austrália, Japão e os países da Europa Ocidental, entre outros. Considerando, portanto, as taxas atuais médias de fluxo de alunos no ensino básico nesses países, bem como o desempenho de seus estudantes nos testes do Programa Internacional de Avaliação de Alunos – PISA –, realizado também pela OCDE, estima-se que um valor de 6 seria a média dos países dessa organização.

O estabelecimento das metas do Ideb sinalizou, portanto, para o Brasil como um todo, o surgimento de um esforço nacional em prol da prestação de contas das unidades educacionais. A partir de então, as escolas passariam, a cada edição da Prova Brasil, a ter seus respectivos valores de Ideb divulgados e comparados com as metas que se lhes haviam traçado. Entretanto, o cumprimento ou não das metas do Ideb por parte das escolas ainda não resultou, na prática, em consequências muito drásticas para as próprias escolas ou para os profissionais nelas envolvidos, a não ser, naturalmente, a exposição pública dos valores reais do Ideb com suas respectivas metas, e as consequentes implicações que isso pode vir a ter em termos de prestígio ou propaganda das escolas junto às suas respectivas comunidades. Entretanto, consequências mais altas, como o impacto que o desempenho escolar possa vir a exercer sobre a remuneração dos professores, é algo que ainda não foi adotado pelos idealizadores do Ideb, até porque esse é um tema profundamente controverso, e de equacionamento político extremamente complicado.

As críticas à prestação de contas com implicações mais sérias para as escolas e para os profissionais da educação provêm de vários lados. Uma das vozes mais eloquentes nesse sentido costuma ser a dos sindicatos dos professores, que criticam essa possibilidade com o argumento de que os docentes e os diretores não podem ser responsabilizados sozinhos pelo desempenho dos seus alunos, aos quais faltam diversas condições necessárias para o seu desempenho acadêmico, que vão desde a falta de apoio familiar à precariedade das condições de ensino nas escolas. Outra crítica, de natureza técnica, costuma vir de alguns estudiosos da eficácia escolar, para os quais há diversos problemas ainda não muito bem resolvidos por aqueles que vêm tentando avaliar o desenvolvimento da educação. Um exemplo dessas questões problemáticas é o fato de que, muitas vezes, os índices propostos para aferir a qualidade educacional, como o próprio Ideb, não levam em conta, por exemplo, a realidade

socioeconômica dos alunos, fazendo com que tanto escolas de clientela mais rica quanto as de clientela mais pobre sejam comparadas como se fossem iguais, o que não seria justo. Outra crítica de natureza técnica é a observação de que essas medidas, sempre ou quase sempre, baseiam-se em dados transversais, ou seja, são produzidas periodicamente, porém, não com base nos mesmos alunos, a fim de averiguar o avanço de seu respectivo aprendizado, mas sim a partir da avaliação de alunos diferentes em anos diferentes, o que compromete, ou mesmo impossibilita, o cálculo do valor agregado da aprendizagem devido à escola. (CARLSON, 2014).

Não obstante todas essas objeções, há também setores relevantes da administração pública que se posicionam a favor de uma maior transparência dos resultados acadêmicos obtidos pelas escolas, argumentando para isso sobre a necessidade de fazer com que as unidades educacionais deixem de ser a “caixa preta” que costumavam ser no passado, onde o que imperava eram as avaliações internas, e onde o público externo raramente sabia do que estava acontecendo no interior dos estabelecimentos de ensino. (ANDRADE, 2008)

### 3.5.2. Políticas de prestação de contas

Essa posição em favor de uma maior abertura dos processos internos escolares para a sociedade tem ganhado força no Brasil nos últimos anos, como se pode observar pelo aumento significativo de casos em que os estados vêm implementando seus próprios sistemas de prestação de contas, além de os atrelar a políticas de compensação para os profissionais da educação. Em Minas Gerais, por exemplo, o estado vem implementando um sistema de pacto de metas, pelo qual os profissionais das escolas podem vir a receber complementos à sua remuneração com base, parcialmente, no desempenho de suas escolas e de suas respectivas superintendências regionais de ensino nos testes do Sistema Mineiro de Avaliação da Educação Básica, o SIMAVE. No Rio de Janeiro, igualmente houve tentativas de atrelar aumentos na remuneração dos profissionais educacionais a uma medida de desenvolvimento escolar que combinava, entre outros elementos, dados de fluxo dos alunos e percentuais de seus alunos cujas proficiências em Língua Portuguesa e Matemática se situavam em níveis mais elevados da escala do Sistema de Avaliação da Educação do Rio de Janeiro, o SAERJ. Outras medidas semelhantes igualmente podem ser observadas em outros estados, como em Pernambuco e no Espírito Santo, entre outros.

A tendência de prestação de contas tem se verificado principalmente nos estados que estiveram entre os primeiros a adotar sistemas próprios de avaliação educacional no Brasil. Tal fato sugere, portanto, que essa tendência pela *accountability* (termo muito utilizado na literatura internacional, e que no Brasil se traduz como *rendição* ou *prestação de contas*, entre outras expressões) se deu no seio de sistemas que já contavam com uma história razoavelmente sedimentada de avaliação externa.

### 3.5.3. Disseminação dos sistemas estaduais de avaliação

Entretanto, paralelamente a isso, outro aspecto tem sido observado no que diz respeito ao terceiro período das avaliações educacionais no Brasil, já comentado no início desta seção: o surgimento de novos sistemas próprios de avaliação, aos quais aderiram estados que antes não contavam entre os adeptos desses procedimentos avaliativos. A figura a seguir apresenta, como um exemplo dessa tendência, os sistemas de avaliação educacional cuja execução ficou, no período compreendido entre os anos de 2009 e 2012, a cargo do Centro de Políticas Públicas e Avaliação da Educação – CAEd – órgão da Faculdade de Educação da Universidade Federal de Juiz de Fora – UFJF (CAED, 2009).



Figura 3.4: Sistemas de avaliação em grande escala com avaliações executadas pelo CAEd entre 2009 e 2012. Fonte: CAEd, 2012.

Por esta figura, percebe-se que a implementação dos sistemas estaduais de avaliação é uma realidade já consolidada em praticamente todo o país.

### 3.6. POSSÍVEIS PROGRESSOS E DESAFIOS DAS AVALIAÇÕES NO CONTEXTO BRASILEIRO

Apesar das críticas sofridas pelos sistemas externos de avaliação, algumas das quais já foram mencionadas neste texto, é inegável o fato de que eles estão dando mostra de ser uma tendência irreversível na atual conjuntura educacional brasileira, e mesmo mundial. Embora tampouco pareça haver dúvidas de que tais sistemas padeçam de problemas que requerem ainda soluções mais adequadas, como uma maior validade dos cálculos de valor agregado das escolas e da adequação das políticas de responsabilização, também é inegável perceber que o simples surgimento dos atuais sistemas de avaliação em grande escala já foi suficiente para produzir algumas mudanças significativas e positivas no quadro da educação nacional. Uma das mais importantes delas é uma maior transparência dos resultados educacionais, que tem permitido, inclusive, dimensionar mais precisamente as grandes lacunas de qualidade que precisam ser devidamente diminuídas ou desfeitas pelos esforços movidos pelas autoridades e profissionais da educação.

Por tudo isso, aqueles que são favoráveis ao movimento de avaliação externa e à prestação de prestação de contas resumem essas vantagens dizendo que os desenvolvimentos recentes que se vêm observando têm sido bem sucedidos em deixar os sistemas educacionais mais visíveis e racionais, bem como em promover um maior senso de participação e responsabilidade entre os profissionais diretamente responsáveis por promover o progresso educacional do Brasil. E essas vantagens, segundo essa visão, mais do que compensam os eventuais problemas desses sistemas, os quais, a propósito, podem – e, com efeito, devem – sempre estar sujeitos a um processo constante de aperfeiçoamento.

E é precisamente para atender a propósitos como esse que se vêm desenvolvendo, tanto os sistemas estatais como os nacionais de avaliação educacional no Brasil, conforme já se comentou. Neste sentido, na continuação do presente texto, analisaremos o desempenho das escolas públicas mineiras do ponto de vista de dois de seus aspectos básicos: o status e o progresso de suas medidas de proficiência, conforme mensuradas por diversas edições do SIMAVE, o teste estadual de Minas Gerais.

## **4. A EFICÁCIA DAS ESCOLAS PÚBLICAS DO SIMAVE: UMA PROPOSTA DE MENSURAÇÃO**

### **4.1. INTRODUÇÃO**

Conforme se disse no capítulo anterior, diversos sistemas de avaliação têm surgido no Brasil nas décadas mais recentes. E, devido a isso, a cada ano que passa, tem surgido uma verdadeira avalanche de dados de mensuração da proficiência das escolas, que muitas vezes aparecem resumidos e interpretados nos relatórios produzidos pelas instituições responsáveis por esses exames.

Entretanto, um ponto que, em geral, chama a atenção nesses resultados, é que eles muitas vezes se apresentam como realidades pontuais, ou seja, como uma espécie de fotografia instantânea do desempenho das escolas. E é precisamente com base nesses “instantâneos” que as escolas costumam ser avaliadas, tanto pelos próprios sistemas educacionais, como também pela imprensa e pela sociedade em geral, que absorve essas informações, e às vezes constroem novas interpretações a partir delas. Isso faz com que seja comum, por exemplo, que uma determinada escola que, digamos, tenha alcançado médias de proficiência extremamente elevadas num determinado ano de testagem, seja vista como merecedora de prêmios e outras distinções a serem concedidas a seus respectivos alunos, professores e gestores.

Contudo, o grande perigo de tudo isso é o fato sobejamente conhecido por parte dos técnicos trabalhando com estatísticas educacionais, que costuma haver um nível considerável de variação não-explicada nos resultados de uma escola, de um ano para outro, conforme já se comentou em partes anteriores deste trabalho. Dessa forma, para aqueles interessados em detectar resultados mais consistentes de desempenho escolar, um caminho praticamente incontornável é a busca por resultados oriundos, não apenas de uma ou poucas avaliações disponíveis, mas sim a partir de observações colhidas ao longo de uma série mais extensa de exames aplicados à rede educacional de interesse.

Nesse sentido, o objetivo do presente capítulo é analisar o desempenho das escolas do SIMAVE, com base em duas dimensões básicas: status e progresso, porém tomando como base uma série mais extensa de resultados, ao invés de somente poucas avaliações. Para este estudo, será considerada em separado a 8ª série ou 9º ano do Ensino Fundamental, em duas disciplinas: português e matemática. Os dados da tabela a seguir informam as quantidades totais das escolas avaliadas, em termos brutos e também em percentuais.

*Tabela 4.1: Dados gerais da população de escolas avaliadas, por rede de ensino (SIMAVE 2000 – 2011)*

	<i>Municipal</i>	<i>Estadual</i>	<i>Total</i>
<i>N</i>	1638	2991	4629
<i>%</i>	35	65	100

Observa-se, portanto, que cerca de dois terços delas são estaduais, e o restante constituído de escolas municipais. Cabe lembrar que as escolas estaduais vêm sendo avaliadas desde o início do SIMAVE, no ano 2000, ao passo que as municipais somente aderiram ao sistema avaliativo em 2006.

#### 4.2. DESEMPENHO GERAL

A Tabela 4.2 apresenta os dados gerais de desempenho, segundo algumas variáveis de interesse. Suas seis primeiras linhas referem-se a dados de progresso ou avanço da proficiência ao longo do tempo. Neste sentido, as duas primeiras linhas apresentam os percentuais totais de vezes em que as médias das escolas tiveram uma variação positiva de uma avaliação para a seguinte. Nota-se, neste caso, que as variações das médias encontram-se pouco acima dos 50% em ambas as disciplinas, com uma ligeira predominância do caso de Matemática, onde as variações positivas chegam a 60,7% das vezes.

*Tabela 4.2: Dados gerais de desempenho das escolas avaliadas, por rede de ensino (SIMAVE 2000 – 2011)*

<i>Desempenho</i>	<i>MUN</i>	<i>EST</i>	<i>TOTAL</i>
% de var. pos. – LP	54,1	57,2	56,4
% de var. pos. – MAT	55,8	62,4	60,7
Inclin. média – LP	1,4	1,5	1,5
Inclin. (dp) – LP	5,3	2,6	3,5
Inclin. média – MAT	3,1	3,2	3,2
Inclin. (dp) – MAT	0,8	1,3	1,2
Média 2006 – LP	240,5	243,1	242,6
DP 2006 – LP	18,7	17,1	17,5
Média 2006 – MAT	242,8	246,7	245,9
DP 2006 – MAT	21,8	20,2	20,6

Outro ponto que se observa nesta tabela é que, ao menos em termos brutos, os dados indicam, numericamente, uma certa superioridade do desempenho nas escolas estaduais, quando comparadas com as municipais. Por eles, observa-se que, um pouco mais da metade

das variações de média de uma avaliação para a seguinte, correspondeu a casos de aumento do desempenho, sendo que, na rede estadual, os números se mostraram ligeiramente maiores que na rede municipal.

Também são informadas as inclinações médias das retas de regressão das escolas individuais, que sintetizam a sua tendência de evolução temporal ao longo das diversas edições do SIMAVE. Observa-se que, para ambas as redes, as médias são positivas tanto para Língua Portuguesa quanto para Matemática, indicando que, em geral, tem havido um ganho na proficiência média de ambas as disciplinas. Entretanto, cabe, aqui, fazer ao menos duas observações. Uma, é que as inclinações médias, principalmente no caso de Língua Portuguesa, parecem ser relativamente modestas, situando-se em torno de um ganho de 1,5 ponto ao ano, na escola de proficiência desta disciplina. Logo, caso esse ritmo de crescimento fosse mantido constante ao longo do tempo, seria necessária uma década para as escolas aumentarem suas respectivas médias de português somente de 15 pontos da escala, o que não parece muito, especialmente quando se observa que muitas delas encontram-se com médias bastante abaixo da média nacional do SAEB.

Um segundo ponto que merece um comentário é a maior consistência de ganhos em matemática do que em português, algo que se depreende do fato de que os desvios-padrão das inclinações temporais das médias tendem a ser, proporcionalmente, bem maiores em Língua Portuguesa do que em Matemática. Em outras palavras, as escolas parecem diferir mais umas das outras quanto ao progresso em português do que em matemática.

Por fim, as quatro últimas linhas da tabela informam os valores totais das médias e dos desvios-padrão das proficiências em ambas as disciplinas.

Também cabe notar que todas essas medidas globais foram calculadas por meio de uma agregação ponderada pelo tamanho das respectivas escolas. E este tamanho, por sua vez, foi calculado a partir da média de alunos do 9º ano EF avaliados na escola ao longo da série histórica de sua respectiva avaliação.

#### 4.3. MEDIDAS DE STATUS

Para o estudo do status ou situação instantânea da proficiência nas disciplinas avaliadas, foram aqui consideradas três variáveis. Duas delas foram os percentuais de examinados que tiveram seus escores situados em um dos dois extremos da escala de proficiência, o inferior e o superior, segundo pontos intervalares de corte determinados por especialistas. No caso da 8ª série/9º ano EF aqui considerado, esses corresponderam

respectivamente a pontos em Matemática abaixo de 225 (caso inferior) ou acima de 300 (superior); em Língua Portuguesa, os pontos de corte foram de 200 e 275 pontos respectivamente. Além desses dois indicadores, que passarão a ser chamados, também respectivamente, de N1 e N3, também foi considerada a média de proficiência das escolas nessas mesmas disciplinas.

Tabela 4.3: Médias totais de N1 (% de alunos no nível inferior), N3 (% de alunos no nível superior) e ME (médias de proficiência) das escolas em Língua Portuguesa e Matemática, por ano de edição do SIMAVE (2000 a 2011) – 8ª série/9º ano EF.

Ano	N1		N3		ME	
	LP	MAT	LP	MAT	LP	MAT
2000	20	46,8	20,7	6,9	238,4	229,7
2002/3	11,9	38,6	19,4	11,7	242,8	245,4
2006	19	33,1	25,9	14,3	242,6	245,9
2007	19,9	31,3	23,1	16	239,9	249,7
2008	12,7	28,7	27,3	18	249,1	253,9
2009	13,4	23,7	29,6	19,2	250	258,3
2010	11,9	11,9	33,1	24,2	253,7	266,2
2011	14,2	14,2	32	19,8	251,4	261,1

Tabela 4.4: Médias dos desvios-padrão de N1 (% de alunos no nível inferior), N3 (% de alunos no nível superior) e ME (médias de proficiência) das escolas em Língua Portuguesa e Matemática, por ano de edição do SIMAVE (2000 a 2011) – 8ª série/9º ano EF.

Ano	N1		N3		ME	
	LP	MAT	LP	MAT	LP	MAT
2000	13,2	17,0	11,2	6,4	16,8	17,7
2002/3	9,6	18,8	10,9	9,6	14,3	22,5
2006	13,6	17,3	13,1	10,7	19,4	21,5
2007	13,9	17,7	13,0	12,4	18,6	22,1
2008	11,2	17,3	14,4	14,0	17,3	22,9
2009	12,3	16,5	15,5	15,0	19,0	22,8
2010	11,1	14,9	16,1	16,4	18,9	22,5
2011	11,9	15,7	16,0	15,6	19,0	22,2

Mas, antes de prosseguirmos, cabe fazermos ao menos três comentários. Um deles diz respeito ao fato de que os seis resultados de status aqui considerados *não* se encontram nas mesmas escalas de medida, entre outras coisas porque as disciplinas de Língua Portuguesa e Matemática possuem escalas diferentes.

Entretanto, quanto às médias de proficiência neste caso, um ponto particularmente favorável à apresentação e análise dos resultados é que ambas essas escalas, especificamente para o 9º ano do Ensino Fundamental aqui considerado, possuíam, ao menos inicialmente na escala nacional do SAEB (de 1997), valores idênticos para suas respectivas médias e desvios-

padrão nas distribuições nacionais de proficiência, e que correspondiam respectivamente a 250 e 50 pontos em cada escala.

Outro ponto a ressaltar é a considerável dificuldade existente em estabelecer uma correspondência exata ou muito precisa entre o que vem a ser um dado nível de desempenho – digamos, o inferior ou o superior – em duas disciplinas diferentes, como português e matemática. Naturalmente, os critérios adotados na determinação desses níveis por especialistas das diferentes áreas ao menos a princípio podem, por variações quanto ao rigor e à precisão, fazer com que haja uma inflação ou deflação dos percentuais de proficiência correspondentes nas diferentes disciplinas avaliadas.

Mas, um terceiro ponto que também merece ser observado é que, apesar das duas ressalvas feitas nas observações anteriores, é possível notar uma associação relativamente robusta entre todas as seis medidas. E, na verdade, é de certo modo previsível que tal aconteça, visto que todas essas variáveis dependem apenas de duas originais: os escores dos alunos em português e matemática, os quais, por sua vez, costumam também se achar forte e positivamente associados entre si.

Dessa forma, observa-se que todas as correlações de Pearson – apresentadas na Tabela 4.5, a seguir – apresentam valores com módulos variando entre as casas de 0,7 e 0,9, mesmo entre indicadores mensurados para disciplinas diferentes. Os valores negativos nas correlações entre N1 (o percentual de alunos no nível inferior de proficiência) e as demais medidas de desempenho deve-se, naturalmente, ao fato de que, à medida que decresce o número de alunos com baixa proficiência numa disciplina, esta tende a aumentar sua respectiva média e também o valor de N3 (o percentual de alunos com proficiência na parte superior da escala).

*Tabela 4.5: Valores das correlações de Pearson entre as médias totais de ME (médias de proficiência), N3 (% de alunos no nível superior) e N1 (% de alunos no nível inferior) das escolas em Língua Portuguesa e Matemática, para toda a série histórica do SIMAVE (2000 a 2011) – 8ª série/9º ano EF.*

		ME_LP	N3_LP	N1_LP	ME_MAT	N3_MAT	N1_MAT
ME_LP	r	1	0,914	-0,917	0,869	0,772	-0,834
	Sig. (bic.)		0	0	0	0	0
N3_LP	r		1	-0,733	0,823	0,799	-0,750
	Sig. (bic.)			0	0	0	0
N1_LP	r			1	-0,779	-0,620	0,793
	Sig. (bic.)				0	0	0
ME_MAT	r				1	0,892	-0,932
	Sig. (bic.)					0	0
N3_MAT	r					1	-0,773
	Sig. (bic.)						0
N1_MAT	r						1
	Sig. (bic.)						

*Obs.: Em todos os casos a correlação é significativa no nível de 0,001 (bicaudal)*

Uma justificativa para a escolha dessas três medidas de status para cada disciplina – N1, média e N3 – deve-se ao fato de que, ao mesmo tempo em que elas se encontram fortemente correlacionadas, o que as aproxima e, portanto, aumenta a fidedignidade de um possível escore global que incorpore a todas, por outro lado, cada uma delas se detém em uma “subdimensão” específica da escala de proficiência: N1 avalia as escolas no tocante às faixas inferiores de desempenho dos alunos; por sua vez, a média, como o seu próprio nome diz, observa especialmente o comportamento das escolas nas faixas intermediárias de desempenho e, finalmente, N3, por medir o percentual de alunos situados no nível mais elevado de proficiência, permite, dessa forma, que se tirem algumas conclusões mais precisas acerca do estado do aprendizado nas regiões superiores da escala de proficiência.

Uma vez feita essa delimitação de quais variáveis observar, passemos agora a analisar os resultados globais de cada uma delas.

#### 4.3.1. O status medido por N1

O Gráfico 4.1, a seguir, apresenta a evolução temporal dos valores globais de N1 simultaneamente para ambas as disciplinas, ao longo de toda a série histórica do SIMAVE. Por ele, ficam evidentes algumas conclusões notáveis, a saber:

Primeiro, que, apesar das restrições quanto à comparabilidade de desempenho entre as duas diferentes disciplinas há pouco mencionadas, ainda assim é possível observar que, de modo geral, os alunos avaliados tendem a se sair melhor em Língua Portuguesa do que em Matemática, visto que esta última, em média, apresenta valores de N1 consideravelmente maiores do que a primeira. Talvez o principal fator explicativo dessa característica seja o fato, sobejamente comentado pela literatura sobre eficácia educacional, de que a linguagem e a leitura em grande parte podem ser aprendidas em outros ambientes e meios sociais que não a própria escola. Este é caso, por exemplo, da aprendizagem ocorrendo na família, ou pelos diversos tipos de mídia, ou na vivência profissional, etc. Diante disso, os resultados dos testes de linguagem, em comparação com os resultados de matemática, podem levar a uma superestimação da eficácia das escolas consideradas, no tocante à primeira disciplina. Já a Matemática, por ter seu domínio formal bem mais dependente do contexto escolar, tende a ressentir-se mais diretamente das deficiências de ensino por este apresentadas.

Outra característica que se observa no gráfico é o fato de que, ao menos no que diz respeito ao comportamento das escolas nos níveis mais baixos de proficiência, o desempenho do sistema mineiro de educação tem sido razoavelmente bem-sucedido. Tal fato se observa no

sentido em que, para ambas as disciplinas, ocorre uma tendência de redução dos valores de N1 ao longo do tempo. No caso de Língua Portuguesa, vê-se que, nas quatro primeiras avaliações, havia cerca de um quinto dos alunos no nível inferior de proficiência. Por outro lado, nas quatro avaliações mais recentes dessa mesma série histórica, esses percentuais reduziram-se para menos de 15%.

Já o progresso em Matemática quanto a esse indicador é consideravelmente mais notável. E isso, em parte, se deve aos valores iniciais bastante elevados desses índices no início de sua série histórica – correspondentes a quase a metade dos alunos no nível inferior de proficiência no primeiro ano da série! Dessa forma, o indicador N1 de matemática, desde o início, possuía mais “teto” para crescer, e isso foi precisamente o que ele fez.

Um ponto de interesse na evolução conjunta de português e matemática é o fato de que, nas duas mais recentes avaliações consideradas, os valores de N1 em ambas as disciplinas coincidentemente foram iguais, indicando, assim, que após pouco mais de uma década de mensuração, as escolas mineiras conseguiram simultaneamente reduzir os percentuais de alunos nos níveis mais baixos de proficiência e também aproximadamente igualar português e matemática quanto a esse mesmo indicador.

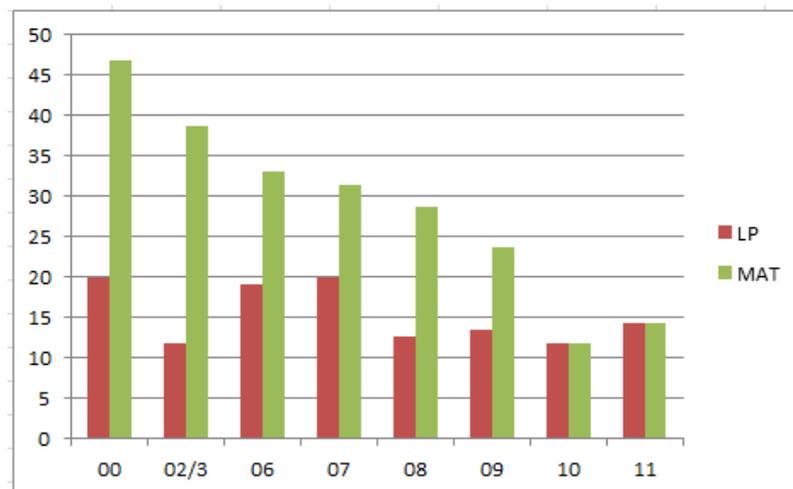


Gráfico 4.1: Evolução temporal das médias totais de N1 (% de alunos no nível inferior de proficiência) das escolas em Língua Portuguesa e Matemática, por ano de edição do SIMAVE ([20]00 a [20]11) – 8ª série/9º ano EF.

#### 4.3.2. O status medido pelas médias de proficiência

Dada a já comentada e significativa correlação negativa entre N1 e a média de proficiência, pode-se ver, no Gráfico 4.2, a seguir, que o decréscimo de alunos nos níveis inferiores de proficiência – visto anteriormente – de certa forma se reflete na elevação das médias de proficiência observadas na ilustração a seguir.

Novamente, é possível aqui observar que as médias ascendentes, principalmente no caso de Língua Portuguesa, obedeceram a um padrão aproximado de dois “degraus”: um inferior, correspondente às quatro primeiras avaliações, e outro superior, vinculado às quatro últimas.

Em ambas as disciplinas, nas quatro primeiras edições do SIMAVE, as médias ficaram abaixo das respectivas médias nacionais do SAEB em 1997. Já nas quatro últimas, elas ficaram aproximadamente iguais ou superiores a essas mesmas médias. E, ao menos em termos numéricos, o progresso em Matemática tem sido, comparativamente um pouco mais alto do que em Língua Portuguesa.

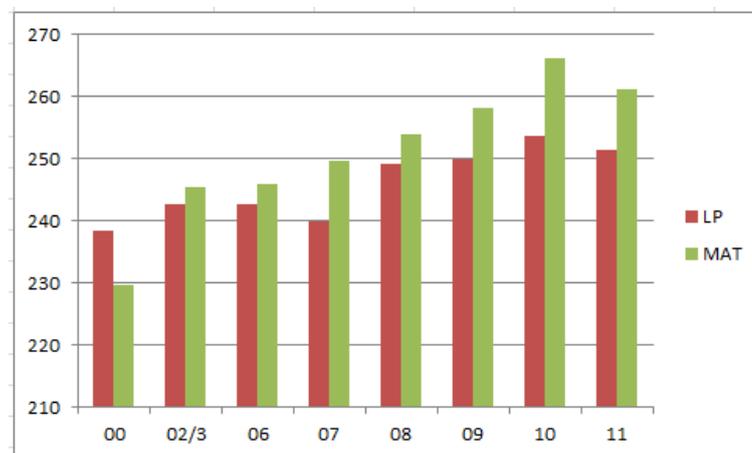


Gráfico 4.2: Evolução temporal das médias totais de proficiência das escolas em Língua Portuguesa e Matemática, por ano de edição do SIMAVE ([20]00 a [20]11) – 8ª série/9º ano EF.

#### 4.3.3. O status medido por N3

Por fim, o Gráfico 4.3, a seguir, apresenta a evolução temporal das medidas de N3. Por ele, percebe-se que os percentuais de alunos com proficiência situada nos níveis superiores das escalas de português e matemática vêm aumentando paulatinamente com o tempo.

No caso de Língua Portuguesa, inicialmente nesta série, cerca de um quinto dos alunos avaliados encontrava-se na faixa superior de proficiência e, nos anos mais recentes, essa proporção subiu para cerca de um terço deles. Em Matemática, também se percebe um progresso, embora menos acentuado, e apesar do “teto” para N3 subir nesta disciplina ser também maior do que em português. E, de fato, observa-se que os percentuais de alunos no nível mais elevado de proficiência em matemática inicialmente nesta série eram consideravelmente baixos – em torno de 10% ou menos – mas, em anos recentes, têm se aproximado de 20%.

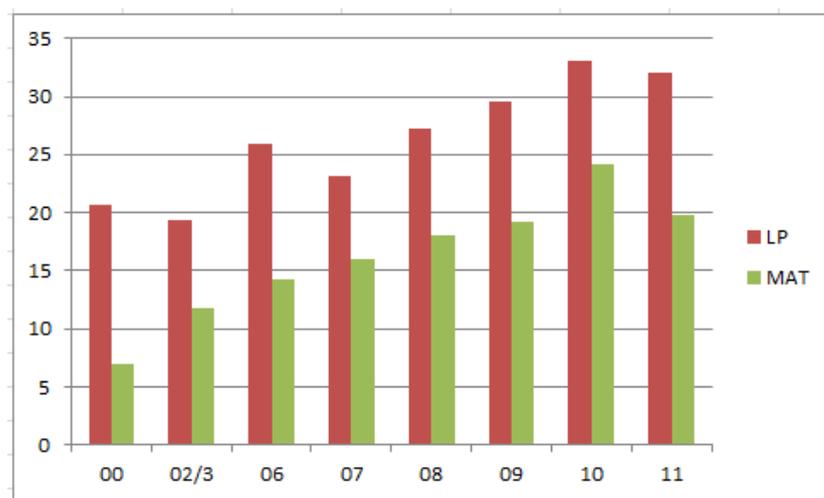


Gráfico 4.3: Evolução temporal das médias totais de N3 (% de alunos no nível superior de proficiência) das escolas em Língua Portuguesa e Matemática, por ano de edição do SIMAVE ([20]00 a [20]11) – 8ª série/9º ano EF.

#### 4.4. MEDIDAS DE PROGRESSO

Outra dimensão de relevo para este estudo da eficácia das escolas do SIMAVE é a evolução temporal dos seus respectivos indicadores de desempenho. Estes, por sua vez, serão os mesmos já considerados na mensuração do status: N1, média e N3. E, da mesma forma que no caso anterior, eles também permitem observar o comportamento das escolas nas regiões inferiores, médias e superiores de ambas as escalas de proficiência, só que, agora, não mais do ponto de vista estático – como no caso do status – mas sim sob um aspecto dinâmico, ou seja, através das taxas de variação dessas mesmas medidas, calculadas entre duas avaliações consecutivas.

As duas tabelas seguintes apresentam uma síntese dessas medidas de evolução, detalhadas para cada par de avaliações consecutivas. Cabe apenas observar que, conforme já se comentou anteriormente neste texto, as avaliações em português e matemática, nos primeiros anos do SIMAVE, não ocorreram de forma perfeitamente uniforme, nem paralela. Desse modo, a disciplina de Língua Portuguesa, após ter sido avaliada inicialmente em 2000, experimentou um salto de dois anos, ao ser avaliada pela segunda vez em 2002, e novamente um salto de quatro anos, quando foi novamente testada em 2006 e, desde então, vem sendo incluída anualmente em todas as avaliações do SIMAVE. Por sua vez, a disciplina de Matemática, depois de ter também sido avaliada inicialmente no ano 2000, teve sua segunda avaliação em 2003 e, a partir de 2006, quando foi avaliada pela terceira vez, passou a apresentar um calendário de avaliação idêntico ao de Língua Portuguesa. Também cabe observar que, a partir de 2006, passaram a ser também avaliadas pelo SIMAVE as redes municipais de ensino, ao contrário de antes, quando só foi avaliada a rede estadual.

As flutuações de aplicações das diferentes disciplinas ocorridas na série inicial do SIMAVE forçaram a adoção de algumas convenções para facilitar a interpretação das medidas de variação de desempenho. Dessa forma, por exemplo, as variações das medidas de Língua Portuguesa entre 2000 e 2002 informadas nas tabelas e gráficos a seguir, na verdade, correspondem à *metade* das variações brutas observadas entre esses mesmos anos. O raciocínio por trás disto é que, uma vez admitida uma tendência linear de variação dessas medidas entre 2000 e 2002, então a taxa *anual* de variação pôde ser inferida a partir da respectiva taxa *bianual*, simplesmente dividindo esta última por dois, e um raciocínio análogo se aplicou a todos os demais casos, inclusive aos de Matemática.

Tabela 4.6: Médias das variações de N1 (% de alunos no nível inferior), N3 (% de alunos no nível superior) e ME (médias de proficiência) das escolas em Língua Portuguesa e Matemática, por ano de edição do SIMAVE (2000 a 2011) – 8ª série/9º ano EF.

Ano	N1		N3		ME	
	LP	MAT	LP	MAT	LP	MAT
2000 - 02/3	-4,1	-2,7	-0,5	1,6	2,4	5,1
2002/3 - 06	1,7	-1,9	1,6	1,0	0,0	0,6
2006 - 07	0,8	-1,8	-2,7	1,8	-2,6	4,0
2007-08	-7,1	-2,5	4,2	1,9	9,1	4,1
2008 - 09	0,4	-5,4	2,5	1,4	1,3	5,0
2009 - 10	-1,4	-4,7	3,5	5,0	3,7	7,7
2010 - 11	2,3	2,2	-1,1	-4,4	-2,3	-5,1

Tabela 4.7: Desvios-padrão das variações de N1 (% de alunos no nível inferior), N3 (% de alunos no nível superior) e ME (médias de proficiência) das escolas em Língua Portuguesa e Matemática, por ano de edição do SIMAVE (2000 a 2011) – 8ª série/9º ano EF.

Ano	N1		N3		ME	
	LP	MAT	LP	MAT	LP	MAT
2000 - 02/3	11,3	17,0	8,6	7,8	12,9	18,0
2002/3 - 06	10,8	16,1	9,7	8,4	13,8	17,7
2006 - 07	10,1	12,6	9,7	8,3	12,9	14,9
2007-08	9,4	12,0	10,2	9,5	12,1	14,6
2008 - 09	8,1	11,5	10,5	10,1	11,6	14,4
2009 - 10	8,4	10,5	11,1	10,5	12,2	13,7
2010 - 11	9,5	10,9	11,7	10,9	13,5	14,3

Já a Tabela 4.8, a seguir, apresenta as medidas globais de correlação entre as taxas de variação ano a ano de cada uma das seis medidas de desempenho aqui consideradas. E novamente se percebem, tanto em relação a uma mesma disciplina quanto entre as duas diferentes, valores significativos de coeficientes R de Pearson, embora, em geral, não tão acentuados quanto nos casos de status anteriormente apresentados.

Tabela 4.8: Valores das correlações de Pearson entre as variações de ME (médias de proficiência), de N3 (% de alunos no nível superior) e de N1 (% de alunos no nível inferior) das escolas em Língua Portuguesa e Matemática, para toda a série histórica do SIMAVE (2000 a 2011) – 8ª série/9º ano EF.

		ME_LP	N3_LP	N1_LP	ME_MAT	N3_MAT	N1_MAT
ME_LP	r	1	0,779	-0,841	0,524	0,422	-0,483
	Sig. (bic.)		0	0	0	0	0
N3_LP	r		1	-0,460	0,483	0,523	-0,445
	Sig. (bic.)			0	0	0	0
N1_LP	r			1	-0,407902607628286	-0,217	0,391
	Sig. (bic.)				0	0	0
ME_MAT	r				1	0,465	-0,347
	Sig. (bic.)					0	0
N3_MAT	r					1	-0,505
	Sig. (bic.)						0
N1_MAT	r						1
	Sig. (bic.)						

Obs.: Em todos os casos a correlação é significativa no nível de 0,001 (bicaudal)

#### 4.4.1. O progresso medido pelas variações de N1

Já se comentou anteriormente acerca da considerável diminuição dos percentuais de alunos nos menores níveis de proficiência, ao longo da série histórica do SIMAVE. E, de fato, percebe-se, no gráfico a seguir, que as variações ocorridas desses indicadores ao longo de avaliações sucessivas tendem, previsivelmente, a ser negativas. Também se observa que, nas sete variações anuais possíveis, Matemática apresentou seis delas com sinal negativo, o que corrobora a existência de, pelo menos, um certo grau de avanço da proficiência dessa disciplina ao longo da série histórica. Já em Língua Portuguesa, os decréscimos de N1 ocorreram quatro vezes em sete. Mas, de todo modo, nesta última disciplina, as variações negativas foram, em módulo, razoavelmente mais acentuadas dos que as positivas, resultando, assim, numa diminuição geral dos percentuais de alunos situados nos níveis inferiores de desempenho.



Gráfico 4.4: Evolução temporal das variações das médias de N1 (% de alunos no nível inferior de proficiência) das escolas em Língua Portuguesa e Matemática, por ano de edição do SIMAVE ([20]00 a [20]11) – 8ª série/9º ano EF.

#### 4.4.2. O progresso medido pelas variações das médias de desempenho

O gráfico, a seguir, apresenta as variações das médias de desempenho nas duas disciplinas, e como elas vêm evoluindo ao longo dos últimos anos. Da mesma forma que com os demais indicadores, observa-se, em geral, uma correlação positiva entre elas, pelas quais os sinais observados nos saltos anuais de resultados de uma disciplina tendem a acompanhar os de outra. E, também da mesma forma que antes, o gráfico possibilita observar a tendência de aumento do desempenho geral ao longo do tempo, devido ao fato de que a maior parte das variações observadas são positivas.



Gráfico 4.5: Evolução temporal das variações das médias de proficiência das escolas em Língua Portuguesa e Matemática, por ano de edição do SIMAVE ([20]00 a [20]11) – 8ª série/9º ano EF.

#### 4.4.3. O progresso medido pelas variações de N3

Por fim, o Gráfico 4.6, a seguir, representa a evolução das taxas de variação de N3, que, conforme se pode observar, vem sendo predominantemente positivas ao longo da série histórica. Entretanto, diferentemente do caso oposto – o do decréscimo de N1 – em que o desempenho de Matemática foi ainda superior ao de Língua Portuguesa, neste novo caso – o do aumento de N3 – os resultados de Língua Portuguesa são ainda superiores aos de Matemática. Isso indica, uma vez mais, uma certa defasagem que o aprendizado em Matemática apresenta em relação a Língua Portuguesa, devido ao fato desta última poder crescer razoavelmente bem devido a outros fatores que não a escola.



Gráfico 4.6: Evolução temporal das variações das médias de N3 (% de alunos no nível superior de proficiência) das escolas em Língua Portuguesa e Matemática, por ano de edição do SIMAVE ([20]00 a [20]11) – 8ª série/9º ano EF.

#### 4.5. UM ÍNDICE DE DESEMPENHO PARA AS ESCOLAS MINEIRAS

Uma vez definidas as seis variáveis de interesse, o passo seguinte foi conceber um ou mais indicadores capazes de mensurar o desempenho das escolas segundo os critérios de status e progresso observados. Isto então foi feito por meio de três etapas: primeiro, a determinação de uma medida única de status; segundo, o cálculo de uma taxa de progresso e, por último, a obtenção de uma combinação de ambas as anteriores, que as sintetizasse. Passemos, então, a detalhar cada uma delas.

##### 4.5.1. A busca por um indicador de status

Uma primeira característica desejável de um indicador de status veio a ser a sua capacidade de sintetizar, em uma única medida, o desempenho da escola nas diferentes regiões do espectro de proficiência consideradas. Portanto, isso levou à proposta de integrar, em uma única medida, os valores de N1, média e N3 para cada escola avaliada. Com isso, consegue-se não somente a vantagem de se obter um indicador mais “completo”, capaz de varrer uma área bem mais ampla do espectro de proficiências, como também aumenta-se a fidedignidade do indicador, ao torna-lo um resultado obtido a partir da combinação de vários outros fortemente associados entre si.

O indicador único de status, por sua vez, foi concebido e calculado a partir de dois indicadores de status calculados separadamente, um para Língua Portuguesa e outro para Matemática. Na obtenção destas duas últimas medidas, um problema a ser superado decorre

do fato de que as medidas de status consideradas – N1, média e N3 – conforme já se comentou, não se encontram todas em uma mesma escala. Isto, portanto, torna impossíveis algumas operações potencialmente necessárias para o cálculo de um indicador unificador de todas essas medidas, como procedimentos de soma de parcelas ou de agregação de médias.

A necessidade de unificação das métricas dos três indicadores de status em cada disciplina foi então satisfeita por meio da adoção, não dos valores brutos dessas medidas, mas sim de distâncias calculadas entre cada uma delas e um valor considerado ideal.

Sabe-se que uma característica altamente desejável de um indicador em geral, e de um educacional em particular, é levar em conta o fato de que ele precisa ser observado e analisado por um grande número de profissionais, como os atuantes numa determinada rede de ensino – e que, via de regra, possuem poucos conhecimentos técnicos de estatística ou psicometria. Dessa forma, é altamente desejável que tais indicadores sejam de fácil apresentação e interpretação. E, caso apresente esta qualidade, o indicador tem, ao menos a princípio, uma maior capacidade de bloquear possíveis resistências contra medidas que, aos olhos daqueles que por elas têm que se pautar, de que estão lidando com uma “caixa preta”, sobre a qual praticamente ninguém tem controle ou compreensão.

Portanto, ao se conceberem indicadores, é inevitável lidar com o desafio de criar medidas que sejam, ao mesmo tempo, tecnicamente adequadas e de fácil compreensão. Por outro lado, um grande problema prático é que essas duas últimas características tendem a variar inversamente uma em relação à outra...

Com tudo isso em mente, um procedimento que se propõe nas medidas a seguir comentadas é a adoção, para os novos indicadores, de uma métrica que teoricamente se estende de 0 a 10, sendo que, quanto maior o valor do indicador, maior a sua proximidade em relação ao ideal desejado. Ou, dito de outra forma, uma escola que obtém um valor 10 em um indicador, atingiu ou superou a sua respectiva meta, ao passo que uma escola que tem um zero nesse mesmo indicador, encontra-se demasiadamente distante de alcançá-la.

Naturalmente, um ponto fundamental na determinação de um indicador assim concebido é definir o que se convencionou chamar de um ponto “ideal” ou “máximo desejável” e, por outro lado, também um ponto “mínimo” de desempenho, segundo a escala original do próprio indicador. A tabela a seguir apresenta os respectivos valores desses pontos críticos para cada uma das três medidas de status, tanto em português quanto em matemática, que foram utilizados no presente caso:

Tabela 4.9: Os limites de desempenho considerados na elaboração dos índices de status, para ambas as disciplinas.

	<i>Limite de desempenho</i>	
	<i>Inferior (pior cenário possível)</i>	<i>Superior (cenário minimamente desejável)</i>
<i>N1</i>	100	10
<i>Média</i>	100	275
<i>N3</i>	0	40

Para N1, o pior cenário possível ocorreria se todos os alunos (100%) de uma dada escola nele estivessem situados quanto à proficiência, ou em português, ou em matemática. Por outro lado, convencionalmente adotou-se como 10% o máximo valor que configuraria, nessa escola, como um cenário minimamente desejável, o que implicaria que ao menos nove em cada dez alunos da escola estariam com sua respectiva proficiência situada, ou no nível intermediário, ou no avançado.

Para a média de proficiência, adotou-se o valor de 100 pontos da escala como o mínimo possível, o que representa uma distância de três desvios-padrão abaixo da média nacional de proficiência do SAEB/1997 em cada uma das disciplinas consideradas. Trata-se de uma medida extremamente baixa para uma média escolar nas escalas de proficiência do 9º ano EF, sendo que há, ou tem havido, pouquíssimos casos de médias escolares inferiores a esse valor, tanto no SIMAVE quanto em outras avaliações, como na Prova Brasil. Por outro lado, uma média escolar de 275 pontos corresponde a 25 pontos ou meio-desvio padrão nacional (segundo o SAEB/1997) acima da média do país. E também corresponde ao ponto que, ao menos em Língua Portuguesa, separa os limites intermediário e avançado de proficiência, conforme já apresentados neste trabalho.

Por último, no caso de N3, o pior cenário possível é aquele no qual não há, na escola, nenhum aluno com proficiência situada no nível superior de desempenho. Além disso, convencionou-se que uma medida de 40% dos alunos situados na categoria superior de desempenho configura-se uma situação minimamente desejável.

Acerca de todos os valores de desempenho superior, ou minimamente desejável, também cabe dizer que, embora sua determinação precisa tenha um certo grau de arbitrariedade, é inegável que todos eles traduzem situações de desempenho, no mínimo, razoável. E, por outro lado, eles também têm o aspecto conveniente de não serem, ao menos em princípio, metas excessivamente elevadas ou ambiciosas, as quais, se vierem a se adotar, podem acarretar o seríssimo problema de desmotivar os avaliados, diante da impossibilidade prática de alcançá-las.

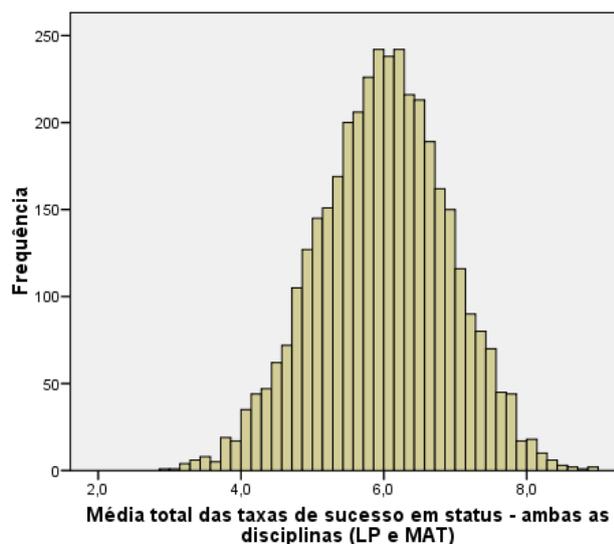
Uma vez obtidos todos esses indicadores em uma mesma métrica, o passo seguinte foi unificá-los. Primeiro, essa agregação se deu no nível de cada disciplina:

$$I_{status} = (X - X_{inf}) / (X_{sup} - X_{inf}) * 10$$

Por fim, obteve-se um único indicador de status agrupando os dois das duas diferentes disciplinas. A Tabela 2.10, a seguir, apresenta algumas estatísticas descritivas sobre esses indicadores. Cabe observar que os valores por ela apresentados referem-se, exclusivamente, às escolas do SIMAVE que obedeceram a dois critérios básicos: (1) tiveram, ao longo da série histórica aqui considerada, um número médio de alunos avaliados pelo menos igual a 10, e (2) foram avaliadas ao menos cinco vezes ao longo das oito avaliações aqui consideradas.

*Tabela 4.10: Estatísticas descritivas das taxas de status do progresso das escolas, em português, matemática e também para um total combinado de ambas – SIMAVE – 2000 a 2011 – 8ª série/9º ano EF (com ponderação pelo número médio de alunos avaliados)*

	LP	MAT	TOTAL
<i>N</i>	3809	3806	3806
<i>Média</i>	6,5	5,8	6,1
<i>DP</i>	0,8	1,0	0,9
<i>Mínimo</i>	2,9	2,9	2,9
<i>1º quartil</i>	6,0	5,1	5,5
<i>Mediana</i>	6,5	5,7	6,1
<i>3º quartil</i>	7,0	6,4	6,7
<i>Máximo</i>	9,2	8,9	8,9



*Gráf. 4.7: Histograma das taxas de sucesso em status das escolas, conjugadas para as disciplinas de português e matemática –SIMAVE – 2000 a 2011 – 8ª série/9º ano EF (sem ponderação pelo número médio de alunos avaliados)*

Tanto pela tabela quanto pelo gráfico acima, é possível perceber que, embora os limites teóricos das taxas assim obtidas se estendam entre zero e dez, os limites práticos aproximadamente se situam entre 3 e 9.

Outras características facilmente observáveis dessas taxas é a elevada correlação positiva entre elas, e também o fato de que os índices de desempenho médio em português são um pouco superiores aos de matemática, o que, de novo, corrobora uma observação já feita neste sentido, quando se falou acerca dos indicadores primitivos de status aqui apresentados.

#### 4.5.1.1. As escolas de maior status

A obtenção do índice de status permitiu que se chegasse a determinados perfis de escolas quanto a esta característica. E um grupo de particular interesse para o presente estudo é o formado pelos estabelecimentos que, ao longo de toda a série histórica ou quase, foram capazes de conservar significativos níveis de eficácia em relação a essa característica. Isso permitiu, também, chegar às escolas de maior desempenho, cuja lista com as 30 primeiras é apresentada a seguir:

Tabela 4.11: As 30 escolas com os maiores índices de status na série histórica do SIMAVE (2000 – 2011)

POS	ÍND_ST	NOME	REDE	MUNICÍPIO	N_AL.	ISE
1	8,9	EE DR LUIZ PINTO DE ALMEIDA	E	SANTA RITA DO SAPUCAI	29	-0,5
2	8,9	EM PROF DORIO BEATO	M	BELO HORIZONTE	115	0,9
3	8,8	COL TIRADENTES PMMG	E	PASSOS	53	0,5
4	8,7	EM JOSE FERREIRA BASTOS	M	UBERABA	82	1,1
5	8,6	EM IRMA MARIA AMALIA	M	ITAMARANDIBA	36	0,4
6	8,6	EM JOSE MARIA DA FONSECA	M	MONTES CLAROS	30	0,1
7	8,6	ESCOLA MUNICIPAL MERIDIONAL	M	ITABIRA	84	0,7
8	8,5	EE DOM HELVECIO	E	IPATINGA	57	1,2
9	8,4	EE CEL XAVIER CHAVES	E	CEL. XAVIER CHAVES	127	0,9
10	8,4	EM PIO XII	M	S. SEBAST. DO PARAISO	138	1,3
11	8,4	EM ANA AMELIA QUEIROZ	M	POCO FUNDO	61	-0,4
12	8,3	COL TIRADENTES PMMG	E	BOM DESPACHO	30	0,7
13	8,3	EE FREDERICO ZACARIAS	E	ABAETE	76	1,2
14	8,3	EE SORAMA GERALDA R. XAVIER	E	BIQUINHAS	44	-0,1
15	8,3	EM SAO SIMAO	M	VESPASIANO	128	1
16	8,2	COL. MUN. ARCEBURGUENSE	M	BELO HORIZONTE	221	0,8
17	8,2	EM PROF OTAVIO BATISTA C. FL.	M	BELO HORIZONTE	58	0,7
18	8,2	EM MSR. JOSE CARNEIRO PINTO	M	MARTINHO CAMPOS	72	0
19	8,2	COL TIRADENTES PMMG	E	PATOS DE MINAS	32	0
20	8,2	EM CORONEL JOAO DOMINGOS	M	LAVRAS	28	1,1
21	8,2	EM PROF SERGIO O. MARQUEZ	M	CONSELHEIRO LAFAIETE	113	0,2
22	8,2	EM DR JOSE VARGAS DE SOUZA	M	ITABIRITO	31	0
23	8,2	EE PROF LUIZ ANTONIO C OLIV.	E	ARAXA	23	-0,1
24	8,2	EM DELFIM MOREIRA	M	TEOFILO OTONI	52	0
25	8,1	EE ALMIRANTE TOYODA	E	IPATINGA	61	0
26	8,1	EE ENG. MARCIO A. DA CUNHA	E	IPATINGA	102	0,9
27	8,1	EE SAO FRANCISCO DE ASSIS	E	DIVINOPOLIS	178	1
28	8,1	EE SAO LUIZ GONZAGA	E	ELOI MENDES	203	0,9
29	8,1	COL TIRADENTES PMMG	E	IPATINGA	20	-0,4
30	8,1	EM M. SALVADOR DE OLIVEIRA	M	PONTE NOVA	132	0,8

Observa-se que, das 30 escolas de maior status, 27 encontram-se fora da capital do estado, Belo Horizonte, e suas proporções quanto às redes de ensino são quase iguais, com 14 escolas estaduais e 16 municipais. Entre as estaduais, observam-se quatro “Colégios

Tiradentes”, pertencentes à Polícia Militar de Minas Gerais, e destinados principalmente aos filhos dos membros daquela corporação. Quanto ao tamanho, é de se observar que essas escolas de maior status de proficiência tendem a apresentar um tamanho mediano, com várias delas possuindo entre 100 e 200 alunos na série considerada, mas nunca indo além disso. Por sua vez, quanto ao índice socioeconômico, um tanto quanto previsivelmente, observa-se que as escolas deste grupo tendem a apresentar maiores índices de nível socioeconômico do que a média estadual. Isto se constata no fato de que apenas 5 dessas 30 escolas apresentam um ISE negativo (e, portanto, abaixo da média estadual), muito embora, ainda sendo negativos, tais índices encontram-se bastante próximos de zero.

#### 4.5.1.2. As escolas de menor status

A Tabela 4.12, a seguir, apresenta as escolas de pior resultado médio de status ao longo da série histórica do SIMAVE. Uma informação que implicitamente aparece na coluna dos municípios é o fato de que a maioria dessas escolas situam-se nas áreas mais pobres do norte do estado, particularmente nas circunscrições das Superintendências Regionais de Educação (SREs) de Januária e Janaúba. Isso também se associa ao fato de que o índice socioeconômico médio de todo o grupo, com pouquíssimas exceções, é negativo, indicando uma situação abaixo do nível das escolas públicas estaduais. Outra característica marcante dessas escolas de menor status é o seu número consideravelmente reduzido de alunos. Isto se percebe no fato de que dois terços delas não atingem, em média, 30 alunos matriculados nesta série por ano. Tal fato também sugere que se trata, em sua quase totalidade, de escolas com somente uma turma no 9º ano do ensino fundamental.

Tabela 4.12: As 30 escolas com os menores índices de status na série histórica do SIMAVE (2000 – 2011)

POS.	ÍND_ST	NOME	RED E	MUNICÍPIO	N_AL.	ISE
-1	2,9	EM SUSSUARANA I	M	JANUARIA	18	-1,6
-2	3,1	EE CESARIO NUNES DOS SANTOS	E	JANUARIA	19	-1,6
-3	3,2	EM DE VARGINHA	M	BONITO DE MINAS	19	-1,6
-4	3,2	EE SANTOS CARVALHAIS	E	JANAUBA	38	-1
-5	3,3	EM SANTOS REIS	M	JANUARIA	14	-1,4
-6	3,3	EM GERINO FERREIRA COSTA	M	RIO VERMELHO	21	-0,2
-7	3,4	EE SAO JOSE DO GIBAO	E	PAI PEDRO	33	-0,9
-8	3,4	EE ELOI FERREIRA DA SILVA	E	IBIRACATU	24	-1,2
-9	3,4	EE GILBERTO ALVES COUTINHO	E	MIRABELA	27	-0,9
-10	3,4	EM DE SAO GERALDO	M	ARINOS	18	-1,2
-11	3,4	EE CARMOSINA DURAES MARTINS	E	CAPITAO ENEAS	35	-1
-12	3,4	EM MAL FLORIANO	M	RIO PARDO DE MINAS	46	-1,1
-13	3,4	EM JORN.BICALHO BRANDAO	M	BONITO DE MINAS	14	-1,2
-14	3,4	EM PE RICARDO	M	URUCUIA	38	-1,4
-15	3,5	EM IRMA MARIA GEMA	M	VARZELANDIA	18	-0,7
-16	3,5	EM DE GROTIHA	M	FRANCISCO SA	57	-1,1
-17	3,5	EM JOAO LOPES CORREA	M	ARINOS	27	-1,3
-18	3,5	EM TOMAS GONZAGA	M	JAIBA	15	-1,1
-19	3,5	EM JACINTO SILVEIRA NETO	M	VERDELANDIA	22	-1,1
-20	3,6	EM DE BARREIRAS	M	RIACHO D/MACHADOS	31	-1,3
-21	3,6	EE BARREIRO DA RAIZ	E	ESPINOSA	76	-1,1
-22	3,6	EM JOSE DE ALENCAR	M	ARACUAI	20	-1,6
-23	3,6	EM MARIA ALVES COUTINHO	M	JANUARIA	26	-1,2
-24	3,6	EE STA MARIA	E	JANUARIA	24	-1
-25	3,7	EE ADOLFO FERREIRA DE BARROS	E	JANUARIA	25	-1,2
-26	3,7	EE WASHINGTON XAVIER MENDES	E	BELO HORIZONTE	68	0,2
-27	3,7	EE FAUSTINO PACHECO	E	ARINOS	19	-1,2
-28	3,7	EM DE AREIAO	M	CONEGO MARINHO	15	-1,2
-29	3,7	EM SEN LEVINDO COELHO	M	MIRAVANIA	23	-0,9
-30	3,7	EM RIVALINO ALVARO DURAES	M	CAPITAO ENEAS	46	-0,8

#### 4.5.1.3. Interpretando os índices de status

Para verificar o comportamento do índice de status, bem como visualizar as diferenças entre as escolas associadas a variações nesta medida, apresentaremos aqui os resultados da comparação entre dois grupos de estabelecimentos situados nos extremos da distribuição deste índice. De um lado, forma-se um grupo de status bastante elevado, correspondente àquelas

escolas ocupando um percentil igual ou superior a 95 da distribuição. De outro lado, situam-se os 5% de escolas com o desempenho mais baixo.

Os três gráficos a seguir apresentam as distribuições, respectivamente, dos percentuais médios de alunos situados no nível inferior de proficiência (variável N1), das médias dos escores dos alunos e, finalmente, dos percentuais médios de estudantes com proficiência no nível superior, para cada disciplina.

É então possível constatar que, visualmente, as diferenças entre esses extremos de escolas são mais marcantes quando observadas do ponto de vista dos alunos situados nas extremidades de desempenho, quando comparadas com as diferenças entre as médias. Estas últimas distâncias, embora também sejam relevantes, ao menos em termos gráficos parecem ser um pouco menos acentuadas.

No primeiro destes gráficos, observa-se que as escolas de maior status, particularmente a partir de 2007, conseguiram baixar para somente um dígito os valores do índice N1 de Matemática. Já em Língua Portuguesa, essa barreira já se encontrava superada desde o início da série histórica. Por outro lado, entre as escolas de status mais baixo, embora os percentuais de alunos nos níveis inferiores estejam razoavelmente diminuindo, esse decréscimo não tem sido tão acentuado como deveria, e nem tão uniforme. Percebe-se, por exemplo, que as proporções de alunos nos níveis mais baixos de proficiência em português estagnaram-se em cerca de um terço ao longo das últimas quatro avaliações da série. E, em matemática, embora pareça ter havido um certo progresso durante esse mesmo período, ainda assim os valores brutos de N1 são bastante altos. Tanto que, somente nas duas últimas avaliações, eles desceram para um pouco menos da metade dos alunos avaliados.

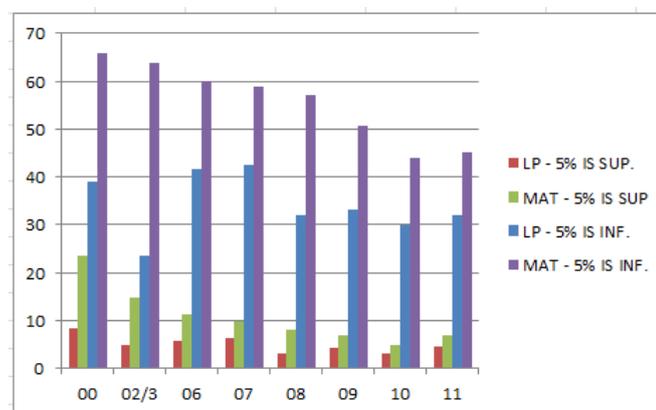


Gráfico 4.8: Distribuição das médias do índice N1 (percentual de alunos com proficiência no nível mais baixo) por disciplina, numa comparação entre os 5% de escolas com desempenho superior e inferior quanto ao índice de status (IS) na série histórica do SIMAVE (2000 – 2011).

Em relação às médias de proficiência, mostradas no gráfico a seguir, observa-se que as escolas de maior status, desde o início da série, apresentam resultados superiores às médias nacionais do SAEB/1997, de 250 pontos em cada uma das disciplinas consideradas. E que esses valores também vêm sistematicamente se elevando ao longo da série histórica, com uma tendência de se aproximar de 300 pontos, o que corresponderia a um desvio-padrão (50 pontos) a mais em relação à média da distribuição nacional do SAEB original.

Por outro lado, os 5% de escolas de pior desempenho vêm apresentando médias sistematicamente baixas, na casa dos 220 pontos ou menos em ambas as disciplinas. Além disso, e o que parece ser ainda pior, é o fato de que, embora a linha histórica dessas médias apresente uma tendência de subida, esta última parece ser bastante tímida, de modo que, caso essa mesma tendência se mantenha, tais escolas ainda levarão um tempo significativo (uma década ou mais depois do final desta série), apenas para alcançar a média nacional brasileira de 1997, um resultado que, por si, já nada tem de excepcional!

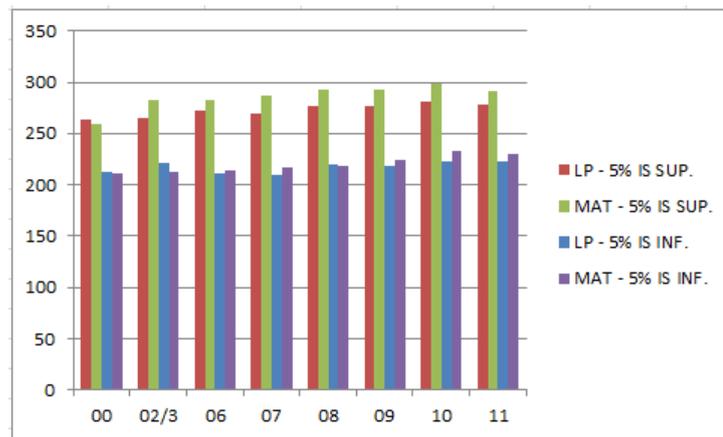


Gráfico 2.9: Distribuição das médias de proficiência por disciplina, numa comparação entre os 5% de escolas com desempenho superior e inferior quanto ao índice de status (IS) na série histórica do SIMAVE (2000 – 2011).

Por fim, o gráfico a seguir apresenta as diferenças entre esses dois extremos de escolas quanto aos percentuais médios de alunos situados no nível mais alto de proficiência, o N3. Observa-se, aqui, basicamente, um comportamento contrário ao do primeiro gráfico, referente às médias de N1.

As escolas de maior status vêm apresentando valores sempre superiores a 40% de seus alunos no nível mais elevado de proficiência em Língua Portuguesa, além de também apresentarem uma tendência de alta. Já em Matemática, embora inicialmente tenha havido certa defasagem quanto a esse índice (ou seja, percentuais relativamente modestos de N3 nos primeiros anos da série), não obstante os percentuais médios também experimentaram uma

inegável tendência de elevação, fazendo com que, nos últimos anos, o valor desse percentual tenha sido sempre superior a 40%, eventualmente aproximando-se da metade dos examinados.

Por outro lado, nas escolas de pior desempenho, os valores médios de N3, em sua maioria, situam-se abaixo dos 10%, sendo que, particularmente em Matemática, em nenhum ano eles saíram da marca de um dígito. Além disso, as tendências dessas medidas também apontam para o problema adicional, de que, embora esses índices estejam aumentando com o tempo para essas escolas, a taxa desse aumento parece ser bastante modesta, fazendo com que, mantendo-se tal perspectiva de crescimento, ainda será preciso passarem-se algumas décadas para que tais escolas atinjam valores desejáveis desses índices.

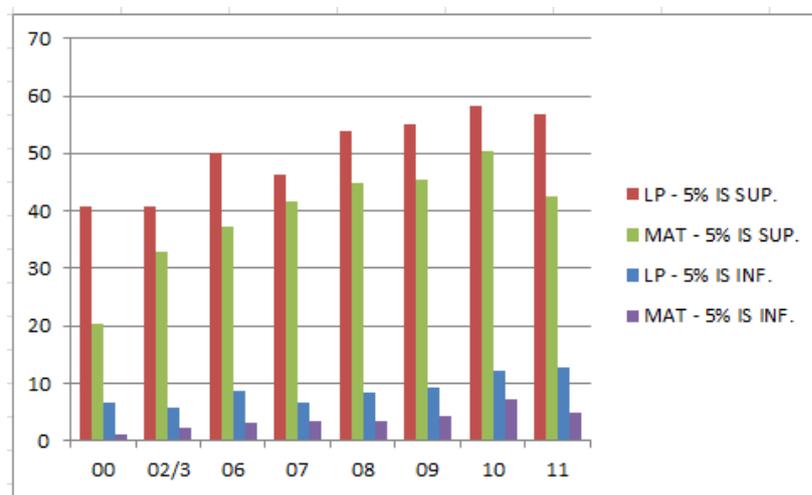


Gráfico 2.10: Distribuição das médias do índice N3 (percentual de alunos com proficiência no nível mais elevado) por disciplina, numa comparação entre os 5% de escolas com desempenho superior e inferior quanto ao índice de status (IS) na série histórica do SIMAVE (2000 – 2011).

#### 4.5.1.4. Uma análise de cluster dos extremos de desempenho quanto ao status

##### *As escolas de maior status*

A técnica estatística de clustering, ou clusterização, costuma ser bastante eficiente para se produzirem classificações de objetos ou sujeitos estudados, com base nas distâncias assumidas pelos diferentes casos observados, segundo variáveis pré-definidas. Assim, os casos que apresentam as menores distâncias médias entre si produzem grupos relativamente homogêneos quanto aos valores dessas variáveis, o que funciona como uma forma útil para a sua classificação segundo um grupo arbitrário de categorias (ALDENDERFER, 1997).

No nosso caso de estudo, uma vez selecionados os 5% de escolas de maior status médio, realizamos uma análise de cluster com o propósito de visualizar o perfil básico desses

estabelecimentos quanto a algumas variáveis fundamentais, que foram quatro: o número de alunos avaliados na série específica (funcionando como uma *proxy* do tamanho da escola), seu respectivo índice socioeconômico médio (conforme calculado em 2011) e as taxas médias de status e de progresso de cada estabelecimento, medidas estas que, conforme já se viu, combinam os índices de N1, média e N3 (ou suas respectivas variações) para ambas as disciplinas.

Os resultados que nos pareceram mais informativos foram os obtidos para a existência de dois clusters ou grupos de escolas, dados pela tabela a seguir:

*Tabela 4.13: Coordenadas dos centroides dos dois clusters nos quais se dividem os 5% de escolas de maior índice de status na série histórica do SIMAVE (2000 – 2011).*

	Cluster	
	1	2
<i>Número de escolas</i>	23	115
<i>Nº médio de alunos avaliados</i>	190	67
<i>Média do ISE</i>	0,8	0,3
<i>Média_status (0 a 10)</i>	7,9	7,9
<i>Média_progresso (0 a 10)</i>	3,2	4,0

Por esta última tabela, constata-se a existência de dois grupos de escolas compartilhando as mesmas características de elevadas taxas de status de proficiência média. Um primeiro grupo, relativamente pequeno, de 23 estabelecimentos maiores (com uma média de 190 alunos avaliados na série considerada), e um segundo grupo bem mais numeroso, com 155 escolas, porém com tamanhos individuais bem menores, correspondendo a 67 alunos em média.

Ambos esses grupos apresentam médias idênticas de índice de status (7,9) e, o que não chega a ser uma surpresa, também mostram ter índices consideravelmente baixos de progresso – 3,2 para o grupo 1 e 4,0 para o grupo 2, numa escala de 0 a 10. Isso mostra, uma vez mais, que as escolas que já possuem um elevado nível de status, não conseguem ir muito além disso, o que faz com que suas taxas de progresso sejam, de fato, moderadas, sem que necessariamente isso resulte num problema ou falha desses sistemas.

Além disso, e também um tanto quanto previsivelmente, as médias de índice socioeconômico desses estabelecimentos – principalmente no caso do grupo 1 – encontram-se acima da média estadual. Isso indica que, muito provavelmente, o elevado índice de status dessas escolas talvez não se deva somente aos méritos intrínsecos desses estabelecimentos,

mas também refletem o fato de que eles estão atendendo alunos provenientes de meios mais privilegiados do que a grande média estadual.

#### *As escolas de menor status*

A tabela a seguir apresenta medidas análogas ao caso anterior, só que agora para as escolas que vêm apresentando sistematicamente os menores índices de status no SIMAVE. Novamente, na clusterização, optou-se por adotar uma classificação em dois grupos de escolas.

*Tabela 4.14: Coordenadas dos centroides dos dois clusters nos quais se dividem os 5% de escolas de menor índice de status na série histórica do SIMAVE (2000 – 2011).*

	<i>Cluster</i>	
	<i>1</i>	<i>2</i>
<i>Número de escolas</i>	43	271
<i>Nº médio de alunos avaliados</i>	94	28
<i>Média do ISE</i>	-0,4	-0,8
<i>Média_status (0 a 10)</i>	4,4	4,2
<i>Média_progresso (0 a 10)</i>	4,0	4,4

Por esta última tabela, observa-se que a grande maioria (271 de 314, ou 86%) das escolas de status muito baixo são bastante pequenas, pois apresentam uma média de somente 28 alunos avaliados. Muito provavelmente, portanto, na maior parte dessas escolas, existe apenas uma única turma de alunos. Por outro lado, há um grupo menor, embora ainda assim significativo, de escolas um pouco maiores, com uma média de 94 alunos avaliados. As médias de status de ambas essas escolas são, naturalmente, muito baixas, com as médias dos grupos 1 e 2 correspondendo a 4,4 e 4,2, respectivamente. Já os índices de progresso desses estabelecimentos são um pouco maiores que os do grupo das escolas de maior status, o que tampouco vem a ser uma novidade. Afinal de contas, as escolas de menor status possuem um teto maior para crescer, o que resulta em valores um pouco maiores quanto ao índice de progresso calculado. Não obstante, essas taxas de progresso não parecem ser tão elevadas assim, visto que, para ambos os grupos, elas se situam na casa de 4 pontos do índice, numa escala que vai de 0 a 10. Isso mostra a existência de uma realidade perversa envolvendo essas escolas: o fato de que, além delas virem sistematicamente apresentando baixas medidas de proficiência, tampouco parecem ter a tendência de sair dessa situação negativa, ao menos a curto ou médio prazos. Por fim, também um tanto quanto previsivelmente, essas escolas

apresentam valores negativos de índice socioeconômico, indicando que as mesmas encontram-se abaixo da média estadual quanto a essa variável. Isso, de certa forma, “atenua” a responsabilidade desses estabelecimentos, dado que, como é sobejamente conhecido na literatura sobre eficácia escolar, costuma haver uma considerável associação entre o nível socioeconômico dos alunos de uma escola e o seu respectivo desempenho acadêmico.

#### 4.5.2. A busca por um indicador de progresso

A concepção de um índice de progresso dos níveis de aprendizado para as escolas do SIMAVE obedeceu, em geral, a princípios análogos àqueles vistos na mensuração de seus respectivos níveis de status. Ou seja, as medidas tiveram como base as mesmas variáveis tratadas anteriormente, de N1, média e N3. Porém a diferença, naturalmente, é que, agora, não estamos tratando dos valores primários dessas variáveis, mas sim das diferenças entre eles calculada entre duas edições consecutivas do teste.

Da mesma forma que antes, são aqui computados os índices de sucesso quanto ao progresso dos resultados, separadamente por disciplina, para depois chegar-se a um índice único que engloba os dois primeiros. Por sua vez, os índices de sucesso, para serem calculados, necessitam de parâmetros mínimos e máximos de comparação, que estão apresentados pela tabela a seguir:

*Tabela 4.15: Os limites de desempenho considerados na elaboração dos índices de progresso, para ambas as disciplinas.*

<i>Variação anual de:</i>	<i>Limite de desempenho</i>	
	<i>Inferior (pior cenário possível)</i>	<i>Superior (cenário minimamente desejável)</i>
<i>N1</i>	0 (ou posit.)	- 5%
<i>Média</i>	0 (ou negat.)	15
<i>N3</i>	0 (ou negat.)	4%

Os valores críticos inferiores acima apresentados são fáceis de entender. A pior situação ocorre quando o desempenho não melhora, ou seja, quando a sua variação vem a ser nula ou negativa. Entretanto, como variações negativas de proficiência não são muito facilmente tratáveis – em parte devido a problemas de medida, e em parte devido a variações das taxas anteriores de proficiência dos alunos avaliados, entre outros fatores mais – adotar-

se-á, neste caso, a convenção de tornar nulas todas as possíveis variações negativas de desempenho das escolas.

Por sua vez, os limites de desempenho para as taxas de progresso foram fixados a partir da observação empírica das taxas de progresso. Optou-se, então, por adotar valores aproximadamente iguais aos terceiros quartis de distribuição das mesmas. Isto faz com que os ganhos mínimos exigidos sejam consideráveis. A ponto de, por exemplo, produzirem um aumento *mínimo* de 150 pontos (= 15 vezes 10) em qualquer das escalas de proficiência – valor correspondente a cerca de 3 vezes o desvio-padrão da distribuição de escores –, caso tal taxa fosse mantida constante ao longo de toda a década abrangendo a série histórica considerada. Analogamente, caso essas mesmas taxas mínimas de progresso fossem mantidas, haveria um salto de 40% (= 4 x 10) de alunos no nível mais alto de proficiência, concomitante com uma diminuição de 50% de alunos situados no nível inferior de proficiência.

As fórmulas para esses índices de progresso são, matematicamente análogas àquelas já fixadas para as taxas de status. Dessa forma, todas elas também são expressas em índices que variam de 0 a 10, e são dadas por:

$$I_{progresso} = (X - X_{inf}) / (X_{sup} - X_{inf}) * 10$$

A Tabela 4.16, a seguir, apresenta as principais estatísticas descritivas dos índices de progresso das proficiências das escolas do SIMAVE. Da mesma forma que no caso do status de proficiência, as estatísticas descritivas e o gráfico a seguir foram obtidos somente com base nas instituições que foram avaliadas pelo menos 5 vezes na série, e também que contaram com um número mínimo de 10 alunos avaliados em média.

*Tabela 4.16: Estatísticas descritivas das taxas de sucesso do progresso das escolas, em português, matemática e também para um total combinado de ambas –SIMAVE – 2000 a 2011 – 8ª série/9º ano EF (com ponderação pelo número médio de alunos avaliados)*

	<i>LP</i>	<i>MAT</i>	<i>TOTAL</i>
<i>N</i>	3811	3811	3726
<i>Média</i>	4,0	4,6	2,8
<i>DP</i>	1,1	1,1	3,5
<i>Mínimo</i>	0	0	0
<i>1º quartil</i>	3,3	4,0	0
<i>Mediana</i>	4,0	4,7	0,3
<i>3º quartil</i>	4,8	5,3	6,3
<i>Máximo</i>	9,9	8,9	9,9

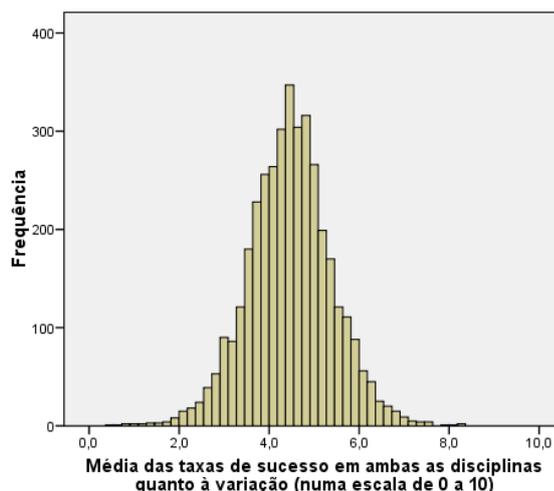


Gráfico 4.11: Histograma das taxas de sucesso quanto ao progresso das escolas, conjugadas para as disciplinas de português e matemática –SIMAVE – 2000 a 2011 – 8ª série/9º ano EF (sem ponderação pelo número médio de alunos avaliados).

Vê-se, portanto, que esse índice varia, grosso modo e na prática, entre 2 e 8. Apresentemos, então, os casos extremos – tanto superiores quanto inferiores – de valores desses índices entre as escolas do SIMAVE.

#### 4.5.2.1. As escolas de maior progresso

A Tabela 4.17, a seguir, apresenta, a título de ilustração, as 30 escolas com os maiores índices médios de progresso ao longo da série histórica do SIMAVE.

Tabela 4.17: As 30 escolas com os maiores índices de progresso na série histórica do SIMAVE (2000 – 2011)

POS.	ÍND_PROGR	ESCOLA	REDE	MUNICÍPIO	N_AL	ISE
1	8,3	EM FAZ LUCINDA	M	MATERLANDIA	21	-0,7
2	8,2	EE PROF HILDEBRANDO PONTES	E	UBERABA	19	0,3
3	8,1	EM VICENTINA DE ABREU SILVA	M	LAVRAS	11	-0,4
4	8	EM PROFA ESMERALDA VIANNA	M	MURIAE	33	-0,2
5	7,6	EM MARIA DE LOURDES OLIVEIRA	M	BETIM	43	0
6	7,6	EM VILMA DE FARIA SILVA	M	IPATINGA	64	0,2
7	7,5	EM JOUBERT DE CARVALHO	M	UBERABA	34	0,3
8	7,5	EM TIAGO MEDEIROS FONSECA	M	FRANCISCO DUMONT	16	-0,9
9	7,4	EM ANTONIO FONSECA LEAL	M	TRES MARIAS	86	0,4
10	7,4	EE MESTRA AURORA	E	CARBONITA	44	-0,3
11	7,3	EM DE SAO PEDRO	M	ACUCENA	10	-1
12	7,3	EM ALIPIO MACIEL DE OLIVEIRA	M	SETE LAGOAS	69	0
13	7,2	EE DR TANCREDO NEVES	E	JOSE GONÇ. DE MINAS	69	-0,7
14	7,2	EM CRISTO REI	M	CORINTO	59	-0,4
15	7,1	EM DE LAVRAS NOVAS	M	OURO PRETO	10	0,3

16	7,1	EM CEL FRANCISCO F DE CARVALHO	M	OLIVEIRA FORTES	24	-0,2
17	7,1	EM ADAIR DE OLIVEIRA PINTO	M	QUARTEL GERAL	37	0,1
18	7,1	EM MARIA CAPRONI DE OLIVEIRA	M	CARVALHOPOLIS	46	0,3
19	7,1	EM PEDRO EPIFANIO	M	FELIXLANDIA	11	-0,8
20	7,1	EE MIGUEL CHIARADIA	E	CAMANDUCAIA	20	0
22	7	EM DE AGUAS CLARAS	M	MARIANA	17	-0,4
23	7	EE SEM. FRANCISCO N. COELHO	E	GUANHAES	58	-0,3
24	7	EM ANISIO ACELINO DE ANDRADE	M	VIEIRAS	21	-0,2
25	7	EM ANTONIO MATIAS PEREIRA	M	COROMANDEL	24	-0,3
26	6,9	EE DE FABIAO	E	JANUARIA	17	-0,7
27	6,9	EM PROFA IRENE PINTO	M	SABARA	50	0,2
28	6,9	EM CARLOS PRATES	M	CENTRALINA	37	-0,1
29	6,9	EE INCONFIDENTES	E	MONTALVANIA	35	-0,8
30	6,9	EM DOM JOSE DE HAAS	M	PADRE PARAISO	47	-1,1
31	6,9	EM PROFA MA. APAREC. M PRADO	M	TIMOTEO	107	0,5

Por esta tabela, é possível observar algumas características da escola “prototípica” mineira, quanto ao sucesso no avanço anual da proficiência em ambas as disciplinas. Para começar, trata-se de escolas consideravelmente pequenas quanto ao número de alunos examinados, com uma média de 37 estudantes avaliados por ano e disciplina. Isto significa também que provavelmente a maioria dessas escolas possui somente uma única turma no 9º ano do ensino fundamental. Além disso, dessas 30 escolas, nada menos que 23 pertencem a redes municipais de educação. Quanto ao índice socioeconômico médio de sua clientela, uma característica marcante deste grupo é a sua grande proximidade da média da rede pública estadual. Isto sinaliza, ao menos em tese, que, apesar dessas escolas não terem um índice socioeconômico médio excepcional, não obstante, elas apresentam taxas quase excepcionais e bastante consistentes de progresso anual médio.

#### 4.5.2.2. As escolas de menor progresso

Pelo índice de progresso, também é possível detectar os estabelecimentos que, ao longo da série histórica do SIMAVE, mantiveram-se consistentemente com taxas consideravelmente baixas de progresso. A tabela a seguir apresenta os 30 casos de menores taxas de progresso da proficiência combinada de português e matemática.

Tabela 4.18: As 30 escolas com os menores índices de progresso na série histórica do SIMAVE (2000 – 2011)

POS.	ÍND_PROGR	ESCOLA	REDE	MUNICÍPIO	N_AL.	ISE
-1	0,4	EE MA. DA GLORIA ASSUNCAO	E	RIBEIRAO DAS NEVES	72	0
-2	0,7	EM OSORIO ALEIXO DA SILVA	M	CATAGUASES	13	-0,3
-3	0,7	EM PRES RAUL SOARES	M	BELO HORIZONTE	110	0,3
-4	0,8	EM ANTONIO DE CASTRO PINTO	M	ITAUNA	96	0,6
-5	1	EE PROFESSOR SOARES FERREIRA	E	BETIM	83	0,3
-6	1,1	EM SUSSUARANA I	M	JANUARIA	25	-1,2
-7	1,1	EE JOAO MARIANO RIBEIRO	E	BETIM	162	0,5
-8	1,3	EM SEBASTIAO F. DE OLIVEIRA	M	SAO BRAS DO SUACUI	49	0
-9	1,3	EE REPARATA DIAS DE OLIVEIRA	E	GOV. VALADARES	19	-0,3
-10	1,4	EM PROF DOMINGOS P.ULHOA	M	ITAPEVA	63	0,5
-11	1,5	EM SALGADO FILHO	M	CAETE	71	0,5
-12	1,5	EM BOAVENTURA ABRITTA	M	RUBELITA	16	-1,3
-13	1,5	EM MURILO RUBIAO	M	VARGINHA	72	0,6
-14	1,7	EE PROFA GERALDA M. L. MELO	E	BARBACENA	136	0,5
-15	1,7	EM DE AREIAO	M	SIMAO PEREIRA	22	0
-16	1,7	EM DSBG. APRIGIO R. OLIVEIRA	M	LAGOA DOURADA	50	-0,1
-17	1,7	EE DR JOSE RODRIGUES SEABRA	E	JANUARIA	18	-1,6
-18	1,8	EE FRANCISCO DE PAULA CASTRO	E	CONC,DO MATO DENTRO	15	-0,7
-19	1,9	EE RUI BARBOSA	E	BETIM	33	0,1
-20	1,9	EM PROFESSORA HELENA REIS	M	SANTA RITA DO SAPUCAI	133	0,7
-21	1,9	EM CONCEICAO APCD. ROSSO	M	GUARARA	22	0,1
-22	1,9	EE ABEILARD PEREIRA	E	LAGOA SANTA	108	0,1
-23	1,9	EE DOUTOR DELFIM MOREIRA	E	UBERLANDIA	130	0,8
-24	1,9	EM FERREIRA MARQUES	M	PATIS	34	-0,8
-25	1,9	EM FRANCISCO SOARES	M	BELO HORIZONTE	81	0,2
-26	2	EM FLORESTAN FERNANDES	M	NOVA SERRANA	85	0,1
-27	2	E M JOSE AMERICO DE LACERDA	M	BURITIS	18	-0,3
-28	2	EE ANALIA C. DOS SANTOS	E	BELO HORIZONTE	225	0,7
-29	2	EM ISAURA COELHO	M	BETIM	79	0,4
-30	2,1	EM FERNANDO DIAS COSTA	M	BELO HORIZONTE	105	0

É curioso observar que as escolas com as piores taxas de progresso também apresentam, em geral, índices socioeconômicos médios bastante próximos da grande média estadual, característica também compartilhada pelas suas contrárias, ou seja, pelas escolas de maiores níveis de progresso médio, conforme detalhado anteriormente. Ou, em outras palavras, ao menos nesses casos, o índice socioeconômico parece se associar mais fracamente

ao desempenho das escolas do que outras possíveis variáveis, entre as quais se presumem algumas relacionadas ao gerenciamento e à condução das práticas educacionais, algo, que, entretanto, teria que ser objeto de uma investigação à parte.

Por outro lado, quanto ao tamanho – medido pela quantidade média de alunos matriculados –, as escolas de menor progresso diferem-se consideravelmente das de maiores taxas de avanço. Enquanto as mais progressistas se apresentam, em geral, consideravelmente pequenas, as escolas estagnadas ou mesmo decadentes são bastante mais heterogêneas quanto a essa medida, variando desde estabelecimentos bastante pequenos – na casa de poucas dezenas de alunos matriculados – até chegarem a casos de mais de uma centena deles, com uma delas chegando a 225 discentes.

#### 4.5.2.3. Interpretando os índices de progresso

Os gráficos a seguir apresentam os resultados médios dos três indicadores já anteriormente comentados – N1, média e N3 – referentes a dois extremos de escolas quanto ao índice de progresso: um grupo formado pelos 5% de estabelecimentos com o desempenho mais alto, e outro com os 5% de desempenho mais baixo.

No primeiro gráfico desta sequência, percebe-se que as escolas com os maiores índices de progresso apresentam, quase sempre, variações negativas de N1, indicando que, a cada avaliação, o percentual de alunos situados no nível inferior de proficiência tende a diminuir, em relação à avaliação anterior.

Por outro lado, as escolas com os piores índices de progresso, previsivelmente, não apresentam essas mesmas características, sendo que, em alguns casos, têm ocorrido situações de aumento substancial de alunos no extremo mais baixo de proficiência.

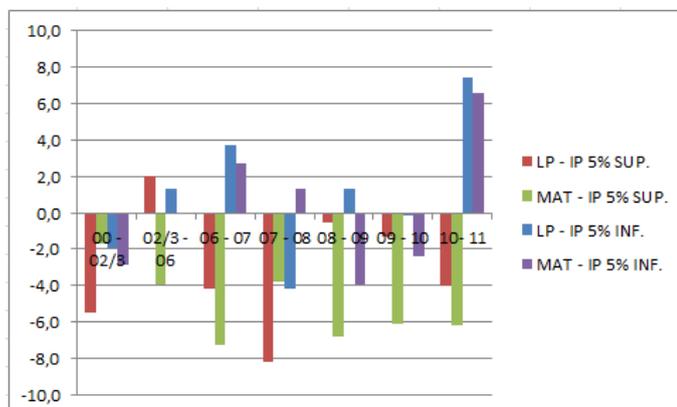


Gráfico 4.12: Distribuição das médias de variação do índice N1 (percentual de alunos com proficiência no nível mais baixo) por disciplina, numa comparação entre os 5% de escolas com desempenho superior e inferior quanto ao índice de progresso (IP) na série histórica do SIMAVE (2000 – 2011).

Já no gráfico a seguir, observa-se que as escolas de maior progresso vêm, em geral, experimentando aumentos consideráveis de suas médias de português e matemática. Esta última disciplina também parece ser a que mais tem aumentado de desempenho médio, chegando a vários casos em que a variação positiva desse indicador se aproximou de médias de 10 pontos na sua própria escala de proficiência.

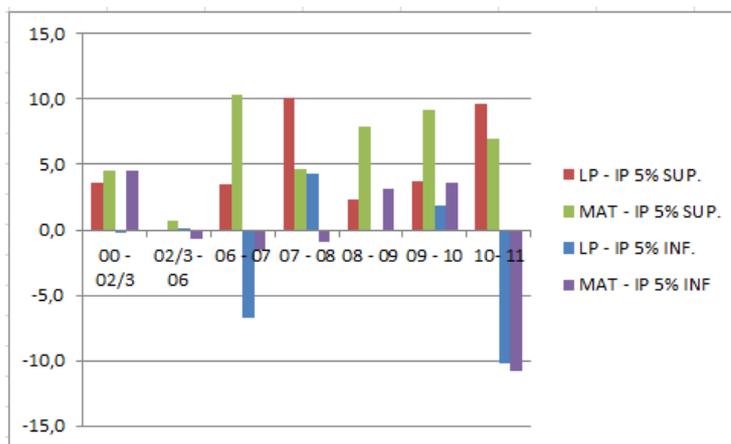


Gráfico 4.13: Distribuição das médias de variação de proficiência por disciplina, numa comparação entre os 5% de escolas com desempenho superior e inferior quanto ao índice de progresso (IP) na série histórica do SIMAVE (2000 – 2011).

Por fim, os gráficos das variações de N3 mostram que, entre as escolas com maiores taxas médias de progresso, estão subindo consideravelmente os percentuais de alunos situados no nível mais elevado de proficiência, notadamente no caso de Língua Portuguesa. Já entre as escolas com os piores índices de progresso, observa-se uma certa alternância entre subidas e

descidas do nível desses percentuais, fazendo com que, quando tomada como um todo, a série histórica apresenta pouca variação geral quanto a esse índice.

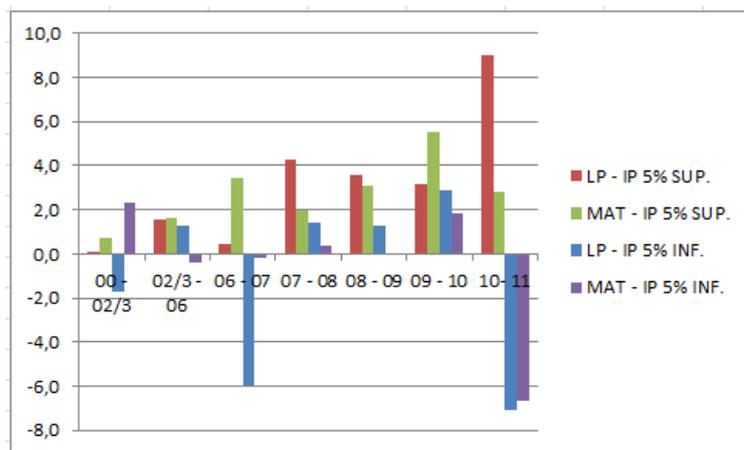


Gráfico 4.14: Distribuição das médias de variação do índice N3 (percentual de alunos com proficiência no nível mais elevado) por disciplina, numa comparação entre os 5% de escolas com desempenho superior e inferior quanto ao índice de progresso (IP) na série histórica do SIMAVE (2000 – 2011).

#### 4.5.2.4. Uma análise de cluster dos extremos de desempenho quanto ao progresso

##### *As escolas de maior progresso*

Apresentamos, a seguir, a tabela com os resultados da clusterização das escolas de maior progresso.

Tabela 4.19: Coordenadas dos centroides dos dois clusters nos quais se dividem os 5% de escolas de maior índice de progresso na série histórica do SIMAVE (2000 – 2011).

	Cluster	
	1	2
Número de escolas	62	219
Nº médio de alunos avaliados	93	30
Média do ISE	0,1	-0,4
Média_status (0 a 10)	5,9	5,9
Média_progresso (0 a 10)	6,2	6,3

Um ponto facilmente observado, quanto às escolas de maior progresso médio, é o seu tamanho relativamente pequeno: a grande maioria delas (cerca de quatro em cada cinco) possuem somente 30 alunos em média avaliados na série em questão. E, mesmo o outro grupo, formado por escolas maiores, possui uma média de 93 alunos avaliados por ano nessa mesma série. Isso parece apontar para o fato de que as ações de progresso educacional que porventura vêm se verificando ao menos em parte das escolas mineiras, são mais fáceis de serem efetivamente implementadas em estabelecimentos menores, cujo menor tamanho possivelmente facilita as atividades de implementação e acompanhamento dos processos de melhoria escolar.

Outro aspecto de interesse, ao observar essas escolas com as maiores taxas de progresso, é o fato de que seus respectivos índices de status são razoavelmente elevados, ainda que não excepcionais. Quanto a eles não serem excepcionais, isto é até previsível, dado que, conforme já se comentou, índices elevadíssimos de status tendem a impedir a ocorrência de taxas demasiadamente altas de progresso. Porém, mesmo apesar dos avanços anuais estarem entre os maiores possíveis, para o caso dessas escolas, as suas taxas de status consideravelmente elevadas indicam que elas, de uma certa maneira, estão conseguindo obter uma saudável e um tanto quanto difícil ou rara combinação de altos valores de status e de progresso.

Por fim, quanto ao índice socioeconômico, cabe observar que elas se encontram aproximadamente na grande média estadual, com um grupo um pouco abaixo e outro um pouco acima desse valor médio.

#### *As escolas de menor progresso*

A tabela a seguir apresenta os resultados do processo de clusterização para as escolas com as menores taxas de progresso ao longo da série histórica observada.

*Tabela 4.20: Coordenadas dos centroides dos dois clusters nos quais se dividem os 5% de escolas de menor índice de progresso na série histórica do SIMAVE (2000 – 2011).*

	Cluster	
	1	2
Número de escolas	54	117
Nº médio de alunos avaliados	143	48
Média do ISE	0,5	-0,2
Média_status (0 a 10)	6,6	5,9
Média_progresso (0 a 10)	2,4	2,4

Por esta tabela, percebe-se que, ao passo que os valores médios dos índices de progresso são bastante baixos, os valores de status, por outro lado, são consideravelmente altos. Portanto, mais uma vez cabe considerar que o fato de que valores muito baixos para o progresso não necessariamente se associam a escolas “ruins”. Isto porque, pelo menos algumas dessas escolas, possivelmente não subiram muito de desempenho porque já vêm apresentando medidas consideravelmente elevadas dos mesmos.

Quanto ao tamanho dos estabelecimentos classificados nesta última categoria, observa-se um grupo com escolas razoavelmente grandes, apresentando uma média de 143 alunos avaliados a cada ano.

#### 4.6. Um indicador unificado de desempenho

Com base nos índices de status e de progresso obtidos para as escolas do SIMAVE, foi então possível obter um índice único de eficácia escolar, correspondente ao produto dos dois primeiros dividido por 10, para ser expresso também em uma escala de 0 a 10:

$$I_{total} = I_{status} * I_{progresso} / 10$$

O gráfico e a tabela a seguir fornecem alguns dos principais aspectos da distribuição deste novo índice. Observa-se que o histograma possui uma acentuada simetria e também tem seu centro quase coincidente com o meio exato de 5 da escala. Entretanto, os valores práticos não se estendem até os limites de 0 a 10, situando-se antes, entre 2 e 7,5. Esse certo distanciamento entre os valores extremos práticos e os limites teóricos, no entanto, não chega a ser surpreendente. Isto porque, na prática, seria – e é – extremamente difícil obter simultânea e consistentemente medidas muito altas de status e de progresso. Tal fato ocorre porque, naturalmente, estas duas variáveis tendem a se desenvolver no sentido inverso uma da outra: uma escola com um baixo status de proficiência tem muito teto para crescer, ao contrário de outra que já vem apresentando níveis elevados de desempenho.

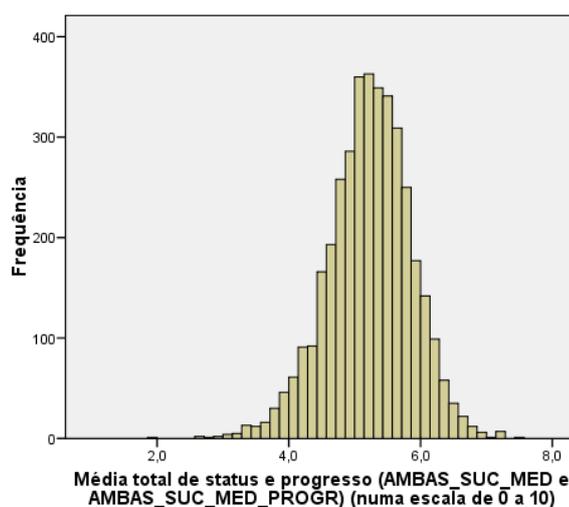
Portanto, é relevante observar que esse índice total de desempenho precisa ser visto de um modo diferente da interpretação dada separadamente aos índices de status e de progresso, antes tratados. Nestes dois últimos casos, de fato, valores pequenos ou moderados dos índices traduziam corretamente a ideia de que, ao menos no que diz respeito àquilo que tais indicadores mediam, era, sim, possível, dizer que faltava eficácia às escolas portadoras de resultados pouco lustrosos. Já no caso do índice total, um valor moderado pode ser obtido por uma escola, digamos, com um desempenho em status consistentemente muito bom ou mesmo excepcional, mas que, no entanto, não tem muito teto para aumentar essas medidas. E, dessa

forma, em seu índice total, essa escola perde terreno quando sua elevada média de status é multiplicada pela relativamente baixa taxa de progresso anual, resultando assim em um produto, ou índice total, não muito elevado.

Ainda assim, também é possível encontrar elevados valores de índice total, indicando casos de escolas que, não somente estão conseguindo bons níveis de status de proficiência, como também encontram-se melhorando ainda mais esse processo ao longo do tempo. Mas, por outro lado, também existem os casos de escolas que, como se não bastassem os desempenhos ruins que vêm apresentando em termos de status, ainda encontram-se com sua proficiência média estagnada ou, num possível cenário ainda pior, diminuindo ao longo do tempo.

*Tabela 4.21: Estatísticas descritivas básicas do índice total de desempenho (status mais progresso) na série histórica do SIMAVE (2000 – 2011), com ponderação pelo número médio de alunos avaliados.*

<i>N</i>	3811,0
<i>Média</i>	5,2
<i>DP</i>	0,6
<i>Mínimo</i>	2,0
<i>Máximo</i>	7,5
<i>1º quartil</i>	4,9
<i>Mediana</i>	5,3
<i>3º quartil</i>	5,6



*Gráfico 4.15: A distribuição do índice total de desempenho (status mais progresso) na série histórica do SIMAVE (2000 – 2011), sem ponderação pelo número médio de alunos avaliados.*

#### 4.6.1. As escolas nas extremidades do índice total de desenvolvimento

Em geral, as escolas que obtiveram os maiores índices totais de desenvolvimento apresentam um comportamento que, simultaneamente, é semelhante ao já descrito para dois grupos: o das escolas com elevadas taxas de status, e também o daquelas que apresentam altas taxas de progresso. E algo análogo também se passa em relação às escolas no outro extremo dessa medida, ou seja, naquelas com os piores índices de desenvolvimento total.

Em decorrência disso, para evitarmos certa repetitividade, não apresentaremos aqui os mesmos tipos de análises já realizadas quando discorrermos sobre o comportamento dos índices anteriores. Não obstante, a fim de traçarmos um perfil um pouco mais fiel das escolas que se encontram nesses extremos de desempenho, forneceremos aqui algumas informações específicas sobre elas, como as listas daquelas com os 30 maiores ou menores índices, bem como uma descrição rápida dos resultados da clusterização das escolas com os melhores e os piores resultados.

Nesse sentido, a Tabela 4.22, a seguir, detalha as 30 escolas com os maiores índices médios totais de desenvolvimento. Examinando essa tabela, é possível perceber algumas características marcantes desse conjunto. Uma delas é o fato de sua grande maioria ser municipal (25 das 30 escolas consideradas). Além disso, em termos do índice socioeconômico, elas se situam aproximadamente próximas da grande média estadual. Outra característica marcante é o pequeno tamanho dessas escolas: 24 delas, ou 80%, têm menos que 50 alunos em média avaliados a cada ano, evidenciando, assim, o fato de que elas, em grande parte, são formadas apenas de uma turma na série em questão.

Este último fato, por sua vez, leva-nos a fazer, ao menos, duas conjeturas mutuamente excludentes que mereceriam uma investigação mais cuidadosa posteriormente, talvez a ser levada a cabo pelos gestores interessados neste sistema. Uma delas se refere a uma possível maior facilidade, por parte das escolas menores, implementar com sucesso processos de aumento ou manutenção de altos índices de qualidade educacional. Como se trata de escolas muito pequenas, talvez esse menor tamanho facilite, por exemplo, as atividades de implementação de programas e monitoramento dos mesmos por parte da gestão escolar. Ou talvez, nesses casos, também haja um clima mais favorável para um maior engajamento de professores e da comunidade nos processos de desenvolvimento escolar.

Por outro lado, uma conjetura oposta, que muito provavelmente não se aplica a todos os casos de desempenho excepcional, mas que talvez se verifique com algumas dessas escolas, seja a presença pura e simples de situações de fraudes nas avaliações... Nesse sentido,

uma ideia razoável que os sistemas educacionais talvez deveriam implementar, poderia ser uma espécie de auditoria, pela qual, posteriormente à obtenção e mesmo divulgação dos resultados, houvesse pelo menos uma seleção aleatória de escolas a serem observadas mais atentamente, com o objetivo de se comprovar – ou não – a plausibilidade das situações de aprendizagem nelas mensuradas através dos testes. E tal prática teria o duplo benefício de, por um lado, vir a detectar casos genuínos de excelência educacional – que poderiam ser então devidamente divulgados e talvez replicados em outros contextos –, como também, ao mesmo tempo, funcionaria como uma freio às possíveis tentativas ou ameaças de fraude nessas avaliações, problema que, inclusive, caso venha a existir, tem a preocupante capacidade de destruir a própria credibilidade do sistema avaliativo.

*Tabela 4.22: As 30 escolas com os maiores índices totais de desenvolvimento (status e progresso combinados) na série histórica do SIMAVE (2000 – 2011)*

POS.	ÍND_TOTAL	ESCOLA	REDE	MUNICÍPIO	N_AL.	ISE
1	7,5	EM VICENTINA DE ABREU SILVA	M	LAVRAS	11	-0,4
2	7,3	C EDUC MUN PROF ALC. R PEREIRA	M	ITABIRITO	106	0,5
3	7,3	EM ANTONIO FONSECA LEAL	M	TRES MARIAS	86	0,4
4	7,3	EM MARIA CAPRONI DE OLIVEIRA	M	CARVALHOPOLIS	46	0,3
5	7,3	EE PROF. HILDEBRANDO PONTES	E	UBERABA	19	0,3
6	7,2	EM PROFA ESMERALDA VIANNA	M	MURIAE	33	-0,2
7	7,2	EM ANTONIO MATIAS PEREIRA	M	COROMANDEL	24	-0,3
8	7,2	EM JOUBERT DE CARVALHO	M	UBERABA	34	0,3
9	7	EE MESTRA AURORA	E	CARBONITA	44	-0,3
10	7	EM PROFA HERMINIA CORGOZINHO	M	DIVINOPOLIS	60	0,4
11	7	EM CEL FRANCISCO F DE CARVALHO	M	OLIVEIRA FORTES	24	-0,2
12	6,9	EM JOSE PAULO DE AMORIM	M	PATOS DE MINAS	14	-0,1
13	6,9	EM ADAIR DE OLIVEIRA PINTO	M	QUARTEL GERAL	37	0,1
14	6,9	EM ANISIO ACELINO DE ANDRADE	M	VIEIRAS	21	-0,2
15	6,9	EM SAO DOMINGOS	M	ST ANT AVENTUREIRO	18	-0,4
16	6,8	EM TOMAS ANTONIO GONZAGA	M	OURO PRETO	47	0,9
17	6,8	EM PROFA ROSAL. A. NOGUEIRA	M	SABARA	25	0,3
18	6,8	EE SORAMA G. RICHARD XAVIER	E	BIQUINHAS	32	0
19	6,8	EM PROFA. GENI CHAVES	M	UBERABA	23	0,2
20	6,8	EM PROFA NIC. BERNARDO	M	POCOS DE CALDAS	21	0,3
21	6,8	EM FREDERICO CAMPOS	M	CEDRO DO ABAETE	17	-0,2
22	6,8	EM JUSCELINO KUBITSCHK	M	SANTANA DO DESERTO	30	0,1
23	6,8	EM DE AGUAS CLARAS	M	MARIANA	17	-0,4
24	6,7	EM JOSE EUSTAQUIO BORGES	M	JAPARAIBA	10	-0,1
25	6,7	EE ABNER AFONSO	E	PATOS DE MINAS	139	0,5

26	6,7	EM PE XISTO	M	BRUMADINHO	25	0,1
27	6,7	EM VILMA DE FARIA SILVA	M	IPATINGA	64	0,2
28	6,7	EE PE JOSE SENABRE	E	VESPASIANO	107	0,8
29	6,7	EM MARLON DOS REIS LOPES	M	STA BARBARA DO LESTE	31	-0,7
30	6,7	EM BARAO DO RIO BRANCO	M	SAO DOM. DAS DORES	40	-0,7

#### 4.6.2. As escolas de menor índice total de desenvolvimento

As 30 escolas de pior desempenho total apresentam um perfil semelhante ao daquelas já apontadas como tendo as menores medidas de status e/ou progresso dos resultados. Observa-se que, inclusive, algumas delas já apareceram entre as de resultados mais fracos segundo um dos dois critérios anteriormente observados.

Dessa forma, uma parte significativa delas encontra-se nas áreas menos desenvolvidas dos estados, e também compõe-se de escolas menores e com baixas médias de índice socioeconômico de seus alunos.

*Tabela 4.23: As 30 escolas com os menores índices totais de desenvolvimento (status e progresso combinados) na série histórica do SIMAVE (2000 – 2011)*

POS.	ÍND_TOTAL	ESCOLA	REDE	MUNICÍPIO	N_AL.	ISE
-1	2	EM SUSSUARANA I	M	JANUARIA	18	-1,6
-2	2,6	EE MARIA DA GLORIA ASSUNCAO	E	RIBEIRAO DAS NEVES	72	0
-3	2,7	EE JOAO MARIANO RIBEIRO	E	CONC. MATO DENTRO	15	-0,7
-4	2,7	EM DE AREIAO	M	JANUARIA	25	-1,2
-5	2,9	EM DE VARGINHA	M	JANUARIA	14	-1,4
-6	3	EM FRANCISCO SOARES	M	PATIS	34	-0,8
-7	3	EM TOMAS GONZAGA	M	MIRAVANIA	23	-0,9
-8	3	EE CESARIO NUNES DOS SANTOS	E	BONITO DE MINAS	19	-1,6
-9	3	EM JOAO LOPES CORREA	M	CON. MARINHO	15	-1,2
-10	3,1	EE RUI BARBOSA	E	RUBELITA	16	-1,3
-11	3,2	EM OSORIO ALEIXO DA SILVA	M	BETIM	83	0,3
-12	3,2	EM FLORESTAN FERNANDES	M	BELO HORIZONTE	81	0,2
-13	3,2	EM IRMA MARIA GEMA	M	ARACUAI	20	-1,6
-14	3,2	EE MANOEL PEREIRA MAGALHAES	E	BONITO DE MINAS	18	-1,1
-15	3,2	EM SEBASTIAO FERREIRA DE OLIVEIRA	M	BETIM	33	0,1
-16	3,3	EE ANALIA CARNEIRO DOS SANTOS	E	BURITIS	18	-0,3
-17	3,3	EE REPARATA DIAS DE OLIVEIRA	E	LAGOA SANTA	108	0,1
-18	3,3	EM SALGADO FILHO	M	BELO HORIZONTE	225	0,7
-19	3,3	EM DE BREJO GRANDE	M	S ANT DO RETIRO	25	-1,2
-20	3,4	EM JORNALISTA BICALHO BRANDAO	M	VERDELANDIA	22	-1,1

-21	3,4	EM N SRA DE FATIMA	M	UNAI	37	-0,5
-22	3,4	EM FERNANDO DIAS COSTA	M	BELO HORIZONTE	105	0
-23	3,4	EM WLADIMIR DE P. GOMES	M	BELO HORIZONTE	135	0,3
-24	3,4	EE IRMA ARCANGELA	E	TEOFILO OTONI	12	-0,2
-25	3,4	EE IONE SILVEIRA MENDES	E	MATO VERDE	73	-0,6
-26	3,4	EM PREF OSW. PIERUCCETTI	M	BELO HORIZONTE	43	0,1
-27	3,4	EM SAO VICENTE	M	ARACUAI	12	-1,1
-28	3,4	EM PRES RAUL SOARES	M	BETIM	162	0,5
-29	3,5	EM DE ARACA	M	JANUARIA	13	-1,3
-30	3,5	EE GILBERTO ALVES COUTINHO	E	VARZELANDIA	18	-0,7

#### 4.7. Uma análise de cluster dos extremos de desempenho

As duas tabelas a seguir apresentam as coordenadas dos centroides dos clusters formados por dois conjuntos de escolas: primeiro, o grupo de escolas com índices totais de desenvolvimento correspondentes, pelo menos, ao 95º percentil da distribuição; e segundo, o das escolas com índices totais iguais ou menores do que o do 5º percentil.

Em ambos os casos, é possível perceber a existência de dois grupos de escolas: um grupo mais numeroso, formado de estabelecimentos menores – com uma ou duas turmas de alunos na série considerada – e um grupo menor, porém constituído de escolas bem maiores.

No caso das escolas de maior desempenho, ambos esses grupos (mostrados na primeira tabela a seguir) são consideravelmente menores do que suas contrapartes de menor desempenho (que aparecem na segunda tabela). Isso novamente reforça a ideia de que o maior tamanho da escola parece funcionar como uma espécie de obstáculo tanto à conservação de um status elevado, quanto à conservação de taxas mais substanciais de progresso (muito embora também seja preciso reconhecer que apenas um tamanho pequeno não seja uma garantia de qualidade educacional, como os dados mostrados vêm inequivocamente demonstrando).

*Tabela 4.24: Coordenadas dos centroides dos dois clusters nos quais se dividem os 5% de escolas de menor índice de progresso na série histórica do SIMAVE (2000 – 2011).*

	Cluster	
	1	2
<i>Número de escolas</i>	41	249
<i>Nº médio de alunos avaliados</i>	130,1	53,0
<i>Média do ISE</i>	0,5	0,0
<i>Média_status (0 a 10)</i>	7,1	6,9
<i>Média_progresso (0 a 10)</i>	5,4	5,8

Tabela 4.25: Coordenadas dos centroides dos dois clusters nos quais se dividem os 5% de escolas de menor índice de progresso na série histórica do SIMAVE (2000 – 2011).

	Cluster	
	1	2
Número de escolas	11	211
Nº médio de alunos avaliados	209,1	67,1
Média do ISE	0,3	-0,3
Média_status (0 a 10)	5,1	4,7
Média_progresso (0 a 10)	2,5	3,0

#### 4.8. Alguns passos adicionais a serem considerados na investigação da qualidade educacional

O presente estudo das escolas do SIMAVE possibilitou o cálculo de índices de status e de progresso, bem como de uma combinação entre eles, que teve o mérito de identificar casos de valores consistentes de proficiência média e/ou de avanço da mesma ao longo dos anos.

Entretanto, ainda assim, alguns pontos ainda necessitam ser abordados mais detalhadamente, como, por exemplo:

##### *A inclusão de mais informações*

No procedimento que se acaba de descrever, foram analisados, para se ter maior clareza e simplicidade, apenas os resultados de duas disciplinas acadêmicas – português e matemática – e apenas uma série específica da escolarização: a 8ª série ou 9º ano do Ensino Fundamental. Entretanto, ao menos a princípio, seria interessante observar como ficariam os resultados gerais das escolas, caso estes fossem calculados levando-se em conta também outras variáveis, como, por exemplo, os resultados de outras disciplinas e/ou de outras séries testadas. Por um lado, isso teria a vantagem de aumentar o número de informações que servem de base para as avaliações, o que aumenta a fidedignidade das medidas. Além disso, tal prática também seria desejável porque poderia promover a inclusão do trabalho de outros profissionais – como o dos professores de outras matérias e séries – no cálculo da qualidade da escola.

Entretanto, por outro lado, quanto mais se acrescentam novas variáveis a um dado índice de qualidade, mais diluída fica a atribuição dos resultados à ação dos profissionais envolvidos. Alcançar, portanto, um meio termo entre esses dois extremos parece ser um desafio considerável para um sistema de avaliação.

### *A implementação de uma série histórica a partir de poucas avaliações*

O estudo aqui realizado beneficiou-se de uma vantagem que muitos, ou mesmo a grande maioria dos sistemas de avaliação educacional atualmente vigentes no Brasil, ainda não possuem: uma série histórica com mais de uma década de avaliações censitárias e colocadas em uma mesma métrica por meio de procedimentos de equalização vertical dos resultados dos testes.

Entretanto, o que fazer quando o sistema dispõe, digamos, de somente uma ou duas ocasiões de testagem? Como construir tipologias de resultados com base em um número tão pequeno de evidências? E, o que talvez seja ainda mais grave e premente: como atribuir consequências às avaliações, diante da escassez de dados de uma série histórica ainda incipiente?

### *O estabelecimento de metas de desempenho*

Os dois problemas apontados anteriormente – a inclusão de novas informações e também o devido tratamento a ser dado aos casos com poucos dados disponíveis – também precisam ser articulados à questão de como traçar metas de desempenho que sejam, ao mesmo tempo, exigentes e factíveis, para que as escolas as atinjam. E tal desafio necessariamente precisa levar em conta vários pontos de grande importância como, por exemplo: (1) o que vem a ser um comportamento adequado; (2) como tratar a volatilidade dos resultados; (3) como equilibrar as exigências de melhorias de um lado e uma certa flexibilização das metas de outro, etc.

Estas e outras questões deverão, portanto, ser abordadas mais detalhadamente na sequência deste texto.

## **5. O ÍNDICE DE DESENVOLVIMENTO DA EDUCAÇÃO BÁSICA (IDEB): INTERPRETAÇÃO, METAS E RESULTADOS**

### **5.1. INTRODUÇÃO**

No capítulo anterior, discorremos sobre o status e o progresso da proficiência das escolas públicas mineiras, com os resultados calculados de modo a cobrir uma série relativamente extensa de edições do teste do SIMAVE. Neste estágio do presente estudo, continuamos interessados especificamente no tema da qualidade educacional. Entretanto, a partir de agora, também pretendemos trazer para essa discussão o tema da mensuração da qualidade das escolas segundo aferida por uma medida que vem adquirindo uma enorme relevância em termos nacionais: o Índice de Desenvolvimento da Educação Básica, ou Ideb, conforme já se comentou em outras partes deste texto. Isso faz com que, neste momento, torne-se adequado discorrer um pouco mais detalhadamente sobre esse índice, para que, posteriormente, abordemos as propostas de aperfeiçoar suas respectivas metas.

Um ponto que cabe mencionar é que a nova metodologia de estabelecimento de metas que se apresentará no presente trabalho, embora esteja, neste caso, especificamente aplicada ao Ideb nas escolas públicas de Minas Gerais, ela, na verdade, pode e deve ser vista como um procedimento mais geral, que, a princípio, poder-se-ia empregar, digamos, em outros estados brasileiros, particularmente naqueles que, à semelhança de Minas Gerais, vêm experimentando sistemas próprios de avaliação educacional ao longo de vários anos, como o SARESP em São Paulo, o SPAECE no Ceará, o SAERJ, no Rio de Janeiro e vários outros. Mas, além disso, igualmente é possível que essa metodologia também se empregue até mesmo nos raros e eventuais casos de estados brasileiros que até hoje porventura ainda não tenham outras avaliações censitárias de proficiência dos alunos que não os testes da Prova Brasil. E, generalizando ainda mais o contexto de aplicação dessa metodologia, ela também poderia, naturalmente com as necessárias adaptações, ser estendida para outras possíveis medidas de avaliação educacional, particularmente as que, da mesma forma que o Ideb, também produzem um índice resultante da combinação de dois fatores extremamente relevantes para a qualidade educacional: o rendimento de uma determinada unidade, dado pelo seu respectivo fluxo de alunos ao longo das séries consideradas, e também a proficiência, mensurada pelos testes em larga escala de disciplinas específicas.

Dessa forma, para avançar um pouco mais nesta questão, passemos agora a discorrer sobre alguns detalhes relevantes envolvendo a concepção do Ideb e suas respectivas metas.

## 5.2. CONSIDERAÇÕES GERAIS SOBRE O IDEB

Em nota técnica divulgada pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP – órgão do Ministério da Educação do Brasil – MEC –, sobre o Índice de Desenvolvimento da Educação Básica, é dito que:

O Ideb é um indicador de qualidade educacional que combina informações de desempenho em exames padronizados (Prova Brasil ou Saeb) – obtido pelos estudantes ao final das etapas de ensino (4ª e 8ª séries do ensino fundamental e 3ª série do ensino médio – com informações sobre rendimento escolar (aprovação). (INEP, 2010)

Desde a sua criação em 2005, o Ideb vem adquirindo uma significância cada vez maior no cenário educacional brasileiro, devido a um conjunto de fatores relevantes. Um deles reside na sua capacidade de combinar, numa única medida, duas informações de caráter crucial para a análise e o monitoramento da qualidade da educação oferecida pelas escolas brasileiras: o fluxo dos alunos através das sucessivas séries ou anos da educação básica – relacionado, naturalmente, às taxas de aprovação – e o desempenho acadêmico dos alunos conforme mensurado por avaliações externas e em grande escala realizadas pelo MEC, na forma do Sistema de Avaliação da Educação Básica – SAEB – e da Prova Brasil.<sup>4</sup>

Outra importância do Ideb encontra-se no fato de que, com ele, metas de desempenho têm sido traçadas pelo MEC para as diversas unidades da educação pública brasileira, que abrangem desde níveis mais elevados de agregação de dados, como os estados da federação, até os vários milhares de escolas públicas – estaduais e municipais – espalhadas por todo o país. Portanto, a cada edição do SAEB e da Prova Brasil, são calculados os valores do Ideb alcançados por essas unidades educacionais, que são então comparados com as metas que lhes haviam sido traçadas. E os resultados dessas comparações – os quais, a princípio, podem apontar desde casos felizes, em que as metas são cumpridas ou mesmo superadas, até aqueles mais problemáticos, em que elas não são atingidas – têm interessado a um público bastante

---

<sup>4</sup> Esses sistemas federais de avaliação educacional, na realidade, mudaram de nome após a concepção original do Ideb. O SAEB passou a ser mais recentemente conhecido como Avaliação Nacional da Educação Básica – ANEB –, ao passo que a Prova Brasil foi rebatizada como Avaliação Nacional do Rendimento Escolar, ou ANRESC. Entretanto, nenhum desses novos nomes parece ter vingado no uso popular, e mesmo técnico, de modo que este texto empregará a nomenclatura original e de uso mais difundido dessas duas avaliações.

amplo e variado, dentro do qual se incluem autoridades e tomadores de decisão, profissionais da educação, alunos e seus pais e órgãos públicos e privados da imprensa, entre outros.

### 5.2.1. Cálculo e interpretação do Ideb

O Ideb, em um ano específico e para uma determinada unidade educacional (como uma escola, município, estado, região ou mesmo para o país como um todo), é calculado, conforme já se disse no Capítulo 3, da seguinte forma:

$$\text{Ideb} = N \times P$$

onde:

$N$  = média da proficiência combinada de Língua Portuguesa e Matemática, padronizada numa escala que varia de 0 a 10;

$P$  = indicador de rendimento, baseado na taxa de aprovação da etapa de ensino considerada, e que corresponde a um número variando de 0 a 1.

A média de proficiência  $N$  é, *grosso modo*, uma simples comparação entre o desempenho da unidade educacional de interesse e a média nacional do Saeb 1997, ano em que se introduziram as escalas de proficiência em Língua Portuguesa e Matemática. Se, digamos, a média de uma escola num ano específico – como, por exemplo, 2011 – foi exatamente igual à média nacional do Saeb 1997 em ambas essas disciplinas, então o  $N$  dessa escola em 2011 corresponderá a 5, valor situado exatamente na metade da escala de  $N$ , que, conforme vimos, varia entre 0 e 10. Desse modo, médias acima da referência (ou seja, maiores que a média do Saeb 1997) correspondem a valores superiores a 5, enquanto que médias abaixo da referência correspondem a valores inferiores a 5. Diante disso, é desejável que, com o tempo, as unidades educacionais consigam obter valores cada vez maiores e superiores a 5, indicando, assim, que seu desempenho médio estaria aumentando progressivamente em relação à média nacional de 1997 em ambas as disciplinas avaliadas.

Entretanto, como se sabe, as notas de português e matemática possuem escalas diferentes, fazendo com que se faça necessário transformar os escores obtidos nas diferentes disciplinas em uma espécie de resultado padronizado. Tal efeito, no caso do Ideb, é obtido por meio das seguintes transformações:

$$n(\alpha) = 10 \cdot (S - S_{\min}) / (S_{\max} - S_{\min})$$

$$N = (n(\text{port}) + n(\text{mat})) / 2$$

Neste caso,  $n(\alpha)$  é a nota padronizada obtida por uma unidade educacional qualquer, como uma escola, numa disciplina  $\alpha$ , que pode ser Língua Portuguesa ou Matemática. Por sua vez, esse valor se calcula a partir da média  $S$  obtida por essa mesma unidade numa determinada edição do SAEB ou da Prova Brasil. E  $S_{\min}$  e  $S_{\max}$  correspondem, respectivamente, aos valores teóricos mínimo e máximo que uma determinada unidade educacional pode alcançar num teste qualquer de uma dessas avaliações. Para a determinação desses valores extremos, tomou-se como base o comportamento das escolas no SAEB de 1997, quando se arbitraram valores fixos para a média e o desvio-padrão dos resultados em nível nacional.

Dessa forma, no caso, por exemplo, da 8ª série ou 9º ano do Ensino Fundamental, que interessa particularmente a este estudo, aquela edição do SAEB teve uma média de 250 pontos e um desvio-padrão de 50 pontos, tanto na escala de proficiência de Língua Portuguesa quanto na de Matemática. E assim, os valores extremos foram determinados como correspondendo a três desvios-padrão abaixo ou acima dessa média, o que resultou em  $S_{\min} = 100 (= 250 - 3 \cdot 50)$  e  $S_{\max} = 400 (= 250 + 3 \cdot 50)$ .

Por sua vez,  $P$ , o outro fator que entra no cálculo do Ideb, é um número variando entre 0 e 1 que mensura o fluxo de alunos, sendo que, quanto mais próximo do ideal (100% de aprovação) estiver o fluxo, maior é o valor de  $P$ , ou seja, mais próximo de 1 ele ficará. Dessa forma, por exemplo:

100% de aprovação;  $P = 1$

80% de aprovação;  $P = 0,8$

Na prática,  $P$  quase nunca chega a atingir os extremos de 1 e 0, visto que são muito raros os casos de 100% de aprovação e virtualmente inexistentes os casos de 0% de aprovação.

Um ponto muito importante a respeito do fluxo dos alunos é que ele é mensurado como referente ao final de um determinado ciclo de escolarização (digamos, ao final da 4ª

série EF<sup>5</sup>). Entretanto, precisamente devido a isso, ele *não* é determinado apenas pelo rendimento dos alunos na última série ou ano do referido ciclo, mas, antes, também pelo rendimento observado em todas as séries anteriores que compõem o ciclo. Dessa forma:

para a 4ª série (5º ano) EF, consideram-se cinco anos de fluxo, correspondentes à atual série inicial e aos anos de 2 a 5 (ou antigas séries de 1 a 4) do ensino fundamental;

para a 8ª série (9º ano) EF, consideram-se as séries de 5 a 8 (do 6º ao 9º ano) EF, e finalmente,

para a 3ª série EM, consideram-se as três séries do ensino médio.

### *As Metas do Ideb*

O Ideb foi concebido não somente para fornecer uma medida da qualidade da educação num ano específico, mas também para servir a outro propósito, de caráter mais estratégico e de prazo mais dilatado: com ele, é possível traçar uma expectativa de evolução temporal das unidades educacionais, como um modo de contribuir para que, no longo prazo, a educação pública no Brasil idealmente atinja um desenvolvimento comparável ao dos chamados países desenvolvidos. Nesse sentido, pensou-se o Ideb como um indicador passível de sofrer evoluções contínuas ao longo das diversas edições do Censo Escolar, e também do SAEB ou da Prova Brasil. E, para que esse objetivo se atinja, o MEC traçou metas do Ideb para as diversas unidades educacionais brasileiras, indo desde as instâncias mais elevadas, como os estados, por exemplo, até as unidades educacionais menores e mais estreitamente associadas ao dia-a-dia educacional, como as escolas.

Um ponto de capital importância para esse estabelecimento de metas foi a realização de um estudo comparativo internacional envolvendo, de um lado o Brasil e, de outro, os países da Organização para a Cooperação e o Desenvolvimento Econômicos – OCDE –, que congrega a maior parte das economias mais avançadas do mundo, como os Estados Unidos, Austrália, Japão e os países da Europa Ocidental. Nesse estudo comparativo, estimou-se que os países da OCDE, segundo suas estatísticas de fluxo escolar e os resultados dos testes de 2003 do Programa Internacional de Avaliação de Alunos – o PISA, do qual o Brasil também

---

<sup>5</sup> Embora, atualmente, expressões como “4ª série do ensino fundamental” estejam em desuso em alguns contextos relevantes, que preferem falar do “5º ano” dessa mesma etapa da escolarização, essa nova nomenclatura ainda não havia prevalecido por ocasião dos trabalhos pioneiros sobre o Ideb, de modo que muitos textos oficiais sobre esse tema têm adotado a nomenclatura antiga.

vem participando como país convidado –, teriam, naquela avaliação, um “Ideb” igual a aproximadamente 6 pontos, numa escala que, conforme se viu, varia de 0 a 10<sup>6</sup>. (INEP, 2012c)

Portanto, um objetivo que se fixou para o Brasil a médio prazo foi a obtenção de um Ideb médio também igual a 6 pontos (a média dos países mais desenvolvidos), algo que, segundo os procedimentos de determinação das metas, deverá, se tudo correr conforme o esperado, ser atingido até o final de 2022, ano de especial simbolismo para o país, visto corresponder aos duzentos anos da independência brasileira, conforme já se comentou anteriormente neste texto. Entretanto, atingir um Ideb de 6 pontos não é um fim em si mesmo, mas, antes, deve corresponder à superação de uma etapa focada num objetivo ainda mais ambicioso, que é fazer com que o Brasil atinja, até próximo ao final do século 21, em 2096, um Ideb médio de 10 pontos. Caso isso seja, de fato, alcançado, a educação básica no Brasil ter-se-á alçado a um patamar extremamente favorável, pelo qual a retenção escolar atingiria níveis desprezíveis no país, ao mesmo tempo em que as notas dos alunos das escolas nos exames nacionais padronizados do SAEB e da Prova Brasil atingiriam níveis, em média, enormemente altos, quando comparados com os resultados médios do país no SAEB de 1997, que serviram de base para o cálculo desse indicador.

O modelo logístico matemático utilizado para traçar essas metas é feito nos moldes das regressões convencionais desse tipo, e que podem, por exemplo, ser encontradas em Aldrich (1993), Cromwell (1994) and Hagenars (1993). Basicamente, ele permite que se calcule uma família de curvas, com cada uma correspondendo a uma unidade educacional de interesse, como um estado ou uma escola pública brasileira, por exemplo, convergindo todas elas para um ponto em comum nesse futuro distante (2096), conforme se vê na ilustração a seguir.

---

<sup>6</sup> Os argumentos utilizados para defender a ideia de que a OCDE tem, em média, um “Ideb” de 6 pontos são desenvolvidos numa nota técnica do INEP, órgão responsável pela realização desse estudo comparativo, que pode ser encontrada nas referências bibliográficas do presente texto (INEP, 2012).

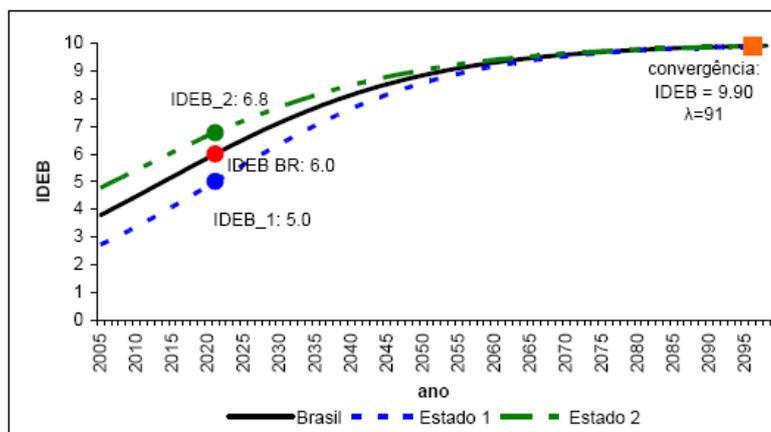


Gráfico 5.1: Trajetória do Ideb para o Brasil e Estados: 2005 e 2096 – 1ª fase do Ensino Fundamental. (INEP, 2012)

O Gráfico 5.1 tem algumas características importantes que cabe aqui comentar. Uma delas é que, conforme se pode perceber pela existência das diferentes curvas mostradas, a trajetória das metas do Ideb não é a mesma para todas as unidades educacionais. Antes, elas dependem, na verdade, dos estágios iniciais de desenvolvimento educacional em que se encontravam essas diferentes unidades. Por exemplo, uma unidade educacional que, no período inicial aqui considerado (ano de 2005) encontrava-se abaixo da média nacional, teria toda a sua trajetória de metas previstas ou estabelecidas para valores também abaixo dessa média nacional no futuro, e o oposto também é verdade, ou seja, as unidades educacionais com valores iniciais acima da média nacional teriam suas metas correspondentemente acima das projeções médias futuras para todo o país.

Dessa forma, as metas diferem, por exemplo, de estado para estado, e de escola para escola, e isso serve para, ao menos, atenuar um pouco a crítica que se costuma fazer ao Ideb, de que ele coloca num mesmo pé de igualdade unidades educacionais inseridas em meios com diferentes índices socioeconômicos médios. A atenuação a essa crítica, naturalmente, decorre do fato de que, unidades com piores desempenhos iniciais têm também, inicialmente, metas mais modestas de evolução, de modo que isso compensa, de certa maneira, as desvantagens por elas vividas no início do período inaugurado em 2005, quando se começou a calcular o Ideb.

Entretanto, outro ponto relevante e possível de ser percebido no Gráfico 5.1 acima, reside no fato de que as metas, com o tempo, tendem a convergir, o que, graficamente, se

traduz numa aproximação das curvas entre si à medida que o tempo avança ao longo do período considerado. Ou seja, espera-se que os esforços empreendidos em prol do desenvolvimento da educação sejam capazes de, no longo prazo, reduzir a desigualdade educacional, fazendo com que, no final, todas as unidades educacionais brasileiras passem a ter um Ideb cada vez mais semelhante entre si, e também mais elevado.

Pois bem, tendo visto como se traçam e funcionam, teoricamente, as metas do Ideb, passemos agora a observar como, de fato, esse índice tem se comportado nacionalmente desde o início da sua série histórica até a sua edição mais recente, correspondente a 2013. E, após essas considerações sobre o Ideb em relação ao país como um todo, passaremos também a apresentar e discutir alguns de seus principais resultados relativos a Minas Gerais, estado que nos interessa particularmente no presente estudo.

### 5.3. O COMPORTAMENTO DO IDEB NO BRASIL

Antes de nos determos especificamente no comportamento do Ideb nas escolas públicas de Minas Gerais, faremos uma análise preliminar da forma como esse índice tem se comportado em termos nacionais. Para isso, consideraremos toda a série histórica do Ideb disponível até o presente momento, englobando as avaliações bianuais entre 2005 e 2013, inclusive. E, neste sentido, esses resultados nacionais são apresentados levando-se em conta tanto os seus valores absolutos como também os referentes à sua evolução em direção às suas metas iniciais, traçadas pelo MEC.

Em geral, observam-se algumas características marcantes nos dados nacionais do Ideb, das quais, pelo menos três, merecem um comentário particular: (1) uma acentuada discrepância de resultados entre as redes pública e particular, favorável a esta última; (2) a ocorrência de um maior progresso na rede pública, paralelamente à existência de um certo efeito-teto na rede particular; (3) as quedas dos índices associadas ao avanço do nível de escolaridade, fazendo com que, apenas para o 5º ano do Ensino Fundamental, parece haver uma tendência real de cumprimento das metas para 2021. Passemos, então, a investigar com um pouco mais de detalhe cada uma dessas características.

### 5.3.1. A evolução do Ideb nacional entre 2005 e 2013

A tabela a seguir apresenta os valores reais do Ideb em termos nacionais por rede de ensino, para cada ano de sua série histórica, bem como suas respectivas metas projetadas pelo MEC, e também as diferenças entre aquilo que se previu e o que se alcançou. A série termina com os valores do ano de 2013, que, conforme já se especificou, são os mais atuais disponíveis no momento de realização deste trabalho. Por outro lado, cabe observar que, naturalmente, não há metas previstas para 2005, que foi o primeiro ano desta série histórica.

Um ponto que logo se observa, ao se examinar um pouco mais atentamente esta tabela, é uma considerável superioridade dos valores dos índices da rede particular em relação aos da rede pública, e também ao total das redes. Por outro lado, essa proximidade entre os valores totais e os da rede pública, naturalmente se deve ao fato de que esta última concentra a grande maioria da população atendida pela escolarização básica no Brasil. Já em relação às diferenças entre o desempenho real e as metas projetadas, é possível perceber alguns pontos de interesse, que se manifestam quando se faz uma análise ao longo das redes de ensino e também das séries de escolarização.

Em termos das séries ou anos de escolarização – 5, 9 ou 12 –, salta aos olhos o fato de que o desempenho geral diminui consideravelmente com o avanço das séries. Ou, dito de outro modo, o desempenho nacional é maior no 5º ano do Ensino Fundamental, onde as diferenças entre os valores reais do Ideb e suas respectivas metas projetadas tendem a assumir valores significativamente maiores do que zero, indicando que tais metas estão sendo atingidas ou superadas. Já no 9º ano EF, embora as diferenças continuem positivas, elas decrescem consideravelmente, até diminuírem ainda mais no final do Ensino Médio.

Já do ponto de vista das redes de ensino, observa-se a ocorrência de um progresso maior nas escolas públicas – e, por conseguinte, também no total – em comparação com a rede particular. Muito provavelmente, ocorre aqui um certo efeito-teto, pelo qual fica consideravelmente mais difícil para a rede particular crescer no mesmo ritmo da pública. E tanto isso é verdade que a rede particular somente conseguiu superar suas respectivas metas para as projeções iniciais em 2007 e, ocasionalmente, em 2009. Depois disso, seu desempenho tem, geral e sistematicamente, ficado aquém do esperado. Por outro lado, a rede pública e a total têm experimentado avanços mais significativos, muito embora sua situação atual não seja muito confortável, devido ao fato de que, no início de suas respectivas séries, os valores de partida eram consideravelmente baixos.

Tabela 5.1: Valores reais e metas projetadas do Ideb nacional por rede de ensino e série, e as diferenças entre eles

SÉRIE	ANO	PARTICULAR			PÚBLICA			TOTAL		
		REAL	META	DIF	REAL	META	DIF	REAL	META	DIF
5 EF	2005	5,9			3,6			3,8		
	2007	6,0	6,0	0,0	4,0	3,6	0,4	4,2	3,9	0,3
	2009	6,4	6,3	0,1	4,4	4,0	0,4	4,6	4,2	0,4
	2011	6,5	6,6	-0,1	4,7	4,4	0,3	5,0	4,6	0,4
	2013	6,7	6,8	-0,1	4,9	4,7	0,2	5,2	4,9	0,3
9 EF	2005	5,8			3,2			3,5		
	2007	5,8	5,8	0,0	3,5	3,3	0,2	3,8	3,5	0,3
	2009	5,9	6,0	-0,1	3,7	3,4	0,3	4,0	3,7	0,3
	2011	6,0	6,2	-0,2	3,9	3,7	0,2	4,1	3,9	0,2
	2013	5,9	6,5	-0,6	4,0	4,1	-0,1	4,2	4,4	-0,2
3 EM	2005	5,6			3,0			3,4		
	2007	5,6	5,6	0,0	3,2	3,1	0,1	3,5	3,4	0,1
	2009	5,6	5,7	-0,1	3,4	3,2	0,2	3,6	3,5	0,1
	2011	5,7	5,8	-0,1	3,4	3,3	0,1	3,7	3,7	0,0
	2013	5,4	6,0	-0,6	3,4	3,6	-0,2	3,7	3,9	-0,2

Fonte: MEC, 2014

O que acaba de se comentar sobre os valores reais e projetados do Ideb pode ser ilustrado nos três gráficos que se seguem, que representam cada uma das séries estudadas. No primeiro deles, referente ao 5º ano EF, é possível visualizar tanto o aumento progressivo dos valores desse índice com o progresso da série temporal, como também o fato de que os valores reais tendem a ser maiores do que suas respectivas metas projetadas. Este último ponto se traduz nas menores alturas das colunas de número par – 2, 4 e 6, correspondentes às metas projetadas –, que tendem a ser sistematicamente mais baixas do que as colunas de número ímpar imediatamente anteriores: 1, 3 e 5 –, as quais correspondem aos respectivos valores reais obtidos pelas diferentes redes.

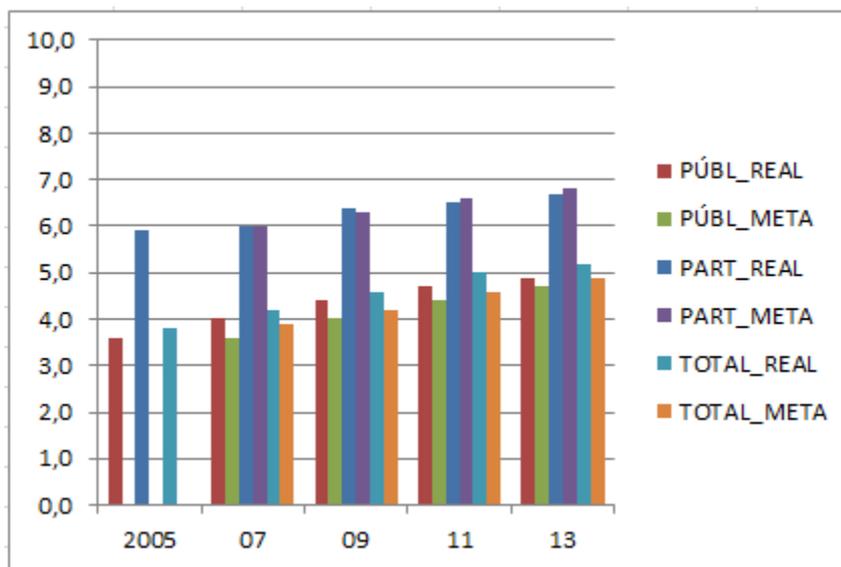


Gráfico 5.2: Valores reais e metas projetadas do Ideb nacional por rede de ensino, e as diferenças entre eles, para o 5º ano EF.  
Fonte: MEC, 2014.

Já no gráfico que se segue, referente ao 9º EF, percebe-se que, embora um certo progresso temporal ainda exista, a tendência de elevação do Ideb é bem menos acentuada do que no caso anterior. Por exemplo, enquanto que antes, o Ideb total do 5º ano EF aproximava-se de 5 em 2013, agora, no 9º ano EF, ele se aproxima de 4. Além disso, também se observa que a tendência de crescimento do Ideb na rede particular é consideravelmente menor que na pública, conforme já se comentou anteriormente.

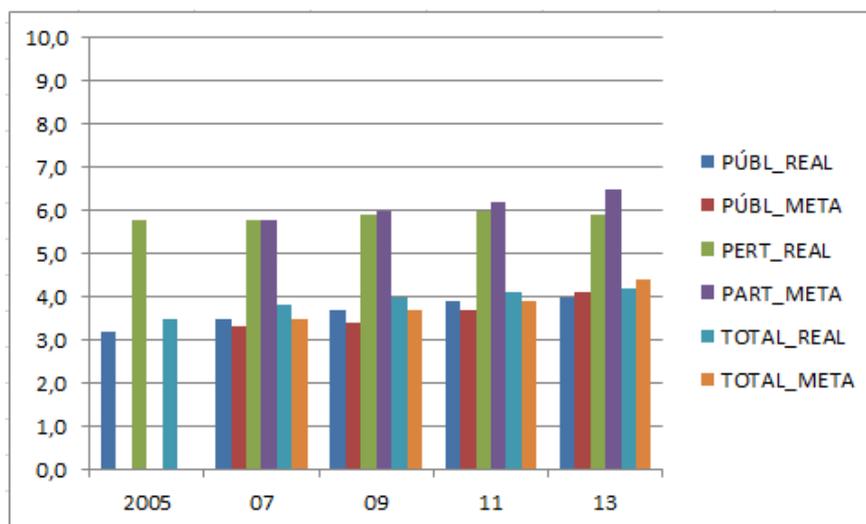


Gráfico 5.3: Valores reais e metas projetadas do Ideb nacional por rede de ensino, e as diferenças entre eles, para o 9º ano EF.  
Fonte: MEC, 2014.

Porém, é no final do Ensino Médio que os resultados tornam-se ainda mais desanimadores. Pode-se observar, no gráfico para esta etapa de escolarização, que as séries temporais, que já começam bem baixo em 2005, apresentam uma subida extremamente débil no caso das redes pública e total, fazendo com que, em 2013, último ano da série, seus valores ainda se encontrem na casa dos 3 pontos. E, na rede particular, o Ideb real praticamente se estaciona ao longo de toda a série, ocasionando um “descolamento” entre os valores reais e as respectivas metas projetadas, que se acentua em 2013, último ano desta série, e que provavelmente continuará ou mesmo aumentará futuramente, caso essa tendência se conserve.

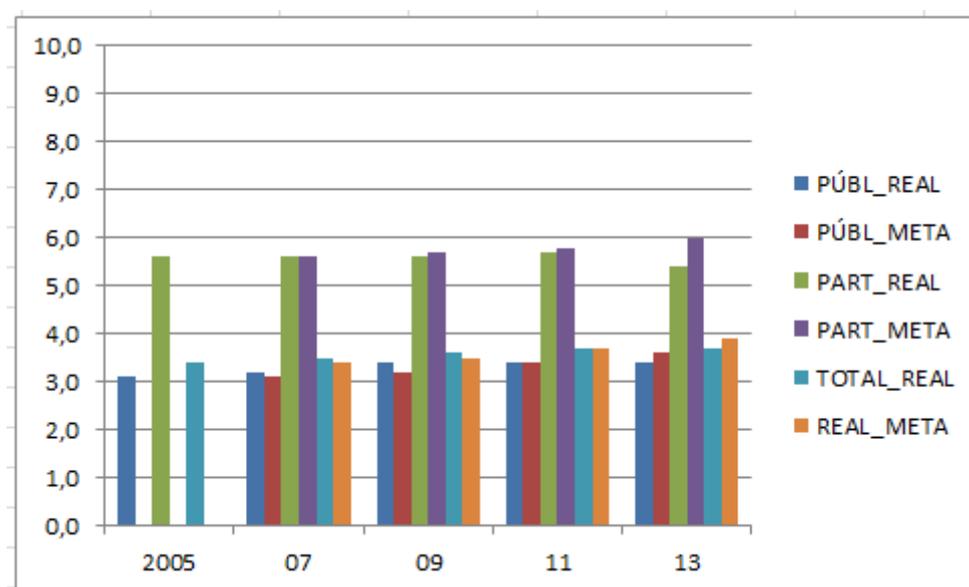


Gráfico 5.4: Valores reais e metas projetadas do Ideb nacional por rede de ensino, e as diferenças entre eles, para o 3º ano EM.  
Fonte: MEC, 2014.

### *A catástrofe do Ideb no bicentenário da Independência*

Como já se comentou, em 2022, o Brasil comemorará os duzentos anos de sua emancipação política de Portugal. Tal fato serviu, em 2005, como um componente simbólico de motivação para se traçarem metas para o Ideb de modo que o país alcançasse, no bicentenário da Independência, uma espécie de valor médio para esse índice de qualidade educacional que hipoteticamente seria encontrado em um país “desenvolvido” da OCDE, a Organização para a Cooperação e o Desenvolvimento Econômicos, um organismo internacional congregando algumas dezenas de países situados consideravelmente acima da média mundial em termos de desenvolvimento econômico, social e humano. E assim, chegou-

se à conclusão de que essa meta nacional brasileira do Ideb para 2021 deveria ser igual a seis pontos na escala do índice, que teoricamente varia entre 0 e 10.

Porém, caso se mantenham as tendências do Ideb até o bicentenário da independência do Brasil – por nós calculadas por meio de uma regressão linear clássica envolvendo os valores reais desse índice entre 2005 e 2013 –, o cenário educacional que o país viverá naquela ocasião provavelmente não dará motivos para grandes comemorações... Com efeito, é possível perceber, analisando o comportamento do Ideb ao longo da série histórica já existente, que apenas no 5º do Ensino Fundamental, os resultados são promissores. Nas demais séries, e também na rede particular de modo geral, as previsões são de uma certa “estagnação” do índice em questão.

No caso do Ensino Médio, por exemplo, observa-se que o Ideb da rede particular simplesmente não apresenta previsão de crescimento, fazendo com que o “descolamento” entre os valores previstos pela tendência real e aqueles projetados como metas só faz aumentar à medida que os anos passam. E, também no Ensino Médio no nível nacional, observa-se que, sem exceção, as previsões apontam apenas para casos em que as metas não conseguirão ser cumpridas, independentemente das redes. Também se observa que tais efeitos provavelmente se intensificarão já a partir de 2015 (cujos resultados serão divulgados em 2016), sendo que, em muitos casos, esse mesmo problema corresponde a uma continuação daquilo que, de fato, já ocorreu em 2013. A tabela e os três gráficos a seguir ilustram essas informações.

Tabela 5.2: Tendências de crescimento e metas projetadas do Ideb nacional por rede de ensino e série, e as diferenças entre elas.

SÉRIE	ANO	PARTICULAR			PÚBLICA			TOTAL		
		TEND	META	DIF	TEND	META	DIF	TEND	META	DIF
5 EF	2015	7,0	7,0	0,0	5,3	5,0	0,3	5,6	5,2	0,4
	2017	7,2	7,2	0,0	5,6	5,2	0,4	6,0	5,5	0,5
	2019	7,5	7,4	0,1	6,1	5,5	0,6	6,5	5,7	0,8
	2021	7,6	7,5	0,1	6,2	5,8	0,4	6,7	6,0	0,7
9 EF	2015	6,0	6,8	-0,8	4,2	4,5	-0,3	4,4	4,7	-0,4
	2017	6,0	7,0	-1,0	4,4	4,7	-0,3	4,5	5,0	-0,5
	2019	6,1	7,1	-1,0	4,6	5,0	-0,4	4,7	5,2	-0,5
	2021	6,1	7,3	-1,2	4,8	5,2	-0,4	4,9	5,5	-0,6
3 EM	2015	5,5	6,3	-0,8	3,5	3,9	-0,4	3,8	4,3	-0,5
	2017	5,4	6,7	-1,3	3,6	4,4	-0,8	3,9	4,7	-0,8
	2019	5,4	6,8	-1,4	3,7	4,6	-0,9	4,0	5,0	-1,0
	2021	5,4	7,0	-1,6	3,8	4,9	-1,1	4,0	5,2	-1,2

Fonte: MEC, 2014

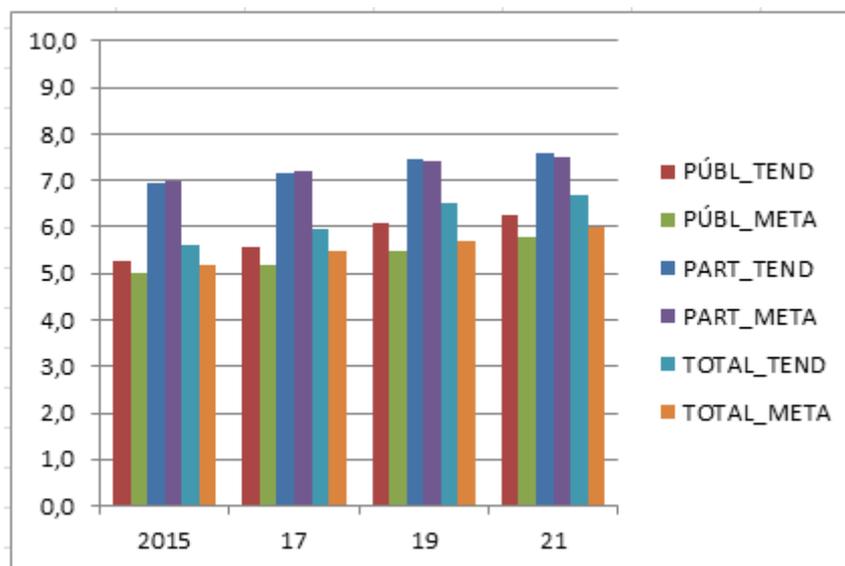


Gráfico 5.5: Tendências de crescimento e metas projetadas do Ideb nacional por rede de ensino, e as diferenças entre elas, para o 5º ano EF, para o período entre 2015 e 2021.

Fonte: MEC, 2014.

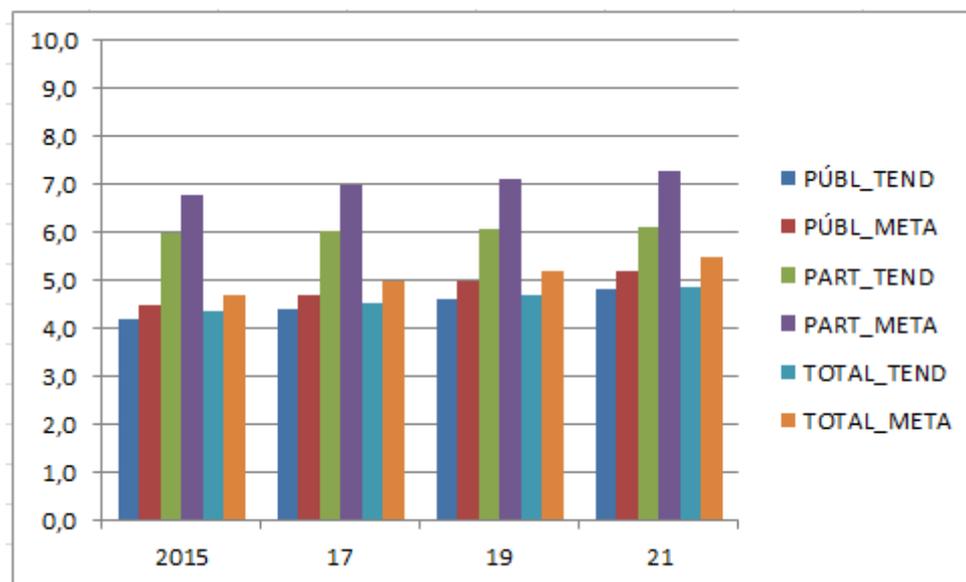


Gráfico 5.6: Tendências de crescimento e metas projetadas do Ideb nacional por rede de ensino, e as diferenças entre elas, para o 9º ano EF, para o período entre 2015 e 2021.

Fonte: MEC, 2014.

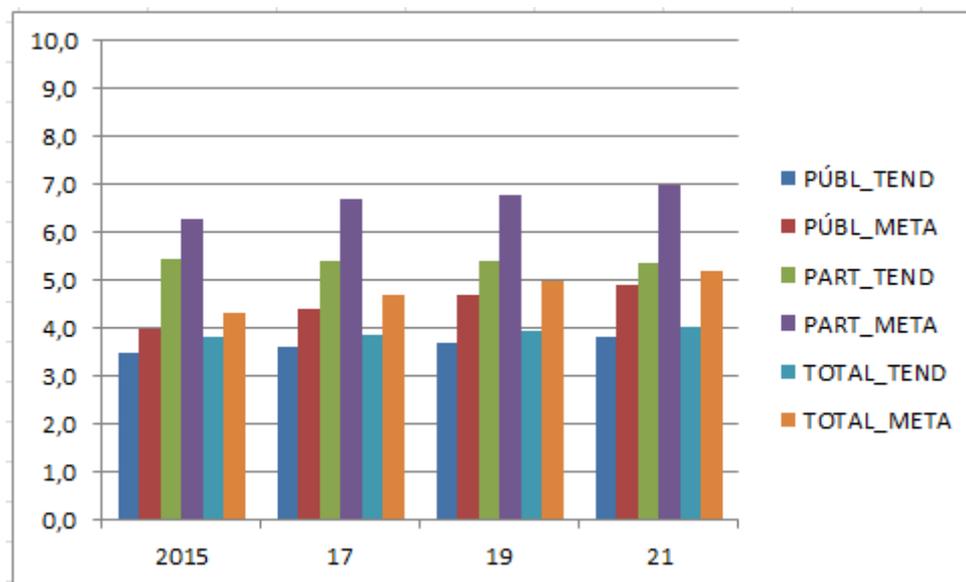


Gráfico 5.7: Tendências de crescimento e metas projetadas do Ideb nacional por rede de ensino, e as diferenças entre elas, para o 3º ano EF, para o período entre 2015 e 2021.

Fonte: MEC, 2014.

#### 5.4. O IDEB EM MINAS GERAIS

A tabela e os três gráficos a seguir apresentam dados referentes à evolução do Ideb por rede de ensino, ao longo de toda a sua série histórica disponível até o presente momento, especificamente para o estado de Minas Gerais.

Da mesma forma que se passou com o Brasil como um todo, também é possível perceber, no caso mineiro, muito do comportamento anteriormente observado para a situação nacional. Um desses aspectos é uma superioridade dos resultados da rede particular em relação à pública e à total, bem como um decréscimo do desempenho no Ideb à medida que se aumentam os níveis de escolarização, fazendo com que tais resultados sejam maiores no 5º ano EF e consideravelmente baixos no outro extremo, correspondente ao 3º ano EM.

Um aspecto de particular interesse no caso mineiro é o que ocorre com o 9º ano do Ensino Fundamental, particularmente no que diz respeito às diferenças entre as redes. No caso público, e também no total, percebe-se uma elevação do Ideb, que faz com que, para ambas

essas formas de agregação, os resultados delineiam um quadro “otimista”, pelo qual os valores reais do Ideb situam-se sistematicamente acima de suas metas, durante o período considerado (entre 2007 e 2013). Além disso, a tendência mostrada pelos gráficos também aponta para o fato de que, para o estado como um todo, e também especialmente para a rede pública, provavelmente as metas do Ideb do 9º ano EF serão atingidas em 2021. Entretanto, por outro lado, nada disso parece se verificar com relação à rede particular, onde não somente os valores reais têm sistematicamente ficado abaixo das metas, como também, nas avaliações futuras até 2021, as previsões são ainda mais pessimistas, sugerindo a possível ocorrência de uma crescente defasagem entre as metas projetadas e o desempenho real desta rede.

Tabela 5.3: Valores reais e metas projetadas do Ideb em MG por rede de ensino e série, e as diferenças entre eles.

SÉRIE	ANO	PARTICULAR			PÚBLICA			TOTAL		
		REAL	META	DIF	REAL	META	DIF	REAL	META	DIF
5 EF	2005	6,9			4,6			4,7		
	2007	6,7	7,0	-0,3	4,6	4,6	0,0	4,7	4,8	-0,1
	2009	7,1	7,2	-0,1	5,5	4,9	0,6	5,6	5,1	0,5
	2011	7,4	7,4	0,0	5,8	5,3	0,5	5,9	5,5	0,4
	2013	7,6	7,6	0,0	5,9	5,6	0,3	6,1	5,7	0,4
9 EF	2005	6,4			3,6			3,8		
	2007	6,4	6,5	-0,1	3,8	3,6	0,2	4,0	3,8	0,2
	2009	6,7	6,6	0,1	4,1	3,7	0,4	4,3	3,9	0,4
	2011	6,5	6,8	-0,3	4,4	4,0	0,4	4,6	4,2	0,4
	2013	6,3	7,0	-0,7	4,6	4,4	0,2	4,8	4,6	0,2
3 EM	2005	6,2			3,4			3,8		
	2007	5,7	6,2	-0,5	3,5	3,5	0,0	3,8	3,8	0,0
	2009	6,1	6,3	-0,2	3,6	3,6	0,0	3,9	3,9	0,0
	2011	6,1	6,4	-0,3	3,7	3,7	0,0	3,9	4,1	-0,2
	2013	5,8	6,6	-0,8	3,6	4,0	-0,4	3,8	4,3	-0,5

Fonte: MEC, 2014

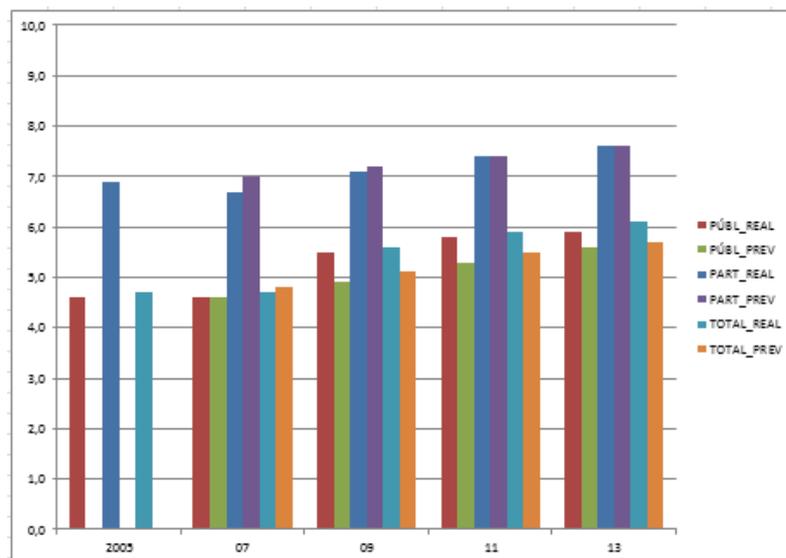


Gráfico 5.8: Valores reais e metas projetadas do Ideb de MG por rede de ensino, e as diferenças entre eles, para o 5º ano EF. Fonte: MEC, 2014.

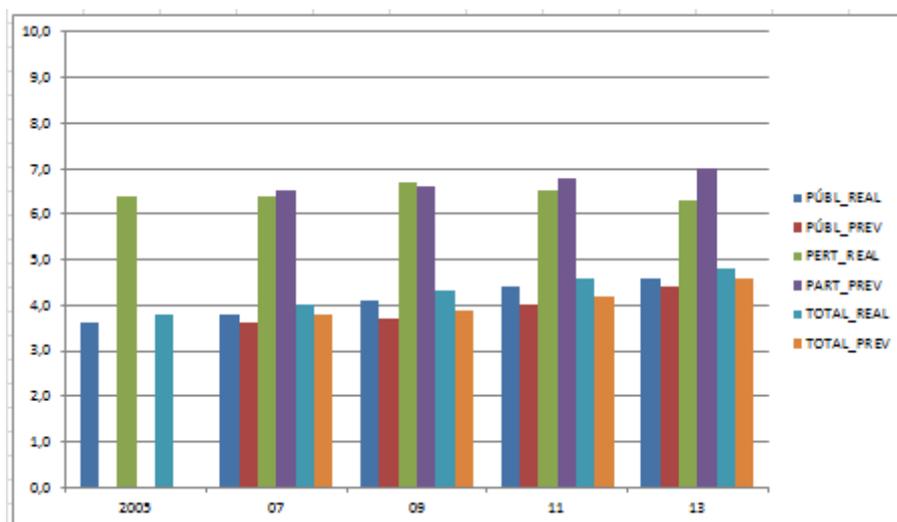


Gráfico 5.9: Valores reais e metas projetadas do Ideb de MG por rede de ensino, e as diferenças entre eles, para o 9º ano EF. Fonte: MEC, 2014.

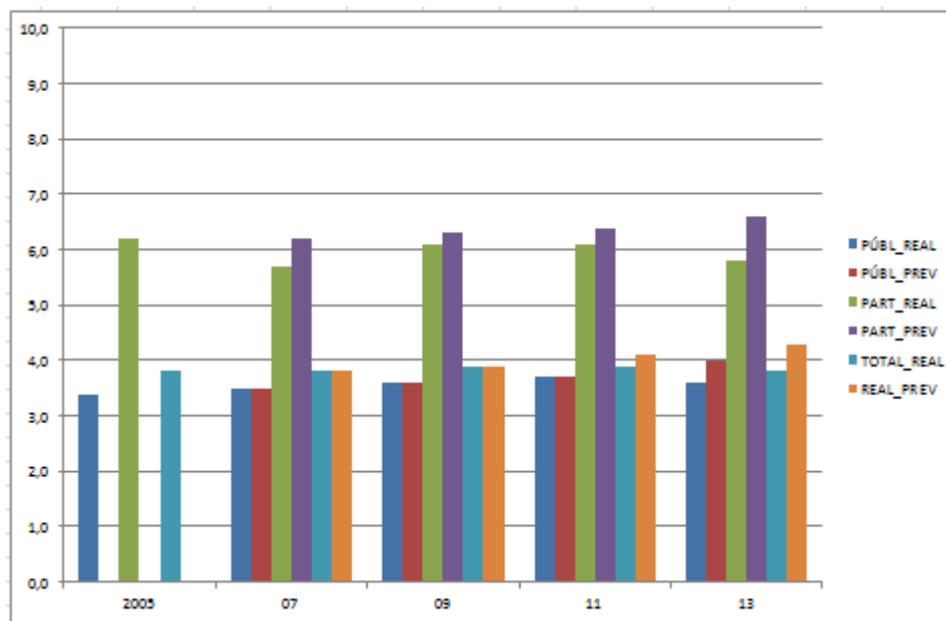


Gráfico 5.10: Valores reais e metas projetadas do Ideb de MG por rede de ensino, e as diferenças entre eles, para o 3º ano EM. Fonte: MEC, 2014.

#### 5.4.1. Perspectivas do cumprimento das metas do Ideb em Minas Gerais

A tabela e os três gráficos a seguir apresentam as metas projetadas para o Ideb em Minas Gerais num futuro próximo (entre 2015 e 2021), lado a lado com as tendências estimadas – igualmente por nós obtidas por meio de uma regressão linear clássica – para esses índices com base nas séries históricas já coletadas (entre 2005 e 2013). Analogamente ao caso geral brasileiro, também em Minas Gerais se observa um sucesso maior do Ideb no 5º ano EF, e igualmente um bom desempenho geral no 9º ano dessa mesma etapa (porém já não acompanhado de um resultado adequado na rede particular de ensino). E, da mesma forma que no caso nacional, um grande problema do Ideb mineiro relaciona-se ao 3º ano do Ensino Médio, cujo desempenho parece estagnado, e não dá mostras de que será alto o suficiente para cumprir as metas futuras até 2021 ou depois.

Tabela 5.4: Tendências de crescimento e metas projetadas do Ideb de MG por rede de ensino e série, e as diferenças entre elas.

SÉRIE	ANO	PARTICULAR			PÚBLICA			TOTAL		
		TEND	META	DIF	TEND	META	DIF	META	PROJ	DIF
5 EF	2015	8,0	7,7	0,2	6,5	5,9	0,6	6,7	6,0	0,7
	2017	8,2	7,9	0,3	6,9	6,1	0,8	7,1	6,2	0,9
	2019	8,4	8,0	0,4	7,3	6,4	0,9	7,5	6,5	1,0
	2021	8,6	8,1	0,5	7,6	6,6	1,0	7,9	6,7	1,2
9 EF	2015	6,4	7,3	-1,0	4,9	4,8	0,1	5,1	5,0	0,1
	2017	6,3	7,4	-1,1	5,2	5,0	0,2	5,4	5,2	0,2
	2019	6,3	7,6	-1,3	5,4	5,3	0,1	5,6	5,5	0,1
	2021	6,3	7,7	-1,4	5,7	5,6	0,1	5,9	5,7	0,2
3 EM	2015	6,0	6,8	-0,8	3,7	4,4	-0,7	3,9	4,7	-0,9
	2017	6,0	7,1	-1,1	3,8	4,8	-1,0	3,9	5,1	-1,2
	2019	5,9	7,3	-1,4	3,8	5,0	-1,2	3,9	5,3	-1,4
	2021	5,9	7,4	-1,5	3,9	5,3	-1,4	3,9	5,6	-1,7

Fonte: MEC, 2014.

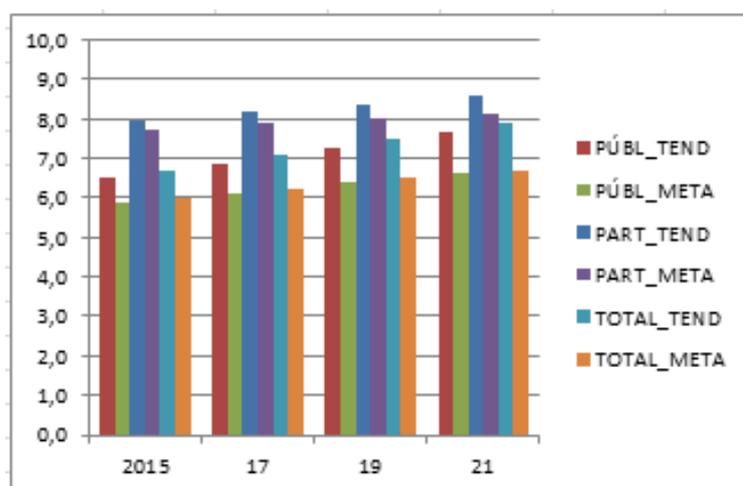


Gráfico 5.11: Tendências de crescimento e metas projetadas do Ideb de MG por rede de ensino, e as diferenças entre elas, para o 5º ano EF, para o período entre 2015 e 2021. Fonte: MEC, 2014.

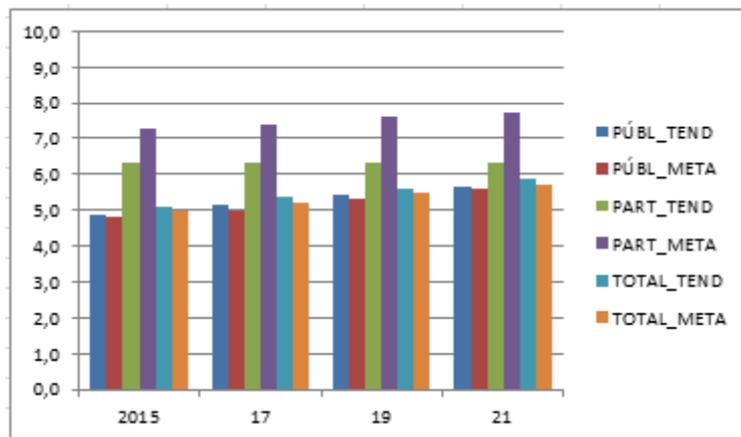


Gráfico 5.12: Tendências de crescimento e metas projetadas do Ideb de MG por rede de ensino, e as diferenças entre elas, para o 9º ano EF, para o período entre 2015 e 2021. Fonte: MEC, 2014.

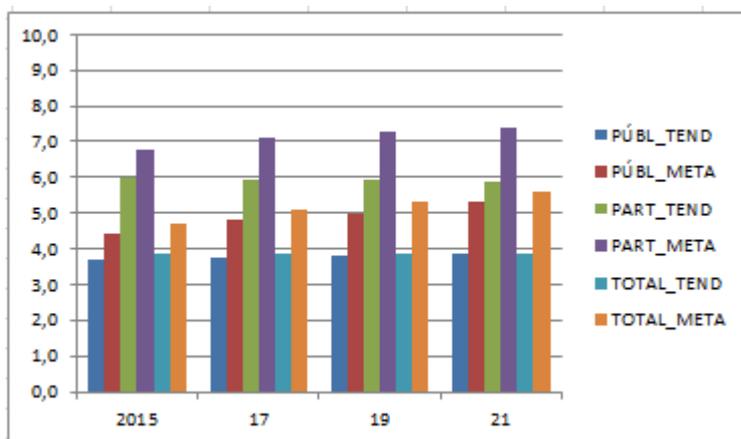


Gráfico 5.13: Tendências de crescimento e metas projetadas do Ideb de MG por rede de ensino, e as diferenças entre elas, para o 3º ano EM, para o período entre 2015 e 2021. Fonte: MEC, 2014.

## 6. A APLICAÇÃO DE MODELOS LINEARES HIERÁRQUICOS COMO UM MECANISMO DE APERFEIÇOAMENTO DAS METAS DO IDEB

### 6.1. INTRODUÇÃO

No capítulo anterior, vimos como o Ideb vem se comportando, tanto em termos nacionais, como no caso específico do Estado de Minas Gerais. Entretanto, uma questão que os dados mostrados até aqui não abordam diretamente é um problema de grande interesse para muitos atores envolvidos neste processo, qual seja, o cumprimento das metas desse índice, não do ponto de vista global, como os agregados nacionalmente ou por estado, mas sim por parte das diferentes instituições educacionais por ele avaliadas, ou seja, das escolas envolvidas nesse processo.

Conforme se tornará explícito em partes posteriores deste trabalho, será utilizada uma metodologia estatística específica para lidar com o problema da determinação das metas para cada escola mineira avaliada, tanto pela Prova Brasil quanto pelo SIMAVE, e que tem a potencialidade de gerar resultados mais acurados e razoáveis do que em muitos casos se vêm observando na experiência real com esse índice. O método estatístico aqui empregado é a chamada modelagem linear hierárquica (*hierarchical linear modelling*, ou HLM, no inglês original) e, mais especificamente, sua versão especialmente concebida para estudos longitudinais, que considerem a escola e suas múltiplas testagens como objeto de análise (MENARD, 2002).

Portanto, neste ponto, faz-se necessário tecer algumas considerações que justifiquem o emprego desta metodologia para tal propósito. E começar-se-á abordando este tópico por meio de uma breve, porém concisa, exposição acerca das limitações dos métodos clássicos de testagem de hipóteses, que aqui são preteridos em favor do HLM.

## 6.2. REVISÃO DE CONCEITOS E MÉTODOS CLÁSSICOS DE TESTES DE HIPÓTESES

### 6.2.1. Os testes $t$ e a ANOVA

Muito frequentemente, e não apenas em ciências sociais, os investigadores se preocupam em saber se dois ou mais objetos de análise – digamos, duas escolas, em estudos de eficácia escolar – são semelhantes ou não no tocante a uma dada variável de interesse, como, por exemplo, a proficiência média de seus respectivos alunos numa disciplina avaliada por um exame externo.

Se o interesse foi especificamente por dois objetos, como duas escolas, um procedimento clássico que se emprega para se analisar a diferença de desempenho entre as duas é o chamado teste de significância  $t$ .

Por sua vez, o funcionamento deste teste, aplicado à diferença entre as médias dos dois grupos considerados, é deveras simples. Considera-se que a diferença específica  $D (= \mu_1 - \mu_2)$  entre as duas médias pertença a um conjunto de possibilidades, sendo que, quando  $D = 0$ , naturalmente as médias são iguais entre si. A partir dessa suposição, uma forma de se tratar o problema é considerar que a hipótese nula, ou seja, de que as médias dos dois grupos são, ao menos aproximadamente, iguais, corresponde a uma distribuição de probabilidades para  $D$  que esteja centrada em zero, e que se delinieie segundo uma curva normal, cuja dispersão é medida pelo erro padrão da diferença entre as duas médias.

Um ponto a se considerar em toda essa situação são os graus de liberdade do problema, que dependem diretamente da quantidade de grupos envolvidos, e também do número de dados ou observações dentro de cada um deles. Levando-se então em conta os devidos graus de liberdade, é possível chegar a valores de corte para o quociente da diferença entre as médias pelo próprio erro da respectiva medida. Naturalmente, quanto mais o resultado dessa divisão se aproxima de zero, maior se torna a probabilidade de que a diferença observada entre as médias tenha talvez se devido ao acaso durante a mensuração, e não propriamente à existência de uma diferença significativa entre os dois grupo, quanto à variável considerada no presente estudo.

Por outro lado, caso a divisão de  $D$  por seu respectivo erro-padrão comece a aumentar radicalmente de tamanho, isto sugere que a distância entre as médias (não importando com que sinal ela esteja sendo medida), está também aumentando de módulo, quando comparada com a imprecisão ou erro da própria medida. Em outras palavras, caminha-se, neste caso, para se aceitar a hipótese alternativa, que, contrariamente à nula, considera que as diferenças entre

os dois grupos são significativamente diferentes entre si. E, uma vez sabendo qual dos dois grupos possui mais do valor sendo medido do que o outro, fica também muito simples saber qual deles tem apresentado um comportamento superior, ou inferior, em relação ao outro, bastando, para isto, ver o sinal da diferença entre as médias.

Para que seja realizado, o teste  $t$  (ou  $z$ ) de Student trabalha com a hipótese de que os dois grupos possuem uma distribuição aproximadamente normal em torno de sua respectiva média  $M$ , e com um desvio-padrão  $DP$ . No entanto, muitas vezes, está-se interessado, nos estudos, em se fazerem comparações, não apenas entre dois grupos, mas sim simultaneamente entre vários deles. Este seria, por exemplo, o caso de, digamos, um município, que possui várias escolas, estar interessado em saber como esses estabelecimentos têm se comportado nos testes de proficiência, uns em relação aos outros (LEVINE, 2000).

Em casos assim, não seria mais possível se fazer o teste  $t$ , porque este comporta somente dois grupos (ou, neste exemplo, duas escolas). Porém, alternativamente, seria possível realizar uma análise de variância, ou ANOVA, que, em termos gerais, comporta-se do mesmo modo que um teste  $t$ , porém, não mais com o objetivo de averiguar a significância da diferença entre as médias de dois grupos, mas sim de vários deles. Cabe dizer que, devido a esta última característica, a ANOVA, em estudos educacionais e outros, costuma ser empregada com bastante frequência, já que, muitas vezes, os pesquisadores têm como objeto de pesquisa, não necessária e exatamente dois grupos de interesse, como duas escolas, por exemplo, mas sim várias dezenas, ou mesmo centenas ou milhares delas (MOORE, 2006).

O princípio norteador de uma análise de variância é uma comparação entre dois tipos de variabilidade que se pode encontrar nos dados disponíveis: uma externa, ou entre os diferentes grupos, e outra interna, ou dentro de cada um deles. Aliás, é precisamente essa ênfase no estudo desses diferentes tipos de variabilidade, que dá nome ao método, o qual emprega uma razão da variância externa por outra interna dos dados, a fim de se chegar a uma estatística  $F$ , que compara essas duas quantidades.

#### *Uma ilustração para a ANOVA*

Um exemplo gráfico ajuda a explicar o funcionamento de uma ANOVA. Suponhamos que, num dado contexto educacional de interesse, desejemos verificar como três escolas estão se comportando em relação aos resultados de seus respectivos alunos.

Uma forma quase intuitiva que um avaliador inicialmente empregaria para averiguar a possível diferença entre as escolas, seria calcular as médias de desempenho de seus

respectivos alunos, e então compará-las entre si. Num caso hipotético, suponhamos, por exemplo, que três escolas, de A a C, obtiveram médias expressas por uma escala arbitrária de proficiência, dadas na tabela e no gráfico a seguir:

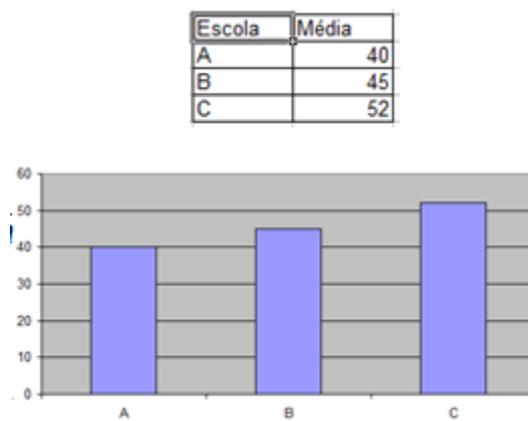


Figura 6.1: Um caso hipotético de diferenças entre médias escolares. Fonte: Elaboração própria.

Cabe observar que o último gráfico é uma solução frequentemente empregada por diversos relatórios, para os mais diferentes tipos de público, que se vêm produzindo sobre os resultados das avaliações em grande escala pelo mundo. Nesses casos, costuma-se, por exemplo, usar as barras do gráfico para comparar o resultado médio de uma escola com o de outras unidades de interesse, como um município ou um estado. Ou então se usam também as barras – ou, alternativamente, linhas – para comparar os resultados médios de uma mesma escola ao longo de alguns anos sucessivos. (OCDE, 2009)

Entretanto, uma limitação considerável dessa representação é que ela não consegue descrever a variabilidade da medida, que se expressa pela distribuição dos resultados individuais em torno da média. Esse ponto, ou esta limitação é, por sinal, extremamente relevante nos casos que precisamente nos interessam, de se desejar saber se as diferenças observadas entre os diferentes grupos são, de fato, significativas ou não.

O diagrama de caixa a seguir, referente às mesmas escolas A, B e C do exemplo anterior, ajuda a explicar um pouco melhor essa situação.

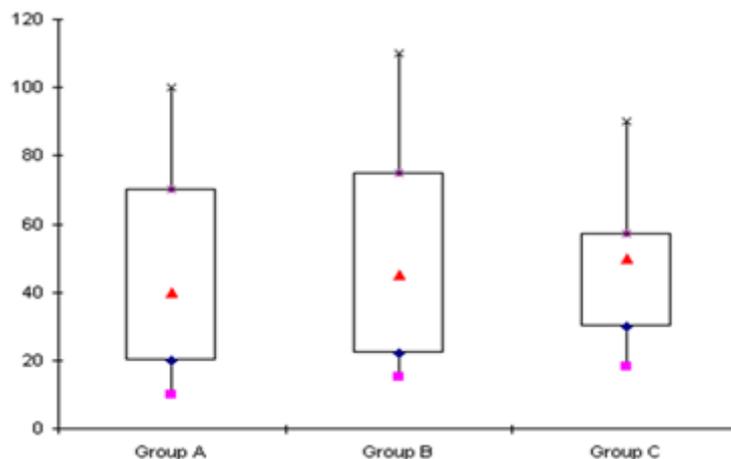


Figura 6.2: O caso da distribuição de resultados dentro das escolas, segundo os dados da figura anterior. Fonte: Elaboração própria.

Em cada uma das caixas dessa figura, naturalmente, as extremidades inferior e superior das linhas verticais indicam, respectivamente, o menor e o maior escores dos alunos daquela escola específica. Já os lados horizontais inferior e superior das caixas correspondem, respectivamente, ao primeiro e ao terceiro quartil da distribuição, ou 25° e 75° percentis. Por sua vez, falta, no interior das caixas, outra linha horizontal, que representaria a mediana da distribuição, e que, a princípio, poderia estar em qualquer altura da caixa, e não necessariamente no seu meio. Além disso, em cada caixa também se representa um triângulo vermelho indicando a posição das médias de cada distribuição, e que, neste caso, correspondem aos valores já mostrados na tabela e no gráfico 6.1.

Por este último diagrama de caixa, é possível perceber que as diferenças de resultado entre as escolas são bem menos acentuadas do que aquela percepção que se teve com o gráfico de barras mostrado anteriormente, embora ambos os casos sejam exatamente idênticos. Isto porque ocorre um acentuado nível de superposição entre os resultados das escolas.

Do ponto de vista do cálculo da estatística  $F$ , o numerador da razão representa a diferença das médias entre grupos, algo que, segundo este último gráfico, corresponderia à distância entre os triângulos vermelhos (ou médias), medida no eixo vertical. Se as médias são muito parecidas, a distância vertical entre elas é pequena, o que diminui o valor de  $F$ , ao fazer decrescer seu respectivo numerador. Por outro lado, caso as diferenças entre as escolas sejam mais acentuadas, então ocorre um maior afastamento entre suas médias, fazendo  $F$  aumentar, junto com seu numerador.

Por outro lado, a estatística  $F$  da ANOVA também possui, para seu cálculo, um denominador, que mede precisamente a variabilidade média de todo o conjunto disponível de

dados, ou seja, de todas as escolas ou grupos considerados na análise. A variância total assim obtida corresponde à soma das variâncias individuais dos grupos considerados. Valores maiores dessa variância indicam a existência de uma maior dispersão dos resultados em torno das diversas médias. Dessa forma, aumenta-se a superposição dos resultados, o que faz com que as diferenças observadas entre as escolas sejam menos nítidas. Em termos matemáticos, ocorre um aumento do denominador da razão, fazendo com que esta última, correspondente à estatística  $F$  da ANOVA, diminua de valor, aproximando-se da conclusão dada pela hipótese nula, segundo a qual não há diferenças significativas de desempenho entre os diferentes grupos.

Inversamente, porém, caso os valores individuais encontrados dentro de cada grupo estejam mais concentrados em torno de suas respectivas médias, então as diferenças entre as escolas se mostram mais destacadas. Um caso assim é precisamente o que aparece na figura a seguir:

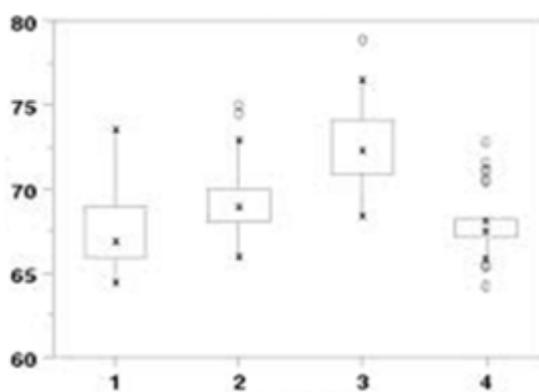


Figura 6.3: Outro exemplo hipotético de distribuição de resultados dentro das escolas, porém com um menor grau de superposição de escores individuais entre elas. Fonte: Elaboração própria.

Neste caso, percebe-se que a superposição de resultados é menos acentuada, fazendo com que algumas escolas se destaquem mais das outras pelos seus resultados, seja para mais ou para menos. Matematicamente, ocorre uma diminuição do denominador da estatística  $F$ , fazendo com que esta última aumente, e leve a decisão para mais próximo de se optar pela hipótese alternativa, ou seja, daquela segundo a qual existe uma diferença significativa de resultados entre os grupos.

Entretanto, uma limitação do procedimento da ANOVA é que, embora ele seja capaz de apontar se está ou não havendo uma variação significativa de resultados entre diversos grupos, nos casos em que se pode perceber isso, ainda assim não se consegue identificar a qual ou quais grupos, especificamente, se deve a existência desse resultado. Ou, em outras

palavras, o método em questão é capaz de descobrir se existe ou não uma diferença significativa entre os diversos grupos; entretanto, uma vez tendo esta sido detectada, ele não consegue identificar quais são os grupos mais discrepantes.

### 6.2.2. Análise de regressão

Outra metodologia estatística muito utilizada para se analisarem diferenças significativas entre grupos quanto a uma ou mais determinadas variáveis de interesse é a análise de regressão, a qual, por sua vez, associa-se diretamente à ANOVA. Basicamente, por uma análise de regressão, torna-se possível avaliar o grau médio de associação de uma ou mais variáveis de interesse, chamadas de variáveis independentes ou explicativas, com um *output*, ou variável dependente, cujo comportamento é, assim, “explicado” pelo citado grupo de covariáveis.

Um ponto que cabe ressaltar é a necessidade de se reconhecer que, via de regra, não se pode chamar as covariáveis de variáveis causais, visto que este último termo implica necessariamente a existência de um processo de causa e efeito respectivamente entre a variável independente e a dependente. Entretanto, o objetivo de estudos envolvendo análises de regressão costumam ser mais modestos, pois estas se detêm, basicamente, no grau de associação entre as variáveis de interesse, deixando o problema da causalidade para outras instâncias do estudo, ou simplesmente não tratando dessa questão.

Com efeito, é notório o fato de que, particularmente nas ciências sociais – entre as quais se incluem as pesquisas sobre educação –, fazer controles em moldes laboratoriais para se tentar averiguar possíveis causas para os fenômenos observados, é algo extremamente improvável de se fazer, tanto por limitações técnicas quanto éticas. Dessa forma, por exemplo, seria extremamente questionável destinar determinadas crianças a certos tipos de experiências educacionais – ou à ausência destas –, apenas para se satisfazer aos critérios de randomização exigidos pelos experimentos controlados em laboratório. Mas, feita essa ressalva, há de observar que a regressão representa um vigoroso mecanismo de análise, o qual também possui diversos tipos de aplicação e também de formulação, a fim de atender a diferentes propósitos, bancos de dados e questões investigadas.

Frequentemente, as variáveis envolvidas numa regressão são contínuas, podendo pertencer ao nível intervalar (onde o zero da escala é arbitrário, conforme comumente ocorre nas escalas de proficiência empregadas nas avaliações em grande escala), ou então ao nível de razão (onde o zero, de fato, indica a ausência da quantidade que se está mensurando).

A expressão matemática para uma regressão associando uma variável dependente  $Y$  a uma ou mais variáveis independentes  $X_i$  é dada pela fórmula:

$$Y = \beta_0 + \sum \beta_i * X_i + e$$

onde:

$\beta_0$ : intercepto da reta ou do plano representado pela equação

$\beta_i$ : o  $i$ -ésimo coeficiente angular da reta ou do plano de regressão

Uma típica aplicação dessa equação ao caso de estudos educacionais refere-se à existência do chamado gradiente socioeconômico, pelo qual a proficiência observada dos alunos avaliados numa determinada disciplina acadêmica – como Língua Portuguesa ou Matemática – possui uma associação positiva e significativa com o índice socioeconômico do aluno e de sua família, normalmente mensurado como uma composição de resultados obtidos a partir de variáveis descrevendo dimensões tais como os níveis de escolaridade, posse, renda e consumo dos membros de sua respectiva família, entre outros dados de possível interesse.

No caso do gradiente socioeconômico prevendo ou “explicando” a proficiência, os coeficientes linear e angular da reta de regressão possuem interpretações bastante significativas para a análise do problema. O coeficiente linear –  $\beta_0$  – corresponde à proficiência esperada de um aluno que possui um índice socioeconômico (ISE) igual a zero. Por sua vez, cabe lembrar que o ISE é medido numa escala intervalar, de modo que o valor nulo de sua respectiva escala é arbitrário. Nesse sentido, é comum calcular-se o ISE por meio de determinados programas ou pacotes estatísticos – como o PARSCALE, o BILOG ou o BILOG-MG, entre muitos outros –, que costumam empregar, nesses cálculos, a metodologia da Teoria da Resposta ao Item (TRI). E, via de regra, essas medidas de ISE são produzidas de modo que a média amostral ou populacional considerada corresponde a um valor nulo e a um desvio-padrão igual a uma unidade.

Na figura a seguir, apresenta-se um típico exemplo de um problema de gradiente socioeconômico em educação. Trata-se de um diagrama de dispersão referente a uma escola específica, em que cada ponto corresponde a um determinado aluno dessa mesma escola, classificado segundo dois eixos ou variáveis. No eixo horizontal, mensura-se o seu índice socioeconômico (*SES* no original inglês), ao passo que, no eixo vertical, encontra-se a sua proficiência (*achievement*) numa escala arbitrária.

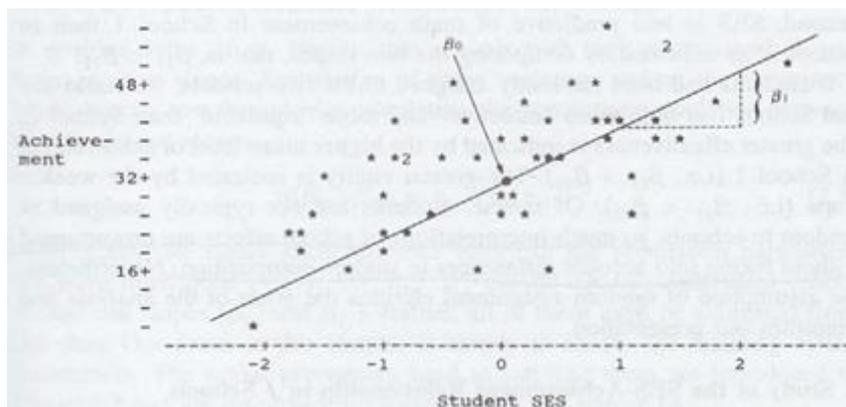


Figura 6.4: Um exemplo de gradiente socioeconômico para o caso educacional. Fonte: BRYK (2002).

Nesta última figura, observa-se que o ponto da reta indicado por uma seta e associado ao parâmetro  $\beta_0$  corresponde ao resultado esperado de um aluno que possui um índice socioeconômico nulo, valor este, por sua vez, corresponde à média do ISE para a escola em questão. Além disso, também é possível observar no gráfico uma associação positiva e consideravelmente elevada entre o aumento do nível socioeconômico e o aumento da nota esperada nesse exame. Matematicamente, essa associação se traduz pela inclinação da reta de regressão, dada pelo parâmetro  $\beta_1$ , que também aparece, na parte superior direita da figura.

Por outro lado, também se observa que, via de regra, os valores reais da proficiência dos alunos quase nunca coincidem exatamente com seus respectivos valores previstos. A distância vertical entre esses dois valores corresponde ao erro – o valor  $e$  na última equação – da medida individual da proficiência dos alunos. Se o valor real estiver acima da reta, o erro é positivo; se estiver abaixo, o erro é negativo, e se for nulo, então o valor real e o estimado são coincidentes.

Outro ponto que merece uma consideração especial é a interpretação dos parâmetros  $\beta_0$  e  $\beta_1$  do ponto de vista daquilo que, digamos, um gestor deseja de um sistema educacional qualquer. Idealmente, os sistemas educacionais devem ser eficazes, isto é, ser capazes de fazer os seus alunos adquirirem um nível consideravelmente elevado de proficiência nas diversas disciplinas trabalhadas em sala de aula, ao mesmo tempo em que também deveriam ser equânimes, ou seja, promover um nível elevado de aprendizado em todos os seus alunos, independentemente das condições socioeconômicas mais ou menos favoráveis dos discentes. Para ilustrar melhor como se interpreta essas duas características do ponto de vista dos parâmetros da reta de regressão “proficiência versus ISE”, apresentamos a figura a seguir, referente a duas escolas hipotéticas, A e B, segundo os mesmos moldes do gráfico anterior:

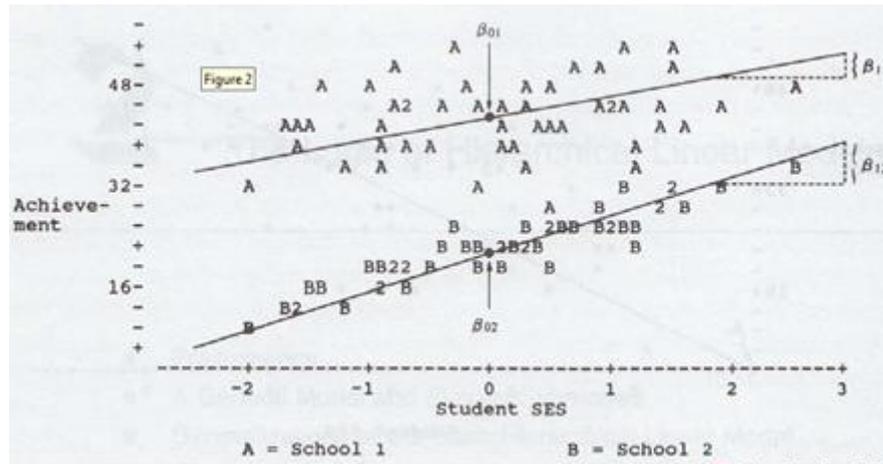


Figura 6.5: Dois diferentes exemplos de gradiente socioeconômico para a situação educacional. Fonte: BRYK (2002).

Neste caso, observa-se que, em termos de eficácia média, os resultados da escola A apresentam-se como superiores aos da escola B, o que se traduz no fato de que o coeficiente linear de A – aqui chamado de  $\beta_{01}$  – é maior, ou mais alto, que o coeficiente linear  $\beta_{02}$  da escola B. Ou, em outras palavras, um aluno da escola A que possui um índice socioeconômico médio nulo (ou seja, igual à média populacional), possui uma nota estimada superior à de um aluno com esse mesmo valor nulo de ISE, porém matriculado numa escola B.

Por sua vez, em termos de equanimidade, a escola A também se encontra numa posição mais favorável que a escola B. Isto porque o seu gradiente socioeconômico – medido pela inclinação de sua respectiva reta de regressão – é menos acentuado que o da escola B. Ou, em outras palavras, na escola A, para efeitos da nota esperada no exame em questão, importa menos o fato do aluno ser mais rico ou mais pobre. E, indo por essa linha, tem-se que uma situação ideal corresponderia àquela onde a inclinação da reta de regressão seria nula, indicando que a mesma estaria na horizontal. Em casos assim, qualquer que fosse o índice socioeconômico do aluno de uma dada escola, a sua média esperada seria a mesma para todos os discentes, independentemente do fato deles serem mais ou menos ricos.

A ilustração que se acaba de comentar ressalta a importância do uso de uma metodologia de análise linear hierárquica em contextos educacionais (e outros). Caso se utilizasse, em situações como a descrita na última figura, um modelo clássico de regressão, o resultado seria uma única reta, aproximadamente no meio do caminho entre as nuvens de alunos das escolas A e B. No entanto, ainda que tal reta única tivesse o mérito de apontar valores globais para a proficiência média e também para a tendência de crescimento das notas

esperadas, em função do ISE, ainda assim, isto, naturalmente, em nada conseguiria captar as diferenças entre as distintas escolas, quanto a esses mesmos aspectos.

### 6.3. A MODELAGEM LINEAR HIERÁRQUICA COMO UMA ALTERNATIVA PARA OS MODELOS CLÁSSICOS DOS TESTES DE HIPÓTESES

Em relação ao que se acaba de dizer na seção precedente – a maior conveniência dos modelos lineares hierárquicos em diversas situações educacionais –, também se pode verificar que o uso desses modelos é bastante relevante devido ao fato de que este último método pode ser empregado mesmo nos casos em que os clássicos falham, devido à não observância dos respectivos pressupostos para estes últimos.

Por exemplo, para que uma regressão clássica seja conduzida, os dados utilizados precisam de:

- 1) Ser normalmente distribuídos ao longo da reta de regressão.
- 2) Possuir variâncias iguais ao longo da reta (ou seja, apresentar homocedasticidade).
- 3) Ser independentes entre si.

Entretanto, todos esses três pressupostos básicos são, via de regra, desobedecidos quando se procede a uma regressão linear clássica ou convencional envolvendo dados de proficiência educacional. O primeiro deles é violado quando as distribuições de escores em torno da reta assumem padrões bastante diferentes da normalidade. Além disso, o valor da dispersão dos erros em torno dos seus respectivos valores previstos também pode variar bastante, dependendo da região da escala de proficiência que se esteja considerando. Por exemplo, nos intervalos medianos de proficiência, a dispersão tende a ser maior do que nos extremos dessa mesma escala, fazendo com que o segundo pressuposto (o da homocedasticidade) seja desobedecido.

Outro problema decorre do fato, nos bancos de dados usualmente utilizados, de que é comum, por exemplo, que alunos matriculados em uma mesma turma ou escola tendem a apresentar escores mais semelhantes entre si, e mais diferentes dos alunos de outras turmas ou escolas, fazendo com que os erros deixem de ser independentes, o que, por sua vez, se constitui numa violação do terceiro pressuposto.

Portanto, o reconhecimento dessas limitações dos métodos clássicos levou à busca por modelos alternativos de análise dos dados, tanto educacionais quanto de outras áreas. E, nessa

linha, uma proposta desenvolvida nas décadas mais recentes por diversos autores tem sido a utilização dos chamados modelos lineares hierárquicos. Basicamente, o que essa nova metodologia faz é considerar a estrutura hierárquica que muitos dados possuem, e então propor uma série de análises de regressão feitas especificamente para cada grupo que surge a partir dessa hierarquização.

Por exemplo, no caso educacional, os alunos avaliados num teste podem ser agrupados em turmas, as quais se reúnem em escolas, e estas, por sua vez, compõem redes municipais e estaduais de ensino. Portanto, ao invés de se utilizarem procedimentos globais de regressão para todos os casos envolvidos – o que, por exemplo, poderia levar à produção de uma única reta de regressão para todo o conjunto de dados –, o que esse modelo propõe é a realização de uma série de regressões aplicadas a grupos individuais. E, como resultado disso, é então possível obter, digamos, centenas ou mesmo milhares de retas de regressão, uma para cada escola considerada.

Segundo Bryk e Raudenbush (2002), a metodologia dos modelos lineares hierárquicos tem recebido diversos nomes desde quando começou a ser inicialmente desenvolvida, muito embora, apesar dessas diferentes denominações, ela se refira, basicamente, ao mesmo tipo de procedimento. Por exemplo, Goldstein e Mason chamam esse método de modelagem linear de multinível; outros, como Elston, Grizzle, Laird, Ware e alguns demais autores, de modelos de efeitos aleatórios. Alguns outros, particularmente preocupados com a aplicação desse método em econometria, chamam-no de modelos de regressão de coeficientes aleatórios, e assim por diante.

Contudo, independentemente da maneira como possa ser chamada, essa metodologia invariavelmente se refere ao procedimento de se realizar uma série de análises de regressão no banco de dados, respeitando-se a estrutura hierárquica destes últimos. Passaremos, a seguir, a expor a construção desse modelo, tomando como base uma situação tipicamente encontrada nos estudos de avaliação educacional: a aferição do desempenho médio das escolas por meio de um ou mais testes de proficiência aplicados a seus alunos.

### 6.3.1. O modelo nulo de dois níveis

Consideraremos aqui, para fins de exposição, o tratamento de dois níveis hierárquicos, algo que, precisamente, será adotado em partes posteriores deste trabalho. E, também nesse sentido, cabe dizer que a expansão desse método para níveis superiores, como três ou mais, funciona, ao menos em termos formais, como uma simples extensão deste mesmo tratamento,

de modo que, embora os cálculos tornem-se paulatinamente mais complexos com o aumento do número de níveis, em essência, o propósito e o funcionamento do método permanecem o mesmo que no caso de dois níveis.

Para exemplificar, então, a utilização desse modelo, consideremos um caso bastante comum em análises desse tipo, que considera os alunos agrupados por escola, num contexto educacional qualquer. Dessa forma, os alunos são o objeto de análise no nível 1, ao passo que as escolas são o objeto do nível 2.

Nesse sentido, no nível 1, pode-se escrever a seguinte equação:

$$Y_{ij} = \beta_{0j} + r_{ij}$$

onde:

$Y_{ij}$ : proficiência  $Y$  do aluno  $i$  na escola  $j$

$\beta_{0j}$ : média de proficiência da escola  $j$

$r_{ij}$ : erro estocástico da medida, ou seja, a diferença entre o valor “previsto” (a média da escola, ou  $\beta_{0j}$ ) e o valor observado para o aluno  $i$  da escola  $j$

Em outras palavras, o que esse modelo quer dizer é simplesmente que a nota de um aluno é igual à média de sua respectiva escola, acrescido de um diferencial  $r$  que pode ser positivo, negativo ou nulo, dependendo da nota observada do aluno ter sido maior, menor ou igual à média de sua escola, respectivamente.

Analogamente, constrói-se um modelo para o nível 2, ou seja, da escola, pelo qual a média desta última é igual à grande média da amostra ou população, acrescido também de um diferencial, segundo a equação:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

onde:

$\beta_{0j}$ : média de proficiência da escola  $j$

$\gamma_{00}$ : grande média amostral ou populacional

$u_{0j}$ : erro estocástico da medida, ou seja, a diferença entre o valor “previsto” (a grande média), e a média da escola  $j$

Algumas suposições deste modelo são que o erro estocástico  $r_{ij}$  tenha um valor estimado nulo e uma variância igual a  $\sigma^2$ . Dessa forma:

$$E(r_{ij}) = 0 \quad \text{e} \quad \text{Var}(r_{ij}) = \sigma^2$$

Algo análogo se passa também no nível das escolas, de modo que:

$$E(u_{0j}) = 0 \quad \text{e} \quad \text{Var}(u_{0j}) = \tau_{00}$$

Uma conhecida propriedade da variância de uma soma é que ela é igual à soma das variâncias de suas respectivas parcelas. Aplicando essa propriedade ao presente caso, temos que:

$$\text{Var}(Y_{ij}) = \text{Var}(\beta_{0j} + r_{ij}) = \text{Var}(\gamma_{00} + u_{0j} + r_{ij}) = \text{Var}(\gamma_{00}) + \text{Var}(u_{0j}) + \text{Var}(r_{ij}) = \tau_{00} + \sigma^2$$

Visto que  $\text{Var}(\gamma_{00}) = 0$ , pois este último parâmetro é fixo, ou seja, a grande média não varia. A partir daí, torna-se estatisticamente possível estimar para esse sistema uma medida de sua desigualdade educacional, através do chamado de *coeficiente de correlação intraclasse* ou CCI – que varia entre zero (maior igualdade possível) e um (maior desigualdade possível).

O ponto de partida para o seu cálculo é considerar a variação apresentada pelos resultados dos alunos, que pode ser de dois tipos. O primeiro deles, que poderíamos chamar de variação *intraescolar*, corresponde aos desvios (para mais ou para menos) que as notas dos alunos apresentam em relação às médias de suas respectivas escolas. O segundo tipo, que chamaremos de variação *extraescolar*, corresponde à variação das médias das escolas em relação à grande média de toda a amostra ou de toda a população avaliada. Dessa forma, as escolas variam entre si quanto ao seu desempenho coletivo médio (variação extraescolar) e, dentro de cada uma delas, os alunos também variam entre si quanto ao seu desempenho individual (variação intraescolar). A soma desses dois tipos de variação resulta na variabilidade total de desempenho observada nos resultados das avaliações dos alunos, conforme se demonstrou acima.

A partir dessas duas variâncias – a do nível do aluno e a do nível das escolas –, o coeficiente de correlação intraclasse (CCI) nada mais é do que a proporção que a variação de desempenho devida às escolas representa em relação à variação total (a intraescolar e a extraescolar) observada. Em termos simbólicos, é comum apresentar o CCI como:

$$\text{CCI} = \text{Var}(\text{escolas})/\text{Var}(\text{total}) = \text{Var}(u_{0j})/\text{Var}(Y_{ij}) = \tau_{00}/(\tau_{00} + \sigma^2)$$

Às vezes, ao invés de se usar uma proporção que varia de zero a um, alguns analistas e autores preferem multiplicar esse valor por 100, expressando, assim, o CCI na forma de um percentual.

### Três ilustrações do CCI

Apresentaremos a seguir algumas ilustrações que, esperamos, sejam capazes de esclarecer o conceito e a interpretação do CCI como um mecanismo de mensuração da igualdade (ou desigualdade) educacional.

#### Caso 1: Máxima desigualdade educacional (CCI = 1 ou 100%)

Nesta ilustração, consideremos que a nossa população educacional se resuma a seis estudantes, com os alunos identificados pelos números de 1 a 3 pertencentes a uma determinada escola (azul) e os outros três alunos, identificados pelos números de 4 a 6, pertencentes a uma segunda escola (vermelha). Na figura 1, podemos ver os resultados obtidos por esses alunos em um determinado exame hipotético, expressos em uma escala arbitrária de proficiência.

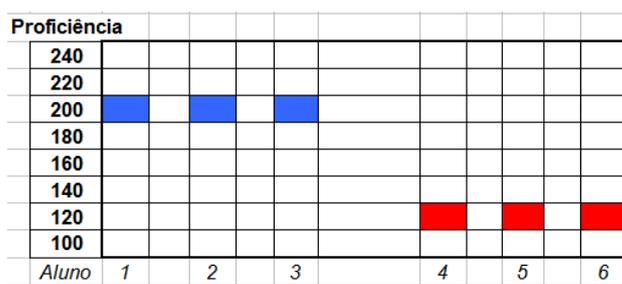


Figura 6.6: Caso de perfeita desigualdade escolar (CCI = 1). Fonte: Elaboração própria.

Por esta última figura, podem-se fazer duas observações notáveis:

- 1) Não existe variação intraescolar nessa situação, visto que as notas dos alunos dentro de cada escola são iguais. Em outras palavras, os desvios dessas notas em relação à média de suas respectivas escolas é zero, ou, ainda, a nota de cada aluno corresponde à média de sua respectiva escola.
- 2) A escola dos casos de 1 a 3 tem um desempenho médio superior ao da outra escola, visto que as médias dessas escolas correspondem, respectivamente, a 200 e a 120 pontos na escala de proficiência.

### Cálculo do CCI na Figura 6.6

Como não há variação das notas dos alunos em torno de suas respectivas médias escolares, então sigma quadrado é nulo. Por outro lado, existem, de fato, variações das médias escolares em relação à grande média. Nesse exemplo simples, a grande média seria 160, correspondendo à média dos seis alunos avaliados; portanto, a escola azul (com uma média de 200 pontos) estaria 40 pontos acima da grande média, e a escola vermelha (com uma média de 120 pontos) estaria 40 pontos abaixo da grande média. Dessa forma, o fato de sigma quadrado ser nulo e de tau zero não o ser, leva-nos ao seguinte resultado:  $CCI = \tau_{00}/(\tau_{00} + \sigma^2) = 1/(1 + 0) = 1$  ou 100%

### Interpretação

Na situação descrita pela Figura 6.6, o fato de um aluno pertencer a uma escola específica determina completamente o resultado que esse aluno terá numa dada avaliação. Portanto, digamos, caso essa avaliação esteja, por exemplo, sendo utilizada para se decidir quem entrará em uma faculdade ou em um emprego, os alunos “azuis” estarão em total vantagem em relação aos “vermelhos”, visto que, nesse caso, é a escola o único agente responsável pelo melhor ou pior resultado acadêmico desses estudantes, ao mesmo tempo em que há, de fato, casos de escolas melhores (como a azul) e piores (como a vermelha).

### Caso 2: Máxima igualdade educacional (CCI = 0 ou 0%)

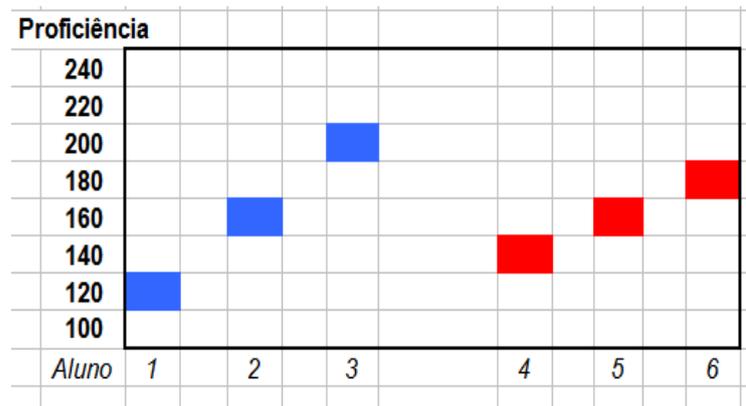


Figura 6.7: Caso de perfeita igualdade escolar (CCI = 0). Fonte: Elaboração própria.

Nessa segunda ilustração, mantendo-se as mesmas convenções do caso anterior, pode-se perceber que:

1) Não existe variação *extraescolar* nesse exemplo, porque as médias das escolas são iguais entre si. (A média, geometricamente, pode ser definida como o ponto do meio de uma distribuição simétrica de valores, como a que ocorre para ambas as escolas no exemplo). Dessa forma, para as duas escolas, a média é igual a 160 pontos, valor que também corresponde à grande média. Logo, não há variação entre as médias escolares e a grande média.

2) Por outro lado, agora existe uma variação *intraescolar*, visto que, dentro de cada escola, há alunos obtendo notas diferentes, que podem ser maiores, iguais ou menores do que as médias de suas respectivas escolas.

#### *Cálculo do CCI na Figura 6.7*

Como, nesse caso, a variação *intraescolar* assume um valor não-nulo, ao mesmo tempo em que a variação *extraescolar* vale zero, substituindo esses valores na equação do coeficiente de correlação *intraclasse*, obtemos:

$$CCI = \tau_{00}/(\tau_{00} + \sigma^2) = 0/(0 + 1) = 0$$

#### *Interpretação*

Na situação dada pela Figura 6.7, observa-se que, a princípio, o fato do aluno pertencer a uma determinada escola, por si só, não é nenhuma garantia de que ele estará abaixo ou acima da grande média observada. Ambas as escolas têm a mesma média e o que distingue um aluno de todos os demais (de sua própria escola ou de outra) é o diferencial da sua nota em relação à média de sua escola, que é igual à grande média da amostra ou população. Dessa forma, o fato de um aluno pertencer a esta ou àquela escola não representa nenhuma garantia de que o seu desempenho será maior, menor ou igual ao dos demais alunos. Essa seria, portanto, uma situação de perfeita igualdade educacional.

*Caso 3: Caso “realista” de desigualdade educacional ( $0 < CCI < 1$ )*

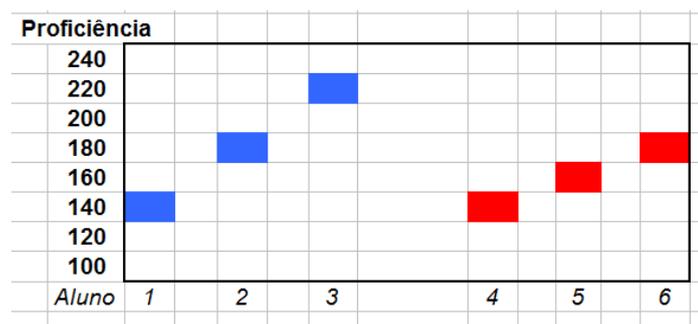


Figura 6.8: Caso “realista” de desigualdade escolar ( $0 < CCI < 1$ ). Fonte: Elaboração própria.

Esta última ilustração funciona como um esboço para o tipo de caso que, de modo geral, encontramos na prática:

- 1) É possível observar diferenças individuais entre os alunos dentro de uma mesma escola (portanto, a variação intraescolar não é nula).
- 2) Ao mesmo tempo, também é possível observar diferenças entre as médias das escolas (portanto, a variação extraescolar também existe, sendo, portanto, igualmente não-nula).

Esses dois fatos fazem com que, ao se calcular o CCI, o resultado obtido venha a se situar entre zero e um (ou entre zero e 100%).

### *Interpretação*

A situação descrita pela Figura 6.8 é, como se comentou, um caso intermediário entre a ocorrência de uma situação de máxima desigualdade (Figura 6.6) e de outra de máxima igualdade (Figura 6.7).

Nessa terceira situação, quanto mais o valor do coeficiente se aproxima de 1 (ou, em termos percentuais, de 100%), mais as médias escolares se afastam umas das outras e, portanto, mais desigual é o sistema educacional em questão. Inversamente, quanto mais o coeficiente se aproxima de zero, mais próximas entre si estão as médias das escolas e mais igualitário é o sistema.

### 6.3.1.1. Vantagens do modelo nulo

Talvez, à primeira vista, um modelo nulo como o discutido até aqui possa parecer simples demais, visto que ele não inclui nenhuma variável explicativa do lado direito das equações que o constituem. Por exemplo, no nível 1, ou do aluno, a proficiência  $Y_{ij}$  é posta a variar apenas em função da média  $\beta_{0j}$  da escola e de um erro aleatório  $r_{ij}$ , porém não depende de nenhuma outra variável que se poderia considerar no nível do aluno, de que são exemplos o seu respectivo índice socioeconômico ou o de sua família, seu gênero, etnia, idade, etc.

De modo análogo, a média de uma escola apenas flutua em torno de uma grande média, segundo um erro estocástico, porém não se encontra modelada por outra variável mensurada também no nível escolar, como, digamos, o índice socioeconômico médio ou o tamanho do estabelecimento (medido pelo número de alunos matriculados), ou a rede de ensino à qual a escola pertence, etc.

Entretanto, mesmo destituído dessa maior sofisticação, o modelo nulo, em sua grande simplicidade, é capaz de responder a questões fundamentais, que se devem considerar, a princípio, em qualquer estudo dessa natureza. Uma dessas questões é que ele permite estimar a significância estatística das variâncias dos erros no nível das escolas, por meio de um “diagnóstico” fornecido pelo seu respectivo valor  $p$ . À medida que o valor  $p$  vai se afastando de zero e se aproximando de um, maiores são as evidências de que as escolas pouco variam entre si quanto ao seu desempenho médio. Inversamente, valores  $p$  muito pequenos, ou seja, bastante próximos a zero, indicam o contrário: que as escolas – ou, ao menos algumas delas – estão variando significativamente entre si em termos de desempenho médio.

Uma segunda utilidade é que esse modelo também permite estimar o tamanho de cada uma dessas variâncias – ou seja, tanto da intraescolar, medida a partir dos escores dos alunos em torno de suas respectivas médias escolares – como também da extraescolar, mensurada a partir das médias escolares em torno da grande média. E, com base nisso, também se chega a uma terceira grande utilidade do modelo nulo: estimar o CCI do sistema avaliado, ou seja, avaliar em que medida a variação dos escores dos alunos se deve à variação das médias das escolas onde eles se encontram matriculados. Caso o CCI seja baixo, isso indica que o sistema tem se mostrado relativamente equânime, ou seja, as escolas têm, em geral, um comportamento aproximadamente igual entre si. E se, de fato, isso se verificar, a questão seguinte seria se preocupar com o status dessas medidas de proficiência das escolas, visto que a equanimidade somente seria algo saudável no caso de um desempenho mais elevado, ao

passo que haver equanimidade no meio de um desempenho em geral baixo, significa tão somente que o sistema educacional em questão está produzindo um nivelamento do aprendizado por baixo, o que, naturalmente, não interessa a ninguém.

Por outro lado, caso o CCI seja consideravelmente elevado, isso significa que existem variações significativas entre os resultados escolares, o que, por si, acena como um campo rico de investigação para os pesquisadores em educação, particularmente para aqueles interessados em compreender melhor os mecanismos que fazem as escolas destacar-se das demais, seja para mais ou para menos, em termos do desempenho acadêmico de seus alunos.

A consideração dessas possíveis diferenças de desempenho entre as unidades escolares leva, naturalmente, à formulação de modelos mais sofisticados que o nulo, segundo a metodologia aqui apresentada. Dessa forma, apresentar-se-á, agora, um passo seguinte que se pode dar na análise de dados educacionais baseada na modelagem linear hierárquica, correspondente ao modelo de nível 1.

### 6.3.2. O modelo de nível 1

No modelo de nível 1, conforme o próprio nome sugere, está-se interessado em modelar a equação do primeiro nível considerado, que, no exemplo aqui exposto, corresponde ao do aluno. Conforme já se comentou, há diversas covariáveis que a literatura sobre eficácia educacional, tanto brasileira como internacional, vem apontando como possuindo, em geral, associações significativas com o desempenho dos alunos nos testes. Alguns exemplos disso são a proficiência anterior dos alunos, o seu índice socioeconômico, o gênero, sua etnia ou, alternativamente, seu status em termos de pertencer ou não a uma maioria racial na sua respectiva sociedade, etc. Dessa forma, para se modelar a equação com base nessas novas informações, o novo modelo se escreve como:

$$Y_{ij} = \beta_{0j} + \sum \beta_{.j} \cdot (X_{ij} - M_{.j}) + r_{ij}$$

Neste caso, percebe-se que, no lado direito da equação, aparece um vetor que corresponde, precisamente, ao efeito das variáveis associadas ao desempenho  $Y_{ij}$ . E esse efeito, por sua vez, manifesta-se pela multiplicação da grandeza  $(X_{ij} - M_{.j})$ , que corresponde à distância entre o valor individual para a variável  $X$  e a média dessa mesma variável, pelo coeficiente  $\beta_{.j}$  correspondente, que pode ser positivo ou negativo, dependendo da direção do

efeito da variável em questão sobre a proficiência, ou então ser aproximadamente nulo, quando não influencia significativamente a proficiência, devendo, assim, ser retirado da equação.

Quanto ao nível 2 desse modelo, ele permanece da mesma forma que antes, ou seja, não se introduzem variáveis explicativas no lado direito da sua respectiva equação. Porém, ainda no nível 2, cabe comentar algumas opções interessantes que podem ocorrer quanto à modelagem de seus parâmetros. Isto porque existe, nesse caso, a possibilidade de se escolher entre duas opções de modelagem dos parâmetros ou coeficientes  $\beta$ , os quais, no segundo nível, como já se disse, aparecem no lado esquerdo das equações. Uma possibilidade é fixar o coeficiente de inclinação, de modo que:

$$\beta_{.j} = \gamma_0$$

Ou seja, o coeficiente é igual para todas as unidades consideradas (neste caso, seria igual para todas as escolas). Um exemplo dessa possibilidade, a título de ilustração, poderia ocorrer no caso do índice socioeconômico, que, normalmente e, nesta ilustração, por acaso, apresenta uma correlação positiva com o desempenho nos testes. Dessa forma, caso o coeficiente beta para o índice socioeconômico fosse fixado, então haveria um valor (neste caso, positivo) para o coeficiente do ISE que regeria a associação dessa variável com a proficiência dos alunos. No entanto, esse valor seria o mesmo para todas as escolas, visto que, pela última equação, o parâmetro é constante, independentemente da escola onde se situa o aluno.

Esse tipo de opção de tratamento do coeficiente costuma ser utilizado quando a sua variável correspondente não é o principal objeto de interesse de estudo, porém sim um controle estatístico. Isto porque, caso a variável em questão fosse de um interesse especial para um(a) pesquisador(a), então o mais indicado é deixar esse parâmetro variar para cada escola, segundo a equação:

$$\beta_{.j} = \gamma_0 + u_{.j}$$

Neste caso, percebe-se que cada escola  $j$  possui um diferente valor para seu coeficiente  $\beta$ , o qual flutua de  $u_{.j}$  em torno da grande média do coeficiente, que é  $\gamma_0$ .

Um exemplo para isso pode talvez ilustrar melhor essa metodologia. Caso uma pesquisa esteja interessada no impacto que o índice socioeconômico do aluno exerce sobre sua respectiva proficiência, então uma opção para o estudo seria deixar o parâmetro de regressão do ISE variar entre as escolas segundo a última equação dada. Dessa forma, os

resultados mais desejáveis, do ponto de vista da equanimidade do ensino, seriam aqueles nos quais os resíduos  $u_j$  fossem negativos, pois assim, combinados com  $\gamma_0$ , que tem um valor positivo, resultaria num coeficiente  $\beta_j$  menor do que a média populacional. E assim, na presente ilustração, isto estaria significando que a associação do ISE com a proficiência, nesta escola específica, é menos importante do que na média populacional, ou seja, a escola é mais equânime do que o restante do sistema tomado como um todo. Inversamente, caso  $u_j$  fosse positivo, aconteceria o contrário:  $\beta_j$ , além de positivo, se tornaria maior ainda, fazendo com que, dentro dessa outra escola, a associação entre o ISE e a proficiência fosse mais impactante do que a grande média.

Naturalmente, o inverso de toda essa interpretação se aplica aos casos em que o coeficiente de regressão beta for negativo. Exemplos disso ocorrem quando a variável em questão opera no sentido de diminuir o resultado, ao invés de aumentá-lo, como é comumente o caso, digamos, de um aluno ter um status de minoria étnica, ser portador de uma deficiência física ou mental, ou ainda ter um histórico de reprovação escolar. Em casos assim, a maior equanimidade se daria de modo inverso, ou seja, quando os resíduos  $u_j$  fossem positivos, pois assim, eles contribuiriam para aproximar de zero o coeficiente angular negativo  $\beta_j$ .

Porém, voltando ao caso da opção de se manter fixo o parâmetro  $\beta_j$ , esta possibilidade é usualmente empregada quando a variável a ele associada não for o objeto principal de interesse da pesquisa, mas sim caso ela estiver funcionando como um controle estatístico da situação considerada. Por exemplo, vimos acima uma ilustração em que o coeficiente angular do índice socioeconômico era deixado variar porque o interesse maior que havia era saber quais escolas estavam lidando de modo mais ou menos equânime com os alunos, do ponto de vista dessa variável.

Entretanto, suponhamos que o principal interesse da pesquisa fosse outro, digamos, o fato da escola tratar de modo igual ou diferente os alunos portadores de uma deficiência física ou mental qualquer. Então, neste caso, o principal interesse do estudo seria a associação estatística, na escola, entre o desempenho a condição física dos alunos, de modo que o parâmetro  $\beta_j$  explicitando essa relação deveria ser deixado variar. E a interpretação desse mesmo parâmetro seria relativamente simples: quanto mais inclinado ele for, mais indesejável é o efeito, visto que, neste caso, as notas dos alunos estariam variando mais em função de sua respectiva condição física. Inversamente, e num caso ideal, caso esse parâmetro de inclinação fosse zero, ou aproximadamente nulo numa determinada escola, isso significaria que esse

estabelecimento atingiu uma grande equanimidade em relação aos alunos portadores ou não do problema físico em questão, dado que os resultados acadêmicos não dependem, ou dependem muito pouco, da existência do referido problema.

Porém, ao mesmo tempo, nada impede – na verdade, é até mesmo desejável – que se acrescentem outras variáveis explicativas no lado direito da equação, como o nível socioeconômico dos alunos. Só que, neste caso, o ISE e sua relação com o desempenho não mais funcionam como o foco do estudo, mas sim como um controle estatístico na condução deste último. E assim, o parâmetro  $\beta_j$  a ele referente é deixado fixo, apresentando o mesmo valor para todas as escolas, ao mesmo tempo em que apenas se refere a um efeito geral observado, qual seja, a influência da condição socioeconômica sobre o desempenho acadêmico, porém sem mais se preocupar com as possíveis diferenças entre as escolas quanto a esse aspecto.

Todas essas considerações são fundamentais no processo de modelagem dos coeficientes  $\beta_j$  no segundo nível das equações, além de também serem extremamente importantes nas decisões que se tomam a respeito da centralização das variáveis explicativas no primeiro nível. Isto porque, como se viu pela última equação, a seguir repetida, no lado direito da equação de nível 1, escreve-se:

$$Y_{ij} = \beta_{0j} + \sum \beta_j \cdot (X_{ij} - M_j) + r_{ij}$$

Percebe-se, portanto, que os coeficientes angulares  $\beta_j$  são multiplicados pelos respectivos fatores  $(X_{ij} - M_j)$  correspondentes à diferença entre o valor observado ( $X_{ij}$ ) para o aluno e a sua respectiva média quanto a essa mesma variável ( $M_j$ ). Porém, uma dúvida que surge é que tipo de média considerar: se a média do grupo (no exemplo, a média da escola), ou a grande média (no mesmo exemplo, a média de todas as escolas).

Pois bem, essa opção de centralização se pauta precisamente pela escolha que, no modelo, se faz acerca do papel da variável explicativa. Se a variável em questão for o principal foco de estudo (como no exemplo acima, seria o caso daquela que informa sobre o fato do aluno ser ou não portador de uma determinada deficiência física), então a sua centralização se dá no nível do grupo, ou da escola e, neste caso, o parâmetro  $\beta_j$  é deixado variar. Por outro lado, caso a variável em questão se restrinja a um controle estatístico (como o ISE no mesmo e já mencionado exemplo), então a sua centralização ocorre na grande média, e o respectivo parâmetro  $\beta_j$  é conservado fixo.

*A redução da variância não-explicada no nível 1*

Conforme se viu, as equações de nível 1 (neste exemplo, referentes à proficiência dos alunos), nos modelos nulo e de nível 1 correspondem, respectivamente, a:

$$Y_{ij} = \beta_{0j} + r_{ij}$$

$$Y_{ij} = \beta_{0j} + \sum \beta_{.j} \cdot (X_{ij} - M_{.j}) + r_{ij}$$

Igualmente, também se comentou sobre  $\sigma^2$ , correspondente a  $\text{Var}(r_{ij})$ , ou seja, à variância não-explicada dos resultados dos alunos. Entretanto, cabe dizer que o significado dessa variância muda significativamente, ao passarmos do modelo nulo para o de nível 1. No caso do modelo nulo, essa variância é a maior possível, pois ela simplesmente corresponde à mensuração das diferenças médias entre as notas dos alunos e as médias de suas respectivas escolas, e calculada a partir dos resíduos ou erros  $r_{ij}$ . Porém, já no nível 1, o erro  $r_{ij}$ , a princípio, torna-se menor do que antes, pois não mais corresponde à diferença “bruta” entre o escore do aluno e a média escolar, porém sim a essa diferença já se descontando o efeito das covariáveis  $X$  de nível 1 presentes no mesmo modelo, e que, poderiam ser, por exemplo, o gênero e o índice socioeconômico dos alunos, entre várias outras possibilidades.

Dessa forma, ocorre uma redução da variância não-explicada no nível do aluno, ao se passar de um modelo nulo para um modelo de nível 1. E cabe observar que essa redução, não é somente previsível, como também desejável, pois, quanto mais intensa ela for, mais preciso é o modelo. Dessa forma, uma possibilidade de mensurar o aumento da capacidade explicativa de um modelo de nível 1 em relação a um modelo nulo, é através da equação:

$$\text{Ganho da capacidade explicativa} = (\text{Var}(\text{mod.}_\text{antigo}) - \text{Var}(\text{mod.}_\text{novo})) / \text{Var}(\text{mod.}_\text{antigo})$$

Caso se queira expressar esse ganho em termos percentuais, basta multiplicar por 100 a proporção expressa no lado direito desta última equação. Apenas a título de exemplo, se, numa situação ideal, um novo modelo, dotado de covariáveis adequadas, conseguisse prever precisamente os resultados reais obtidos, não haveria diferença entre os valores previstos e os observados. Dessa forma, os erros  $r_{ij}$  das medidas seriam todos nulos, e nula também seria sua respectiva variância. Assim sendo, a última equação ficaria como:

$$\text{Ganho da capacidade explicativa} = (\text{Var}(\text{mod.}_\text{antigo}) - 0) / \text{Var}(\text{mod.}_\text{antigo}) = 1 \text{ ou } 100\%$$

Ou seja, toda a variância que inicialmente não era explicada pelo modelo anterior, agora passou a sê-lo. Naturalmente, trata-se de um caso ideal e extremo, que praticamente nunca acontece em situações reais. Não obstante, em casos práticos, é possível encontrar modelos bastante eficientes e capazes de explicar, digamos, um terço, a metade ou até mais, da variância não-explicada pelas modelagens mais “brutas”, realizadas de modo exploratório no início dos estudos.

### 6.3.3. O modelo de nível 2

Na seção anterior, discutiu-se a possibilidade de se modelar o nível 1 das equações, que, no exemplo aqui seguido, refere-se a características mensuradas no nível do aluno, como seu índice socioeconômico, gênero, etc. Entretanto, é perfeitamente possível e legítimo considerar que a proficiência acadêmica também dependa de outros fatores que não sejam individuais, porém sim institucionais e vinculados à escola. Por exemplo, talvez possa haver diferenças de rendimento médio entre escolas pertencentes a diferentes redes de ensino – estaduais, municipais ou particulares – mantendo-se todas as demais condições controladas. E, de forma semelhante, também pode acontecer que variáveis medidas no nível do aluno, porém agregadas para a escola como um todo, também venham a exercer um impacto sobre o resultado individual. Um caso bastante comum desta última situação é o índice socioeconômico médio da escola, que, segundo várias pesquisas, chega a influenciar a proficiência dos alunos ainda mais intensamente do que o ISE individual dos mesmos.

Tudo isso leva à possibilidade ou conveniência de, nos estudos de eficácia escolar, considerarem-se também variáveis no nível da escola. E, segundo a metodologia por nós adotada, esse novo modelo poderia ser escrito como:

Nível 1:

$$Y_{ij} = \beta_{0j} + \sum \beta_{.j} \cdot (X_{ij} - M_{.j}) + r_{ij}$$

Nível 2:

$$\beta_{0j} = \gamma_{00} + \sum \gamma_{.j} \cdot (X_j - M_{.j}) + u_{0j}$$

$$\beta_{.j} = \gamma_{.0} + \sum \gamma_{.j} \cdot (X_j - M_{.j}) + u_{.j}$$

Portanto, de forma análoga ao que ocorreu no caso do nível 1 anteriormente apresentado, aqui também é possível modelar tanto o intercepto  $\beta_{0j}$  quanto as inclinações  $\beta_{.j}$  do nível 2 através de covariáveis mensuradas neste mesmo nível.

Igualmente, também é possível, no nível 2, fixar os parâmetros aleatórios  $u$ , ou então deixa-los variar, exatamente devido aos propósitos descritos anteriormente. Apenas há, em relação ao segundo nível, uma diferença quanto à centralização das variáveis, que costuma ser mais simples do que no nível 1. Isto porque, no segundo nível, não há mais a opção de se centralizarem as medidas na média do grupo porque, neste caso, o modelo já é externo à escola, e não interno a ela. Por exemplo, a média de ISE de uma escola costuma ser centralizada em relação à grande média (amostral ou populacional). Porém, não é possível centralizar a média escolar de ISE em relação à própria média grupal, porque, se assim fosse, o resultado seria sempre nulo, já que essas duas medidas são coincidentes.

Outra possibilidade de centralização, ainda não mencionada, que seria possível tanto no nível 2 como no nível 1 seria deixar a variável não centralizada, ou seja, adotar a média do conjunto como zero, de modo que:

$$(X_j - M_{.j}) = (X_j - 0) = X_j$$

Contudo, essa prática somente costuma ser utilizada caso a variável em questão de fato tenha um zero significativo, como no caso, por exemplo, de situações em que, digamos,  $X$  representa um determinado período de tempo mensurado a partir de um momento inicial específico identificado como instante zero. Em casos assim, o coeficiente linear  $\gamma_{.0}$  das equações passa a significar o valor inicial estimado da medida em questão.

#### 6.3.4. O caso longitudinal de dois níveis

O exemplo de situação multinível até agora descrito considera, como se viu, o aluno no nível 1 e a escola no nível 2. Trata-se de uma boa ilustração de uma estrutura hierárquica de dados. No entanto, há outras possibilidades de aplicação dessa metodologia, que levam em conta especificamente o acompanhamento longitudinal dos dados, algo que é, precisamente, aquilo que nos interessa no presente estudo.

Isto porque estamos interessados na evolução temporal dos resultados das escolas, seja de desempenho, como Língua Portuguesa ou Matemática, seja de rendimento, como o fluxo de alunos avançando em sua escolaridade. E, para casos assim, também é possível utilizar um modelo hierárquico de dois níveis, porém, com uma diferença fundamental: o primeiro nível passa a ser a trajetória de evolução temporal dos resultados, e o segundo nível é o indivíduo ao qual esses resultados pertencem. Desse modo, formalmente, costuma-se mudar as letras do caso convencional há pouco apresentado, com o propósito de se ressaltar a natureza distinta, agora longitudinal, dos dados.

Essa situação, matematicamente, se pode escrever de várias maneiras, dependendo do modelo que se deseja, ou precisa, empregar. Exemplos disso são o quadrático e o logístico, sendo que ambos esses casos são um pouco mais complicados, porque neles ocorre uma aceleração das taxas mensuradas. Entretanto, cabe dizer que, embora alguns modelos mais sofisticados possam, de fato, fornecer resultados mais precisos do que modelos mais simples, por outro lado é preciso observar que, na prática, esses modelos mais elaborados precisam de mais dados para funcionarem adequadamente, algo que com frequência não se observa na vida real (GLENN, 2005).

Em termos estatísticos, esse problema se expressa, por exemplo, no fato de que o número de observações, ou seja, de pontos no espaço matemático onde a medição é realizada, deve ser, no mínimo, igual à quantidade de parâmetros envolvidos nas previsões mais uma unidade. Dessa forma, por exemplo, uma reta, cuja equação tem dois parâmetros – os seus coeficientes linear e angular – poderia ser esboçada por três observações, e uma curva quadrática, que envolve três parâmetros, poderia ser traçada a partir de quatro observações, e assim por diante.

Entretanto, embora tais números deem algum balizamento sobre como proceder em casos como esse, ainda assim é preciso considerar outros fatores na escolha dos modelos. Estes tendem a se tornar mais robustos à medida que o número real de observações aumenta em relação ao mínimo necessário. Por razões assim, muitos estudos são feitos com base em

equações lineares, que reúnem todas essas vantagens acima mencionadas, e que também será o caso do presente estudo. Contudo, cabe ressaltar que, no modelo aqui utilizado, especificamente o modelo corresponde à proficiência média de uma escola, obtida em uma avaliação da Prova Brasil ou do SIMAVE, e o nível 2 será a própria escola. Matematicamente, isso se escreve como:

Nível 1:

$$Y_{it} = \pi_{0i} + \pi_{1i} \cdot a_{it} + e_{it}$$

Nível 2:

$$\pi_{0i} = \beta_{00} + \sum \beta_{0q} \cdot X_{qi} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \sum \beta_{1q} \cdot X_{qi} + r_{1i}$$

Neste caso, no primeiro nível, temos que:

$Y_{it}$ : média da escola  $i$  no instante  $t$

$\pi_{0i}$ : média da escola  $i$  no instante 0

$\pi_{1i}$ : taxa linear de evolução temporal da média da escola  $i$

$a_{it}$ : tempo  $t$  decorrido em relação ao instante 0 para a escola  $i$

$e_{it}$ : erro estocástico

Algumas observações cabem, aqui, ser mencionadas, em relação a essa modelagem. Uma delas é que não necessariamente a escola avaliada precisa ter dados para todos os exames, nem tampouco estes precisam estar igualmente espaçados no tempo. Ao contrário, pode haver lacunas de dados na série temporal, bem como intervalos desiguais de tempo entre resultados consecutivos, pois nada disso impede a realização da regressão linear.

Outro ponto que cabe ressaltar é a escolha da medida zero para o tempo. Embora o senso comum talvez diga que o zero devesse corresponder à observação inicial da série, a escolha do tempo nulo para esta aplicação é arbitrária, e nada impede que, por conveniência, o zero seja um ponto, digamos, no meio da série, ou em sua extremidade posterior. Dependendo dessas escolhas, naturalmente os números envolvidos na medição mudam, porém os resultados que eles representam continuam sendo a rigor os mesmos.

Outro ponto que merece ser destacado é a interpretação dos coeficientes linear e angular. O coeficiente linear  $\pi_{0i}$  corresponde, como se disse, no exemplo dado, à média da escola no instante zero, o qual se, de fato, for o início da série, corresponderia ao ano 2000 no caso do SIMAVE, ou 2005, na Prova Brasil. Já o coeficiente angular  $\pi_{1i}$  refere-se à taxa de crescimento anual dessa média escolar ao longo da série considerada. Como dá para ver pelos termos estocásticos das últimas equações, tanto  $\pi_{0i}$  quanto  $\pi_{1i}$  variam, ou seja, cada escola tem a sua própria média inicial e também a sua própria taxa de crescimento anual.

Por sua vez, ambos esses coeficientes são modelados no nível 2, da escola, da seguinte forma:

Para  $\pi_{0i}$ :

$\beta_{00}$ : grande média das escolas no instante 0

$X_{qi}$ : valor da variável explicativa  $q$  da escola  $i$

$\beta_{0q}$ : coeficiente angular da variável explicativa  $q$  para  $\pi_{0i}$

$r_{ii}$ : erro estocástico

Para  $\pi_{1i}$ :

$\beta_{10}$ : grande média das taxas temporais de variação das médias escolares

$X_{qi}$ : valor da variável explicativa  $q$  da escola  $i$

$\beta_{1q}$ : coeficiente angular da variável explicativa  $q$  para  $\pi_{1i}$

$r_{ii}$ : erro estocástico

Naturalmente, os resultados, em termos de proficiência acadêmica, tornam-se mais favoráveis à medida que os coeficientes lineares aumentam, indicando que as escolas já começaram na série em patamares superiores. E também são mais favoráveis quando as taxas de evolução são positivas e maiores, indicando que, a cada ano que passa, as médias experimentam maiores incrementos. Um exemplo de uma variável explicativa  $X_{qi}$  no nível da escola que será utilizada no presente trabalho, tanto para modelar o coeficiente linear quanto o angular, é o índice socioeconômico médio das escolas, o qual, conforme se verá posteriormente nos dados analisados, possui uma associação significativa e positiva com a proficiência média inicial das escolas, indicando que, quanto maior é o ISE médio da escola, maior em média tende a ser o seu resultado inicial. Não obstante, a associação do ISE médio com o coeficiente angular  $\pi_{1i}$ , embora significativa, costuma, como se verá, assumir um sinal negativo. Isto porque as escolas de menor ISE, embora tenham um início mais baixo na escala

– na verdade, ao menos em boa parte por isso mesmo – crescem mais, ao longo da série histórica, do que aquelas que já começam mais no alto, pois estas últimas se deparam com as limitações de um efeito-teto.

Passaremos agora, no capítulo seguinte, a apresentar uma situação que ilustra a aplicação desse modelo estatístico, ao mesmo tempo em que, a partir dele, apresenta, de modo mais específico, o problema que se quer resolver com a utilização dessa metodologia.

## **7. VOLATILIDADE DOS RESULTADOS EDUCACIONAIS: MENSURAÇÃO E CONSEQUÊNCIAS NO CASO DAS ESCOLAS PÚBLICAS DE MINAS GERAIS**

### **7.1. INTRODUÇÃO**

No capítulo anterior, discorreremos sobre a técnica de modelagem linear hierárquica que será utilizada para se analisar a adequação do estabelecimento das metas do Ideb, bem como para sugerir propostas de seu respectivo aperfeiçoamento. Chegou, então, o momento de começar a ver, mais detidamente, de que modo essa técnica pode nos ser útil nesta tarefa.

A título de uma introdução a este tema específico, apresentar-se-ão, na sequência deste texto, algumas características e resultados de um estudo preliminar que se fez e que relaciona, a um só tempo, diversos aspectos relevantes dos sistemas de avaliação até aqui discutidos. De um lado, há as metas do Ideb para as diferentes unidades educacionais, brasileiras em geral, e mineiras em particular, que se definiram com base nos resultados dessas mesmas escolas na Prova Brasil de 2005. E, de outro lado, há as considerações já levantadas anteriormente por este texto, acerca da existência de uma inevitável instabilidade das medidas educacionais, e que resultam numa flutuação dos resultados obtidos em relação a valores teóricos previstos por um modelo estatístico específico. É precisamente neste momento que entra em cena a utilização da modelagem linear hierárquica apresentada no capítulo anterior. Basicamente, será agora feita uma sequência dos seguintes passos:

Primeiro, são tecidas algumas considerações sobre o banco de dados a ser utilizado no presente estudo. E, neste sentido, já aqui aparece um detalhe relativamente inovador da metodologia ora proposta. É fato sobejamente conhecido que os sistemas estaduais de avaliação vêm se proliferando e se firmando de forma aparentemente irresistível no cenário educacional brasileiro, conforme já se comentou. E um ponto de interesse particular nesta questão é o fato de que tais sistemas estaduais de mensuração da proficiência dos alunos geralmente se pautam por técnicas psicométricas elaboradas em estreita consonância com as que, historicamente, se vêm utilizando nas testagens nacionais, como na Prova Brasil. Por exemplo, em vários casos – e o de Minas Gerais não é uma exceção –, os sistemas estaduais de avaliação utilizam, nas provas de Língua Portuguesa e Matemática, as mesmas escalas de proficiência dos exames nacionais do SAEB e da Prova Brasil. Tal efeito se obtém mediante a adoção de itens, nas provas estaduais, retirados dos testes nacionais. E assim, é possível calibrar tanto a proficiência dos alunos nos testes estaduais, como também os parâmetros dos

demais itens utilizados nestas últimas avaliações, todos expressos na mesma escala de proficiência nacional do SAEB e da Prova Brasil.

Portanto, um objetivo que se consegue, com a inclusão dos resultados estaduais na reformulação das metas do Ideb, é dar mais uma grande utilidade aos sistemas próprios de avaliação dos estados, ao mesmo tempo em que se logra também obter medidas mais precisas e úteis no que diz respeito ao estabelecimento das metas nacionais do Ideb.

Uma vez apresentados os dados utilizados nesse exame exploratório, passa-se então à formulação de algumas perguntas básicas que se fazem em um estudo multinível desta natureza. E, seguindo por essa mesma linha, apresentam-se também alguns modelos estatísticos principais que, a princípio, poderiam servir para se responder às questões assim formuladas.

A seguir, em meio às possibilidades de modelagem apresentadas, decide-se por empregar um modelo longitudinal hierárquico específico para modelar o crescimento das médias, baseado em equações simples nas quais o tempo decorrido das séries históricas em questão é modelado de maneira linear. Com isso, é possível estimar, para cada escola pública de Minas Gerais, uma tendência linear e específica de crescimento, tanto em relação às suas médias de português, quanto de matemática.

Uma vez obtidos os valores previstos para os resultados médios de cada escola em ambas essas disciplinas, passa-se então a estimar as distâncias entre esses valores previstos pelo modelo e os valores reais obtidos nas avaliações. Neste ponto, introduz-se também a questão do tamanho das escolas, do ponto de vista da quantidade de alunos nelas matriculados. Como se sabe sobejamente em estatística, a estabilidade das estimativas de um dado parâmetro, como uma média, por exemplo, tende a variar inversamente com o tamanho das amostras. Ou seja, quando as amostras são muito pequenas – como, no caso dessas escolas, compostas de um número muito reduzido de alunos – então as estimativas de desempenho tendem a oscilar com uma maior amplitude, quando comparadas com casos em que essas amostras são maiores.

No entanto, como adiante se verá, há dois fatos notáveis que se percebem em relação às escolas de Minas quanto a esse aspecto: Um deles é que, com efeito, as oscilações encontradas são, mesmo, consideravelmente grandes para os casos das escolas menores. No entanto, ainda para os casos de escolas consideravelmente maiores – como as que têm, por

exemplo, médias de alunos avaliados anualmente que chegam à casa das centenas –, mesmo assim, a volatilidade não tende a desaparecer, mas, antes, estabiliza-se em patamares consideravelmente altos, pelos quais as médias dos erros das estimativas – ou seja, dos desvios entre os valores previstos e os realmente obtidos – ainda permanecem na casa dos dois dígitos, quando medidos nas escalas do SAEB para esses exames.

Neste ponto, entra em cena um fator crucial nesta argumentação, que se articula a partir de duas constatações. A primeira delas, já mencionada, é o fato de que as metas do Ideb, calculadas para todas as edições bianuais da Prova Brasil a partir de 2007, têm como base apenas os resultados da Prova Brasil de uma única edição do teste, a de 2005. E a segunda é a observação de que, via de regra, existe uma distância – às vezes, bastante considerável – entre o resultado real obtido por uma escola numa dada avaliação, e o resultado para ela previsto com base numa série histórica. Portanto, quando se traçam metas com base somente num único resultado de um teste, caso o erro estatístico dessa medida seja considerável – seja para mais, ou para menos –, poderá ocorrer um efeito altamente indesejável, das metas assim traçadas se situarem muito acima ou abaixo daquilo que, na prática, se poderia esperar das escolas, quando se analisam seus resultados durante um período de tempo maior.

Precisamente nesta linha é que então se apresentam os dados seguintes deste capítulo. Primeiramente, são mostrados os resultados das escolas em termos do cumprimento ou não das metas do Ideb para as duas edições para as quais tais metas foram calculadas: 2007 e 2009. E, a seguir, tenta-se relacionar esse sucesso ou fracasso ao erro estatístico da medida de proficiência em 2005 dessas mesmas escolas, descrito anteriormente. Constata-se, então, uma forte associação entre as essas duas medidas, que é quantificada por meio de uma regressão logística, que modela a chance de um dado evento acontecer ou não (DEMARIS, 1992). No presente caso, temos como variável dependente o logaritmo da chance de uma escola fracassar no cumprimento de suas respectivas metas, e como variável independente, o erro médio entre os valores reais e os previstos pela reta de regressão, para o ano de 2005. Passemos, portanto, a detalhar cada um desses passos no restante deste capítulo.

#### Os dados utilizados

O presente capítulo, como se disse inicialmente, tem por objetivo apresentar os resultados de uma análise exploratória, relacionada ao estudo do problema da volatilidade dos

resultados das avaliações educacionais em grande escala. Para este propósito, utilizar-se-á uma base de dados com a série histórica de Língua Portuguesa e Matemática do SIMAVE, que cobre todas as suas edições entre os anos 2000 e 2011, bem como os resultados da Prova Brasil dos anos de 2005, 2007 e 2009.<sup>7</sup>

A tabela a seguir apresenta o número de testes nessas duas disciplinas aplicados em todas as edições do teste, para o conjunto dos três anos de escolarização por ele considerados: 5º e 9º anos do ensino fundamental e 3º ano do ensino médio. Como já se observou, o SIMAVE iniciou-se no ano 2000, apresentando, em geral, edições anuais voltadas para a avaliação de Português e Matemática. As exceções a esse padrão foram os anos de 2001, quando se aplicaram testes de ciências naturais e sociais – não considerados na presente análise –, 2002, quando só houve o teste de Língua Portuguesa; 2003, quando só houve o de Matemática, e 2004 e 2005, quando não se aplicaram testes. Observa-se que a quantidade total de testes aplicados nessas duas disciplinas ultrapassa a casa dos 10 milhões.

Tabela 7.1: Número de testes de Português e Matemática aplicados no SIMAVE em todas as suas edições, conjuntamente para as três séries consideradas (5º e 9º anos EF e 3º ano EM), entre 2000 e 2011.

<b>ANO</b>	<b>PORT</b>	<b>MAT</b>	<b>TOTAL</b>
2000	490.808	484.295	975.103
2002	467.297	0	467.297
2003	0	459.253	459.253
2006	642.683	625.025	1.267.708
2007	673.031	662.530	1.335.561
2008	662.066	647.313	1.309.379
2009	675.229	669.990	1.345.219
2010	763.823	669.436	1.433.259
2011	764.613	744.245	1.508.858
	5.139.550	4.962.087	10.101.637

Fonte: CAEd, 2012

A Tabela 7.2, a seguir, divide os totais da tabela anterior entre as redes de ensino às quais pertencem as escolas avaliadas ao longo desse mesmo período. Observa-se que quase três quartos dos testes foram aplicados a alunos de escolas estaduais, e o restante, aos de escolas municipais.

<sup>7</sup> Em outras partes deste texto, produzidas posteriormente à ocasião em que se realizou o estudo descrito no presente capítulo, foram utilizados dados do SIMAVE e da Prova Brasil que também incluíram edições posteriores desses testes, que vão até o ano de 2013, inclusive. Entretanto, as conclusões obtidas com base nos resultados aqui apresentados referem-se ao comportamento típico do banco de dados, e não dependem, portanto, do fato do ano final da série considerada ser 2011 ou 2013, ou outro ano qualquer.

Tabela 7.2: Quantidade total de testes de Português e Matemática aplicados por rede de ensino no SIMAVE 2000-2011.

<b>Rede</b>	<b>N</b>	<b>%</b>
<i>Estadual</i>	7.351.106	72,8
<i>Municipal</i>	2.750.531	27,2
<i>Total</i>	10.101.637	100

Fonte: CAEd, 2012

## Níveis e resultados globais de desempenho no SIMAVE

A tabela a seguir apresenta a convenção adotada para se delimitarem os níveis de proficiência nas diversas séries e disciplinas avaliadas no SIMAVE. Cabe dizer que esses níveis foram criados especificamente para os propósitos deste estudo, não sendo, necessariamente, os oficiais utilizados na produção dos resultados do SIMAVE em seus resultados mais recentes, os quais tendem a dividir o espectro da proficiência em quatro níveis. Por outro lado, os três níveis aqui adotados tornaram algumas análises mais fáceis de serem realizadas e comunicadas, sendo esta a principal razão de sua adoção.

Tabela 7.3: Delimitação dos níveis de proficiência no SIMAVE 2000-2011

<b>ANO</b>	<b>NÍVEL</b>	<b>PORT</b>	<b>MAT</b>
<i>5 EF</i>	1 (insuficiente)	abaixo de 175	abaixo de 175
	2 (intermediário)	de 175 a 225	de 175 a 225
	3 (recomendável)	acima de 225	acima de 225
<i>9 EF</i>	1 (insuficiente)	abaixo de 200	abaixo de 225
	2 (intermediário)	de 200 a 275	de 225 a 300
	3 (recomendável)	acima de 275	acima de 300
<i>3 EM</i>	1 (insuficiente)	abaixo de 250	abaixo de 250
	2 (intermediário)	de 250 a 300	de 250 a 300
	3 (recomendável)	acima de 300	acima de 300

Apresentamos, a seguir, alguns resultados da proficiência, calculados para cada edição do SIMAVE, e referentes a cada uma das séries consideradas. Desse modo, são mostradas as médias e desvios-padrão globais, além do percentual de alunos com proficiência se enquadrando nos três níveis acima delimitados.

Tabela 7.4: Resultados globais de Português no SIMAVE 2000-2011.

<b>ANO</b>	<b>5</b>	<b>9</b>	<b>12</b>
<i>Média</i>	199,6	246,0	273,1
<i>Desvio-padrão</i>	45,9	44,8	46,6
<i>% no nível 1</i>	30,2	15,4	29,8
<i>% no nível 2</i>	40,9	58,1	40,8
<i>% no nível 3</i>	29,0	26,5	29,5
<i>Número de al. aval.</i>	2.027.846	1.870.682	1.241.022

Fonte: CAEd, 2012

Tabela 7.5: Resultados globais de Matemática no SIMAVE 2000-2011.

<b>ANO</b>	<b>5</b>	<b>9</b>	<b>12</b>
<i>Média</i>	210,2	251,4	278,3
<i>Desvio-padrão</i>	51,4	50,1	52,8
<i>% no nível 1</i>	26,2	29,7	64,7
<i>% no nível 2</i>	34,8	53,7	32,2
<i>% no nível 3</i>	39,1	16,6	3,1
<i>Número de al. aval.</i>	2.010.428	1.803.896	1.147.763

Fonte: CAEd, 2012

## 7.2. OS RESULTADOS DA 8ª SÉRIE/9º ANO DO ENSINO FUNDAMENTAL

O presente estudo irá se deter nos resultados da 8ª série/9º ano do Ensino Fundamental. O principal propósito desta seleção é evitar uma, até certo ponto, desnecessária repetição de procedimentos e análises, visto que os padrões gerais observados numa série específica em boa medida tendem a se repetir nas demais séries. Outro motivo que levou à escolha do 9º ano EF foi o fato de que, nele, ambos os tipos de redes de ensino (estadual e municipal) aparecem bem representadas, ao passo que, nos demais casos, há uma preponderância muito elevada de uma rede sobre outra (no 5º ano EF, há um predomínio significativo de escolas municipais em relação às estaduais, e o contrário se passa no 3º EM). E como um dos objetivos do estudo era detectar possíveis diferenças de desempenho entre as redes, o 9º ano EF tornou-se, portanto, a opção mais indicada.

Apresentamos, a seguir, as estatísticas descritivas de proficiência divididas por ano de edição do SIMAVE.

Tabela 7.6: Resultados de Português para o 9º ano EF por ano de edição do SIMAVE (2000-2011)

<b>EDIÇÃO</b>	<b>Média</b>	<b>DP</b>	<b>% nível 1</b>	<b>% nível 2</b>	<b>% nível 3</b>	<b>N</b>
2000	238,3	44,5	20,1	59,1	20,8	192.654
2002	241,9	39,2	12,3	69,1	18,6	213.986
2006	242,2	48,0	19,2	55,2	25,6	230.654
2007	239,7	46,4	20,1	57,0	23,0	233.786
2008	249,0	41,9	12,8	59,9	27,3	235.033
2009	249,9	44,2	13,5	57,0	29,6	241.204
2010	253,8	44,9	11,9	54,9	33,2	266.184
2011	252,1	46,0	13,8	53,7	32,5	257.181

Fonte: CAEd, 2012

Tabela 7.7: Resultados de Matemática para o 9º ano EF por ano de edição do SIMAVE (2000-2011)

<b>EDIÇÃO</b>	<b>Média</b>	<b>DP</b>	<b>% nível 1</b>	<b>% nível 2</b>	<b>% nível 3</b>	<b>N</b>
2000	229,1	46,6	47,4	45,8	6,8	194.990
2003	241,9	49,2	37,0	51,0	12,0	200.875
2006	245,5	51,2	33,4	52,5	14,1	224.140
2007	249,5	50,3	31,4	52,6	16,0	230.257
2008	253,9	50,3	28,7	53,2	18,1	230.164
2009	258,4	48,6	23,7	57,0	19,2	239.466
2010	266,4	47,6	18,8	56,7	24,4	230.986
2011	262,3	46,5	20,4	59,2	20,4	253.018

Fonte: CAEd, 2012

Quando se separam as informações por rede de ensino e ano de edição dos testes, percebe-se que a Rede Estadual apresenta médias consistentemente superiores às da Rede Municipal, embora a distância entre essas médias não seja, aparentemente, muito elevada. Posteriormente nesta análise, a significância desses resultados será devidamente considerada.

Tabela 7.8: Resultados de Português por rede de ensino e ano de edição do teste – 9º ano EF – SIMAVE (2000-2011)

<b>EDIÇÃO</b>	<b>REDE</b>	<b>MÉDIA</b>	<b>DP</b>	<b>N_1</b>	<b>N_2</b>	<b>N_3</b>	<b>N</b>
2000	<i>E</i>	238,3	44,5	20,1	59,1	20,8	192.654
2003	<i>E</i>	241,9	39,2	12,3	69,1	18,6	213.986
2006	<i>E</i>	242,7	47,8	18,8	55,4	25,9	181.660
2006	<i>M</i>	240,1	48,9	20,8	54,6	24,6	48.994
2007	<i>E</i>	240,6	46,1	19,4	57,2	23,4	178.578
2007	<i>M</i>	236,8	47,1	22,3	56,3	21,5	55.208
2008	<i>E</i>	250,2	41,6	12,0	60,0	28,0	178.421
2008	<i>M</i>	245,2	42,7	15,4	59,8	24,8	56.612
2009	<i>E</i>	252,1	43,5	12,0	57,1	31,0	175.395
2009	<i>M</i>	243,9	45,7	17,4	56,6	25,9	65.809
2010	<i>E</i>	255,6	44,2	10,8	54,9	34,4	192.649
2010	<i>M</i>	248,8	46,3	14,8	55,2	30,0	73.535
2011	<i>E</i>	253,9	45,4	12,7	53,7	33,7	184.607
2011	<i>M</i>	247,2	47,2	16,9	53,8	29,3	72.574

Fonte: CAEd, 2012

Tabela 7.9: Resultados de Matemática por rede de ensino e ano de edição do teste – 9º ano EF – SIMAVE (2000-2011)

EDIÇÃO	REDE	MÉDIA	DP	N_1	N_2	N_3	N
2000	E	229,1	46,6	47,4	45,8	6,8	194.990
2003	E	241,9	49,2	37,0	51,0	12,0	200.875
2006	E	246,3	51,1	32,7	52,8	14,5	176.552
2006	M	242,3	51,6	35,7	51,5	12,8	47.588
2007	E	250,9	50,1	30,4	53,0	16,6	175.598
2007	M	245,1	50,5	34,7	51,3	14,0	54.659
2008	E	255,8	49,9	27,3	53,9	18,8	174.585
2008	M	247,9	51,1	33,1	51,2	15,7	55.579
2009	E	261,4	47,6	21,4	58,0	20,6	174.310
2009	M	250,2	50,0	29,9	54,4	15,7	65.156
2010	E	268,9	47,0	17,3	56,9	25,8	169.316
2010	M	259,8	48,8	23,1	56,2	20,7	61.670
2011	E	264,1	46,0	19,0	59,7	21,3	182.994
2011	M	257,2	47,4	24,0	58,1	17,9	70.024

Fonte: CAEd, 2012

*Perguntas iniciais do estudo multinível*

Apresentamos, a seguir, algumas perguntas fundamentais que nos dão um ponto de partida para a realização de um estudo multinível como este:

- 1) Haverá diferenças significativas de rendimento médio entre as escolas públicas de Minas Gerais no 9º ano do Ensino Fundamental, particularmente no ano mais recente de sua edição (2011)?<sup>8</sup>
- 2) Haverá, entre as escolas do SIMAVE, uma tendência significativa de variação das médias de proficiência ao longo do tempo? Em caso afirmativo, seria possível estimar seu valor?
- 3) No caso de haver tendências significativas de variações das médias de proficiência, as escolas do SIMAVE se diferem significativamente entre si quanto a essas taxas de crescimento? Em outras palavras, haverá escolas cujo desempenho está crescendo mais rapidamente do que outras?

<sup>8</sup> O motivo de se considerar um ano específico – aqui, 2011 – para a análise dessas diferenças de desempenho entre as escolas se deve ao fato de que, numa regressão multinível longitudinal – conforme no presente caso e, como de resto, em qualquer análise temporal de regressão –, é preciso especificar para qual instante (ou, neste caso, para qual ano) serão calculados os interceptos das retas de regressão de cada escola, sendo que esse intercepto representa a média da escola prevista para o ano considerado. Ora, as escolas municipais somente começaram a ser sistematicamente avaliadas pelo SIMAVE a partir de 2006. Por essa razão, caso o intercepto das retas de regressão correspondesse ao ano inicial da série histórica do SIMAVE (2000), a interpretação dos interceptos para as escolas municipais, embora, teoricamente, possível, padeceria de falta de realismo. Por esse motivo, optou-se por, quanto a este aspecto específico, analisar as séries “de trás para frente”, fazendo com que o intercepto fosse levado para o ano final da série aqui considerada – 2011, neste caso –, fazendo com que, para esse último ano, houvesse em geral valores reais tanto para as escolas estaduais quanto para as municipais.

### *As Variáveis:*

Neste estudo, serão utilizadas as seguintes variáveis, especificadas por nível:

#### Nível 1:

**MEDIA:** proficiência média da escola em Língua Portuguesa ou Matemática, mensurada em sua escala original (SAEB)<sup>9</sup>.

**EDICAO:** ano da edição do SIMAVE

**N:** número de alunos avaliados por escola numa disciplina (LP ou MAT)

#### Nível 2:

**REDE:** rede de ensino (Estadual ou Municipal)

### *Os Modelos*

Na modelagem, foram testadas três possibilidades de crescimento, frequentemente encontradas na literatura internacional sobre estudos longitudinais hierárquicos (BRYK, 2002; SINGER, 2003), e que foram as seguintes:

#### *Modelo 1: Linear*

Neste caso, o primeiro nível é o tempo (medido em anos e correspondente às diversas edições do SIMAVE) e o segundo nível é a escola. Este modelo pode ser expresso pelas seguintes equações:

---

<sup>9</sup> Em geral, variáveis quantitativas contínuas, como a proficiência nos testes, costumam ser padronizadas para se rodar uma regressão, visto que tal procedimento tem a vantagem de unificar a interpretação dos coeficientes das diferentes variáveis. Entretanto, no presente caso, como há um interesse específico pela própria interpretação das escalas do SIMAVE, achamos mais conveniente rodar a regressão nas escalas originais, pois isto favorecerá, assim o esperamos, a compreensão dos resultados por parte de quem já esteja habituado a trabalhar com resultados do SIMAVE e da Prova Brasil (dois testes que adotam as mesmas escalas de Matemática e Português).

Nível 1:

$$Y_{it} = \pi_{0i} + \pi_{1i} a_{it} + e_{it}$$

onde:

$Y_{it}$ : proficiência média (em Matemática ou Português) da escola  $i$  no instante  $t$

$\pi_{0i}$ : proficiência final média (ou seja, no ano 2011) da escola  $i$

$\pi_{1i}$ : taxa de crescimento da escola  $i$  ao longo do período considerado

$a_{it}$ : momento da avaliação da escola  $i$ , onde  $t$  pode assumir valores correspondentes aos anos das diversas edições do SIMAVE, entre 2000 e 2011

$e_{it}$ : erro idiossincrático da média da escola  $i$  no momento  $t$ , sendo que os erros são independentes e possuem uma distribuição normal com uma variância comum  $\sigma^2$ .

O que esta última equação quer dizer é que a média  $Y_{it}$  de proficiência de uma escola  $i$  num instante  $t$  sofre uma variação temporal que pode ser descrita ao menos razoavelmente por uma linha reta cujo coeficiente linear  $\pi_{0i}$  é a média (prevista pelo modelo) da escola em 2011 e o coeficiente angular  $\pi_{1i}$  (ou inclinação da reta) corresponde à taxa média anual de variação da proficiência média da escola.

O coeficiente linear  $\pi_{0i}$  passou, no modelo, a representar o resultado médio da escola em 2011 mediante uma recodificação da variável EDICAO, pela qual se obteve uma variável transformada de modo a ter o zero correspondente a 2011. Teria sido, aparentemente, mais natural adotar o zero, ou início da série, no ano de 2000, que correspondeu à primeira edição do SIMAVE. Entretanto, conforme há pouco se comentou, tal procedimento teria o inconveniente de não representar um dado real para uma grande quantidade de escolas que não tomaram parte da primeira avaliação, sobretudo devido ao fato de que muitas escolas cobertas pelo SIMAVE atualmente são municipais e, portanto, pertencentes a uma rede que não havia sido avaliada nas primeiras provas desse sistema (as redes municipais somente começaram a participar do SIMAVE a partir de 2006).

Quanto à inclinação da reta de regressão  $\pi_{1i}$ , ela representa a taxa de variação anual das médias de cada escola. Portanto, o modelo adotado permite que cada escola tenha a sua própria equação, cujos parâmetros foram calculados com base num processo de centralização das variáveis nos seus respectivos grupos, e não na grande média amostral. Essa centralização

nos grupos é a mais indicada neste caso, visto que se está justamente interessado em detectar possíveis diferenças entre as escolas quanto às suas respectivas taxas de evolução da proficiência.

Nível 2:

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + r_{1i}$$

onde:

$\beta_{00}$ : grande média das proficiências médias finais em 2011

$r_{0i}$ : erro idiossincrático da proficiência média final da escola  $i$  em 2011

$\beta_{10}$ : grande média das taxas de crescimento

$r_{1i}$ : erro idiossincrático da taxa de crescimento da escola  $i$ .

Como se pode ver das equações acima, no nível 2, ou da escola, os dois coeficientes beta do nível 1 são deixados variar, permitindo, assim, que o modelo capture as diferenças idiossincráticas das escolas em relação às grandes médias amostrais do desempenho em 2011 ( $\beta_{00}$ ) e da taxa de variação temporal ( $\beta_{10}$ ). Para isso, cada escola apresenta um resíduo em relação, ou à proficiência média global em 2011 ( $r_{0i}$ ), ou em relação à média global dos coeficientes angulares de regressão ( $r_{1i}$ ).

### *Modelo 2: Quadrático*

Nível 1:

$$Y_{ii} = \pi_{0i} + \pi_{1i} a_{ii} + \pi_{2i} a_{ii}^2 + e_{ii}$$

Nível 2:

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + r_{1i}$$

$$\pi_{2i} = \beta_{20} + r_{2i}$$

Os comentários feitos para o modelo linear permanecem, em grande parte, válidos para o caso da opção quadrática, em termos do significado e interpretação das variáveis e dos coeficientes. No nível 1, a diferença notável do modelo quadrático está na introdução de um termo associado a uma possível aceleração das taxas de evolução das médias de proficiência, aceleração esta que, a princípio, tanto pode corresponder a um aumento da velocidade quanto também o contrário, ou seja, a uma atenuação da variação. Matematicamente, esse efeito é expresso por meio da introdução do termo quadrático na equação de Nível 1. Já no Nível 2, o das escolas, a forma das equações continua como antes, apenas havendo, agora, o acréscimo de uma equação para o coeficiente  $\pi_{2i}$  de aceleração, e cuja interpretação é análoga à dos demais casos.

### *Modelo 3: Intervalar (piecewise)*

Neste terceiro modelo, ocorre uma volta ao padrão linear, porém por meio da divisão da série temporal em intervalos, com cada um destes últimos apresentando uma inclinação, em geral, significativamente diferente das demais.

No nível 1, sua equação é dada por:

$$Y_{it} = \pi_{0i} + \sum \pi_{ki} a_{kti} + e_{it}$$

onde as variáveis  $a_{kti}$  apresentam uma composição diagonal como a do exemplo a seguir, com uma sequência temporal de três pontos:

Tabela 7.10: Valores das variáveis temporais no modelo intervalar

<i>Tempo</i>	$a_{k1i}$	$a_{k2i}$	$a_{k3i}$
<i>1</i>	1	0	0
<i>2</i>	1	1	0
<i>3</i>	1	1	1

No nível 2, as equações têm a mesma forma das equações do mesmo nível nos modelos anteriores. A diferença, entretanto, é que, agora, a quantidade de parâmetros aumenta consideravelmente, em decorrência das diversas inclinações diferentes contidas no mesmo modelo.

## Considerações sobre o ajuste

O ajuste de cada modelo pode ser mensurado através do valor da sua respectiva estatística de *deviance*, sendo que, quanto menor é esta última, melhor é o ajuste. Portanto, foi necessário observar as deviances de todos os modelos, e que foram as seguintes:

Tabela 7.11: Valores de deviance para os diferentes modelos considerados

<i>Modelo</i>	<i>Deviance</i>
<i>Linear</i>	240.307.672.384
<i>Linear com REDE no nível 2</i>	240.230.426.087
<i>Quadrático</i>	241.894.563.242
<i>Quadrático com REDE no nível 2</i>	241.504.189.443
<i>Intervalar ou piecewise</i>	239.141.234.920

Como se vê acima, os modelos considerados ainda contaram com algumas variantes, relacionadas à inclusão ou não, no nível 2 (escola), da variável REDE (Estadual ou Municipal). Observa-se que a inclusão de REDE melhora o ajuste, conforme se vê pela diminuição dos respectivos valores de deviance em relação aos modelos nulos no nível 2. Isso, de certa forma, é previsível, visto que as redes tendem a apresentar diferenças entre as médias de desempenho, conforme já se explicitou ao se tratar da parte descritiva desses dados.

Observa-se que o modelo com o pior ajuste (maior deviance) foi o quadrático. Por sua vez, o modelo linear apresentou uma deviance menor que a do quadrático, ao passo que o modelo *piecewise* apresentou a menor deviance de todas. Baseando-nos apenas neste critério, teríamos que optar aqui pelo modelo *piecewise*. Entretanto, a escolha de um modelo nessas circunstâncias não é função apenas do seu respectivo ajuste, mas, antes, de um certo equilíbrio entre o poder explicativo do modelo de um lado (representado pelo seu ajuste) e a sua respectiva fidedignidade, do outro, visto que esta tende a diminuir à medida que se aumenta o número de parâmetros nas equações da regressão. Dessa forma, no modelo *piecewise* aqui considerado, haveria somente duas observações – correspondentes às proficiências médias de uma escola em duas edições consecutivas do teste – para cada intervalo específico, o que invalida, neste caso, a conveniência de sua adoção. Ficou-se, assim, com o modelo linear, que representa uma ferramenta consideravelmente mais poderosa, visto que utiliza toda a série temporal a fim de obter estimativas para somente dois parâmetros, que são os coeficientes linear e angular da reta de regressão calculada para cada escola.

Para o modelo linear e incondicional (sem a variável REDE) no nível 2, os resultados são:

Tabela 7.12: Estimativas dos efeitos aleatórios – Modelo linear, nulo no nível 2 – 9º ano EF – SIMAVE

Efeito aleatório	LP			MAT		
	Componente da		valor- <i>p</i>	Componente da		valor- <i>p</i>
	DP	Variância		DP	Variância	
INTERCEPTO, <i>r<sub>0</sub></i>	1.444.410	20.863.195	<0.001	1.680.818	28.251.502	<0.001
INCLINACAO, <i>r<sub>1</sub></i>	116.504	135.733	<0.001	151.079	228.249	<0.001
nivel-1, <i>e</i>	1.166.717	13.612.281		1.116.696	12.470.105	

Neste ponto, cabe lembrar a primeira pergunta que se fez antes de se rodarem os modelos multinível aqui apresentados:

*Haverá diferenças significativas de rendimento médio entre as escolas públicas de Minas Gerais no 9º ano do Ensino Fundamental, particularmente no ano mais recente de sua edição (2011)?*

A resposta a esta pergunta, em relação a ambas as disciplinas (PORT e MAT) é fornecida na linha do efeito aleatório do intercepto, onde se informa que são extremamente baixos os valores *p* para o teste de que as variâncias desses resultados sejam nulas. Em outras palavras, há uma variância não-nula entre as escolas públicas de Minas Gerais, tanto para os resultados de Português quanto para os de Matemática, o que implica que algumas delas se saíram significativamente melhor no teste de 2011 do que outras.

A segunda pergunta formulada foi:

*Haverá, entre as escolas do SIMAVE, uma tendência significativa de variação das médias de proficiência ao longo do tempo? Em caso afirmativo, seria possível estimar seu valor?*

Esta pergunta pode ser fornecida pelos resultados dos parâmetros fixos da regressão multinível, que são os seguintes:

Tabela 7.13: Estimativas dos efeitos fixos – Modelo linear, nulo no nível 2 – 9º ano EF – SIMAVE

	PORT		MAT	
<i>Efeito fixo</i>	<i>Coefficiente</i>	<i>valor-p</i>	<i>Coefficiente</i>	<i>valor-p</i>
<i>Para INTERCEPTO1, <math>\pi_0</math></i>				
<i>INTERCEPTO,</i> $\beta_{00}$	243.569.308	<0.001	251.758.780	<0.001
<i>Para INCLINACAO, <math>\pi_1</math></i>				
<i>INTERCEPTO2,</i> $\beta_{10}$	1.730.964	<0.001	3.252.343	<0.001

Observa-se, na última linha da tabela, para Língua Portuguesa, que a média das inclinações vale aproximadamente 1,7, e que esse coeficiente é significativamente diferente de zero, devido ao seu baixíssimo valor  $p$ . Dessa forma, conclui-se que as escolas no SIMAVE estão, em média, acrescentando quase dois pontos à sua respectiva média esperada de proficiência em português a cada edição anual do teste. Por sua vez, a grande média estadual corresponde, neste modelo, a aproximadamente 243 pontos na escala de proficiência do SAEB. Em Matemática, a cada ano as escolas ganham, em média, 3,2 pontos. Entretanto, vale lembrar que os valores de português e matemática estão expressos em escalas de proficiência diferentes, uma para cada disciplina, de modo que estes números, ao menos na forma como aqui estão, não são diretamente comparáveis entre si.

A terceira pergunta era:

*No caso de haver tendências significativas de variações das médias de proficiência, as escolas do SIMAVE se diferem significativamente entre si quanto a essas taxas de crescimento?*

Na tabela das variâncias acima apresentada, é possível ver, na linha do parâmetro de inclinação, que a resposta a esta pergunta é positiva, tanto para português quanto para matemática: as escolas do SIMAVE diferem-se umas das outras quanto à taxa de evolução de suas respectivas médias de português, visto que, novamente, é aproximadamente nulo o valor  $p$  para o teste de que essas taxas sejam iguais.

### A partição da variância dos resultados

As diferenças observadas entre as médias das escolas ao longo de todo o período do SIMAVE podem ser atribuídas a duas fontes possíveis de variação, segundo o modelo aqui adotado. Uma delas se relaciona à própria diferença de desempenho médio entre as escolas, e a ela está associado o termo aleatório  $r_0$  na equação de nível 2. A outra fonte de variação ocorre dentro de cada escola, estando associado às flutuações de suas médias em relação à linha temporal prevista para a sua evolução. A essa fonte, está associado o termo de erro  $r$  na equação de nível 1.

Observando os números das variâncias desses dois termos na sua respectiva tabela, concluímos que cerca de 60% das variações observadas entre as médias das escolas se devem a diferenças entre os seus interceptos, e quase todo o restante das variações decorre do surgimento de resultados médios que fogem ao que o modelo que havia sido previsto para cada escola com base na sua respectiva evolução temporal.

### A introdução da variável REDE no nível 2

Conforme já se comentou, a variável REDE foi posteriormente adicionada ao nível 2, para modelar tanto a média de proficiência das escolas quanto a sua taxa de evolução temporal. As tabelas a seguir apresentam os principais resultados deste modelo.

Tabela 7.14: Componentes da variância: modelo linear com Rede no nível 2 – SIMAVE – 9º ANO EF

LP				MAT			
Componente da				Componente da			
Efeito aleatório	DP	Variância	valor-p	DP	Variância	valor-p	
INTERCEPTO, $r_0$	1.440.027	20.736.770	<0.001	1.614.083	26.052.646	<0.001	
INCLINACAO, $r_1$	115.914	134.360	<0.001	0.70564	0.49793	<0.001	
nível-1, $e$	1.165.949	13.594.372		1.271.543	16.168.210		

Observa-se que, embora tenha havido a introdução da variável REDE no nível 2, a queda da variância inexplicada do intercepto  $r_0$  foi de apenas 0,6%, aproximadamente, em Língua Portuguesa:

$$\text{QUEDA DA VARIÂNCIA} = \frac{\text{VAR}(r_0) \text{ EM 1} - \text{VAR}(r_0) \text{ EM 2}}{\text{VAR}(r_0) \text{ EM 1}}$$

$$= \frac{20.863.195 - 20.736.770}{20.863.195} \approx 0,6\%$$

Em Matemática, esse efeito foi mais de dez vezes maior, correspondendo a, aproximadamente, 7,8% da variação observada. A conclusão a que se chega com isto é que as escolas mineiras, em relação ao 9º ano EF, são consideravelmente semelhantes entre si quanto aos resultados brutos em Língua Portuguesa, independentemente de serem da rede estadual ou municipal. Já nos resultados de Matemática, ocorre uma diferenciação maior entre as duas redes, embora ainda continue a haver também uma enorme superposição de resultados, ou seja, alunos com desempenho similar em ambas as redes públicas, o que talvez não se constitua numa surpresa.

Cabe frisar aqui que, quando se fala de uma considerável igualdade entre as redes, isto, naturalmente, não quer dizer que haja igualdade dentro de cada rede. Ao contrário, conforme já se mostrou, foi possível constatar que existe, de fato, uma considerável variação de resultados entre as escolas públicas, sejam elas estaduais ou municipais.

### 7.3. ANÁLISE DOS RESÍDUOS: O EFEITO PARTICULAR DA ESCOLA

Ao se rodarem os modelos, foi possível também obter os resíduos dos resultados, que correspondem às distâncias entre os valores previstos pelo modelo e os respectivos valores reais. Dessa forma, no nível 1 (o do tempo), calcularam-se resíduos indicando a distância entre a nota média obtida pela escola  $i$  no tempo  $t$ , e o seu respectivo valor previsto pelo modelo linear com a variável REDE modelando o nível 2, da escola. Neste segundo nível, por sua vez, os resíduos corresponderam à diferença entre as médias das escolas na avaliação do SIMAVE de 2011 (a edição mais recente da série) e a grande média das escolas neste mesmo teste.

Um resíduo pode ser positivo, nulo ou negativo. É positivo quando o valor real se encontra acima do valor previsto. Neste caso, pode-se dizer que o modelo subestimou a nota média da escola, prevendo para ela um valor menor do que aquele que realmente se obteve. Inversamente, um resíduo pode ser negativo quando o valor real se encontra abaixo do valor

previsto. Neste caso, houve uma superestimação da média da escola. E, por fim, naturalmente, o resíduo é zero quando os valores real e previsto da média escolar são coincidentes. Este é um caso de ajuste perfeito do modelo. Cabe dizer que, nos modelos empregados, há a suposição de que os resíduos se concentrem ao redor da linha de previsão segundo uma distribuição normal e de variância constante ao longo de sua própria escala. Estes são, respectivamente, os pressupostos da normalidade e da homocedasticidade herdados da teoria clássica da testagem, e que nos modelos lineares hierárquicos são observados, não no nível global da amostra, mas no nível de cada unidade de interesse (neste caso, no nível de cada escola). (O'CONNELL, 2008)

#### *O desvio-padrão dos resíduos*

A fim de sintetizar uma estimativa da oscilação de cada escola em torno da sua média prevista, foi calculado o desvio-padrão dos resíduos de cada uma delas. Dessa forma, tem-se, assim uma medida que, *grosso modo*, significa a quantidade de pontos na escala de proficiência que a escola em média varia, para mais ou para menos, já descontados os efeitos da sua evolução temporal e também da rede de ensino à qual ela pertence.

O desvio-padrão dos resíduos, neste caso, é uma medida natural da precisão das estimativas. Menores valores de desvios-padrão indicam a existência de menores resíduos e, conseqüentemente, de uma maior precisão do modelo. Também se pode dizer que os maiores desvios-padrão indicam uma maior volatilidade dos resultados.

Os dados descritivos a seguir referem-se somente às escolas que, no decorrer da série histórica do SIMAVE aqui considerada, foram testadas, ao menos, três vezes.

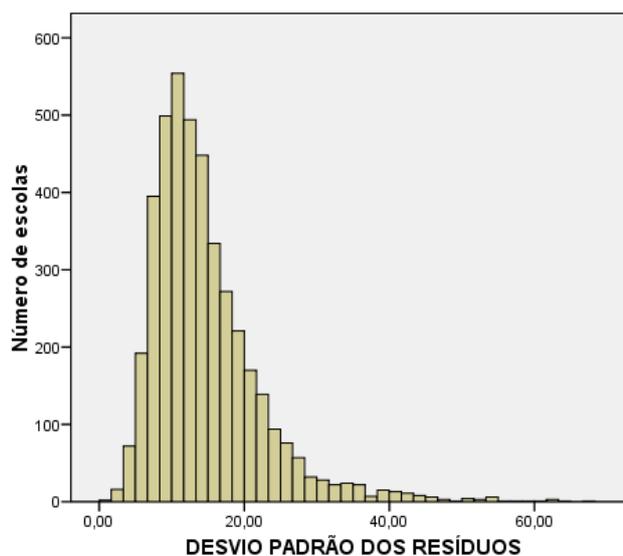


Gráfico 7.1: Desvios-padrão dos resíduos das médias escolares de proficiência em Matemática no SIMAVE

Tabela 7.15: desvios-padrão dos resíduos das médias escolares de proficiência em Matemática e Português no SIMAVE

		MAT	PORT
<i>N</i>	<i>Válidos</i>	4249	4252
	<i>Faltantes</i>	0	0
<i>Média</i>		14,7	12,0
<i>DP</i>		7,6	6,3
<i>Mínimo</i>		1,6	0,8
<i>Máximo</i>		68,2	78,0
<i>Percentis</i>	25	9,7	7,8
	50	13,0	10,4
	75	17,7	14,5

Os dados mostram que as escolas públicas mineiras, no 9º EF, variam aleatoriamente, em média, 12 pontos para mais ou para menos na escala de Língua Portuguesa, em relação à sua respectiva média prevista pela evolução temporal de seus próprios resultados. Em Matemática, essa variação média chega a quase 15 pontos da escala de proficiência.

O número de alunos avaliados por escola, série e edição do teste

A volatilidade dos resultados pode, em parte, ser explicada pela quantidade de alunos avaliados. Escolas muito pequenas estão mais sujeitas a apresentarem médias extremas, seja para mais ou para menos. Este é um motivo pelo qual é comum, ou, pelo menos, recomendável, que, nos sistemas de avaliação educacional, somente sejam considerados os resultados de escolas em que um número mínimo de alunos, como 20 ou 30, foram avaliados. Cabe, entretanto, dizer que, no presente caso, estamos, na verdade, não nos referindo ao número oficial de estudantes nessas escolas, determinado pela quantidade de matrículas, mas sim ao número de alunos que fizeram o teste do SIMAVE, o que, aliás, é precisamente o valor que interessa a um estudo de volatilidade como este. Em relação a esse número, a população de escolas mineiras tem a seguinte distribuição no 9º ano EF:

Tabela 7.16: Número médio de alunos avaliados por escola no SIMAVE, no período 2000-2011 – 9º EF – Matemática

<i>N</i>	<i>Válidos</i>	4249
	<i>Faltantes</i>	0
<i>Média</i>		59,0
<i>DP</i>		44,7
<i>Mínimo</i>		2,7
<i>Máximo</i>		479,4
<i>Percentis</i>	25	23,7
	50	48,7
	75	82,2

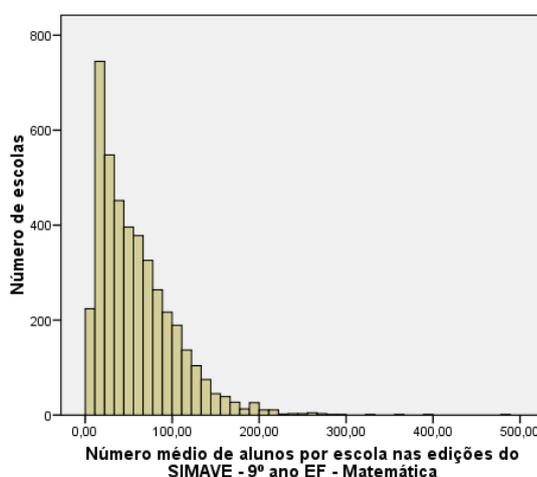


Gráfico 7.2: Número médio de alunos avaliados por escola no SIMAVE, no período 2000-2011 – 9º EF – Matemática

Percebe-se que, em geral, as escolas no SIMAVE têm, em média, cerca de 60 alunos avaliados a cada ano<sup>10</sup>. Também se observa, pelo gráfico, que a distribuição possui uma forte assimetria à direita, pela qual ocorre uma maior concentração de escolas menores na população.

#### Uma categorização de escolas quanto ao tamanho

A ocorrência desta distribuição das escolas quanto ao número de alunos avaliados permite-nos, para fins de descrição e análise dos dados, dividir esta escala em três níveis, que são os seguintes:

Tabela 7.17: Tipo de escola quanto ao número médio anual de alunos avaliados no SIMAVE, no período 2000-2011 – 9º EF

<b>Escola</b>	<b>N médio de al. avaliados</b>	<b>N</b>	<b>%</b>	<b>% cumul.</b>
<i>Pequena</i>	até 40	1750	41,2	41,2
<i>Média</i>	entre 40 e 80	1297	30,5	71,7
<i>Grande</i>	acima de 80	1205	28,3	100
<i>Total</i>		4252	100	

A seguir, cabe investigar como a volatilidade dos resultados se apresenta para os diferentes tamanhos de escolas. Para isso, para cada um dos três grupos de escolas – pequenas, médias e grandes –, obteve-se a média dos seus respectivos desvios-padrão dos resíduos, que correspondem à medida de volatilidade aqui considerada. Os resultados obtidos foram:

Tabela 7.18: Média dos desvios-padrão dos resíduos por tamanho de escola no SIMAVE, no período 2000-2011 – 9º EF

<i>Escola</i>	<b>Média dos desvios-padrão dos resíduos</b>	
	<b>MAT</b>	<b>PORT</b>
<i>Pequena</i>	18,4	15,6
<i>Média</i>	11,8	9,5
<i>Grande</i>	12,1	9,3

<sup>10</sup> Este valor foi obtido a partir do número de alunos avaliados em Matemática. Entretanto, os resultados para a presença de alunos nos testes de Língua Portuguesa foram muito semelhantes a estes, não sendo aqui apresentados para evitar um certo grau de repetição de informações.

Observa-se, com os resultados acima que, em Português, as escolas pequenas no SIMAVE têm uma variação aleatória média superior a 15 pontos da escala de proficiência em relação à respectiva sua trajetória de crescimento prevista pelo modelo. Já entre as escolas médias e grandes, o efeito é aproximadamente o mesmo em média, correspondendo a cerca de 9 pontos na escala. Em Matemática, ocorre algo semelhante, no sentido em que as menores escolas apresentam as maiores flutuações – 18 pontos em média –, ao passo que as escolas de porte médio e grande tiveram uma flutuação média próxima de 12 pontos da sua respectiva escala.

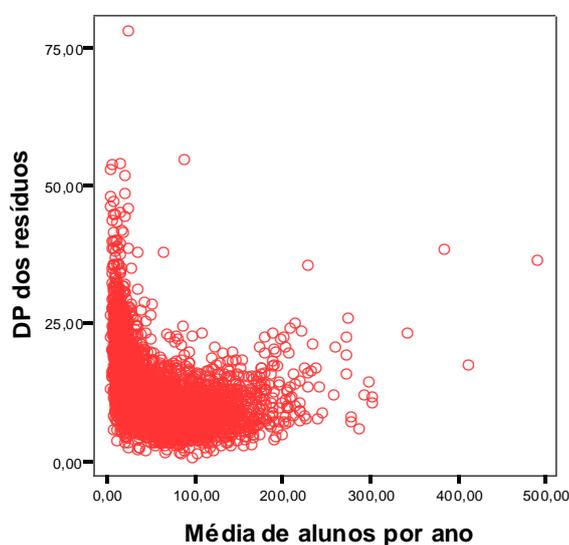


Gráfico 7.3: Resíduos das médias de português em relação ao número médio de alunos avaliados nas escolas do SIMAVE.

#### 7.4. REGRESSÃO DOS RESÍDUOS DAS MÉDIAS EM RELAÇÃO AO NÚMERO MÉDIO DE ALUNOS AVALIADOS

O gráfico de dispersão acima apresenta a distribuição dos desvios-padrão dos resíduos das médias de português em função do tamanho das escolas, medido pelo número médio de alunos avaliados a cada edição do SIMAVE. Embora o padrão seja não linear, ele também é claramente descendente. Isso permite que façamos uma estimativa aproximada da associação entre a volatilidade dos resultados médios das escolas e o seu número médio de alunos avaliados, utilizando para isso uma regressão linear clássica. A tabela abaixo apresenta estes resultados:

Tabela 7.19: Resultados da regressão dos desvios-padrão dos resíduos sobre o tamanho das escolas no SIMAVE, no período 2000-2011 – 9º EF

		Coef. não-padronizado		Coef. Padronizado		
		B	Std. error	Beta	t	Sig.
PORT	Intercepto	14,798	0,151		98,169	.000
	Inclinação	-0,046	0,002	-0,342	-23,727	.000
MAT	Intercepto	17,488	0,186		94,264	0
	Inclinação	-0,048	0,003	-0,281	-19,071	0

Observa-se, por esta última tabela, que, em média, um aluno a mais avaliado em cada escola faz, *grosso modo*, a flutuação das notas médias dessa escola cair cerca de um vinte avos de ponto de proficiência aproximadamente, tanto em matemática quanto em português. Na verdade, entretanto, o efeito é muito mais intenso do que isso na região das menores escolas – que perfazem, aliás, a maioria da população –, segundo o padrão não-linear já comentado a partir do último gráfico acima.

#### *Questões a serem levantadas sobre a volatilidade encontrada*

Na divulgação dos resultados das avaliações em grande escala, é comum as escolas receberem boletins informando a média que cada uma delas obteve naquela edição do teste, quase sempre também com uma menção ao desempenho médio em anos anteriores. No entanto, há que se mencionar que tal prática possui uma série de limitações, ou mesmo problemas, relacionados à produção de informações válidas para diagnosticar o desempenho dessas escolas.

Vimos que, mesmo levando-se em conta a evolução temporal das médias das escolas – por meio de uma modelagem linear hierárquica longitudinal que leva consideração a evolução de cada escola individualmente, e não apenas uma tendência geral de crescimento –, ainda assim, os resíduos que se observam nas médias escolares são consideravelmente elevados, sendo da ordem de uma dezena de pontos em média, em qualquer das escalas de proficiência consideradas (matemática e português). E tal fato é, como já se devia esperar, ainda mais acentuado nas escolas menores (com 40 alunos ou menos, em média).

Em relação às escolas de porte médio e grande, também é possível tirar duas outras conclusões: uma é que o efeito da volatilidade ainda existe para elas, e está longe de ser

desprezível; outra é que, ao menos em geral, não há grandes diferenças quanto a esse efeito entre as escolas médias e as grandes.

Diante desse quadro, fica difícil afirmar que, digamos, um aumento na nota média de uma escola de um ano para outro, ainda que aparentemente significativo, tenha sido o resultado de mudanças em alguma dimensão (ou mais de uma delas) da escola que tenha impactado o aprendizado dos alunos. Isto porque o erro da medida é muito maior do que o tamanho dos efeitos, quaisquer que sejam eles, que se querem identificar. E, se essa situação já gera um desconforto considerável no que diz respeito aos estudos especificamente acadêmicos de eficácia escolar, o que não dizer quando, aos resultados das avaliações, começam a se atrelar consequências práticas derivadas dos diversos esforços de prestação de contas que vêm acontecendo nos últimos anos, no Brasil e no mundo? É, portanto, exatamente um desses casos que examinaremos mais detidamente a seguir.

#### *7.4.1. A fixação de metas do Ideb e o problema da volatilidade dos resultados*

Conforme já se disse anteriormente nesta exposição, as metas do Ideb foram determinadas com base num modelo de regressão logística que levou em conta os resultados de fluxo e proficiência das escolas no ano 2005, tendo esta proficiência média sido mensurada pela Prova Brasil desse mesmo ano. Segundo esse modelo, às escolas que apresentaram um desempenho médio menor em 2005, foram apresentadas metas menores para 2007, devido ao fato de que, neste caso, a curva de regressão passava, em 2005, por um ponto mais baixo, quando comparado às escolas que, nessa mesma edição do teste, tiveram resultados melhores.

Ora, segundo o que acabamos de detalhar quantitativamente nos dados do SIMAVE, o fato de uma escola ter um desempenho pior ou melhor num determinado ano é em grande parte influenciado por variáveis sobre as quais a escola possui pouco ou nenhum controle, especialmente no modelo transversal de avaliação que se costuma empregar nas avaliações do Brasil, e de que são exemplos a maioria dos testes estaduais em grande escala, além da própria Prova Brasil realizada pelo INEP. Portanto, um ponto que merece ser considerado nos estudos quantitativos de eficácia escolar é a associação entre essas flutuações aleatórias das médias e as suas implicações para a determinação das metas do Ideb.

### *O banco de dados*

Para realizar este estudo, foi montado um banco de dados que se gerou a partir do SIMAVE, sendo, portanto, o mesmo utilizado nas análises descritas até aqui neste capítulo. Entretanto, a ele foram inseridos os resultados das escolas na Prova Brasil de 2005 (lembrando que, neste ano, não houve avaliação do SIMAVE em nenhuma das duas disciplinas consideradas), além dos resultados, também das escolas na Prova Brasil, de 2007 e 2009, que substituíram os resultados do SIMAVE para estes mesmos anos. Essa mescla de resultados, a rigor, não representou um grande problema, devido ao fato de que os testes do SIMAVE têm seus resultados calibrados, precisamente, segundo os parâmetros das escalas da Prova Brasil. Por outro lado, a introdução das notas da Prova Brasil permitiu que se traçasse uma linha evolutiva da média das escolas levando em conta todos os resultados já disponíveis para essa avaliação, acrescidos dos dados das mesmas escolas recolhidos por um sistema de avaliação praticamente análogo ao federal.

Com isso, foi possível rodar o mesmo modelo longitudinal de crescimento linear e com a variável REDE no nível 2 – das escolas –, que possibilitou também a obtenção dos resíduos das escolas em cada edição da Prova Brasil ou do SIMAVE. Mais particularmente, neste caso, nosso interesse se concentrou nos resíduos das médias escolares na Prova Brasil de 2005, devido ao papel decisivo que as médias das escolas nesse ano tiveram sobre o cálculo das suas respectivas metas do Ideb. O raciocínio por trás desse procedimento é simples, e pode ser expresso por meio da seguinte pergunta:

Considerando que as escolas públicas mineiras apresentam desvios aleatórios significativos entre suas médias reais e as suas respectivas médias previstas pelo modelo de crescimento longitudinal, haverá então uma associação significativa entre esses desvios observados em 2005 – ano de fixação das metas do Ideb – e o fato das escolas atingirem ou não suas metas do Ideb em anos posteriores?

### *O cumprimento das metas do Ideb em 2007 e 2009*

A tabela a seguir mostra o percentual das escolas consideradas na amostra que se encontraram abaixo de suas metas para o Ideb nos anos de 2007 e 2009.

Tabela 7.20: Escolas públicas de Minas Gerais que se encontraram abaixo da meta do Ideb em 2007 e 2009  
*Abaixo da meta:*

Ano		Não	Sim	Total vál.	Faltantes
2007	<i>N</i>	1856	818	2674	31
	<i>% vál.</i>	69,4	30,6	100	
2009	<i>N</i>	1990	608	2598	107
	<i>% vál.</i>	76,6	23,4	100	
<i>ambos</i>	<i>N</i>	2213	370	2583	122
	<i>% vál.</i>	85,7	14,3	100	

Observa-se, portanto, que cerca de 30% das escolas públicas mineiras deixaram de atingir a meta do Ideb para 2007 no 9º EF, e que esta proporção reduziu-se para pouco menos de um quarto na avaliação seguinte – a de 2009. Por outro lado, aproximadamente uma em cada sete escolas não atingiu a meta em nenhuma das duas vezes.

*A divisão das escolas segundo o erro médio em 2005: diferentes pontos de partida para o estabelecimento das metas*

Conforme vimos, a distribuição das médias reais de proficiência das escolas é grande o suficiente para permitir a ocorrência relativamente frequente de grandes desvios em relação aos valores previstos. Entretanto, em relação especificamente ao Ideb, cabe comentar que os desvios de Matemática e de Português não podem ser analisados separadamente, mas sim em conjunto, visto que ambos entram simultaneamente no cálculo das variações desse índice. Uma proposta aparentemente válida seria a de se tirar a média dos desvios nessas duas disciplinas. Assim, por exemplo, se uma escola teve um determinado resíduo positivo na nota de português, mas teve um resíduo negativo na de matemática, o efeito médio dessas variações sobre o Ideb seria atenuado, porque tais flutuações estariam ocorrendo em sentidos opostos. Entretanto, cabe dizer que este caso *não* é o mais frequente. O que se percebe, em geral, é uma elevada correlação *positiva* entre os resultados das duas disciplinas, de modo que as escolas que obtêm médias maiores em matemática, também obtêm, em geral, médias maiores em português, e algo semelhante se passa com as escolas de desempenho mais baixo. Portanto, poder-se-ia usar a média simples dos dois resíduos para exprimir o resíduo médio de cada escola. Este, de fato, foi o procedimento aqui adotado, com somente uma modificação: no nosso caso, calculou-se, na verdade, a *média padronizada dos resíduos padronizados*,

tendo a padronização dos resíduos sido feita devido ao fato de que as escalas das duas disciplinas são diferentes entre si.

A seguir, dividiu-se a população de escolas em quatro grupos de mesmo tamanho e dispostos em ordem ascendente de resíduos. Dessa forma, os grupos 1 e 2, chamados respectivamente de “Negativos Acentuados” e “Negativos”, correspondem às escolas que tiraram uma nota média real abaixo daquela que havia sido prevista para elas em 2005. A do grupo “Negativos Acentuados” corresponde, assim, ao primeiro quarto da distribuição. Nelas, portanto, as escolas ficaram muito abaixo do que deveriam ter ficado segundo o modelo de crescimento aqui proposto. Por outro lado, há também os grupos 3 e 4, e que correspondem respectivamente aos resíduos “Positivos” e “Positivos Acentuados”, e que indicam as escolas cujas médias em 2005 ficaram acima dos seus respectivos valores previstos, correspondendo o último grupo à parte superior da distribuição de resíduos.

Tabela 7.21: Tipos de escolas públicas mineiras quanto aos resíduos de sua média medidos em relação aos valores previstos da suas respectivas médias na Prova Brasil de 2005

<b>Resíduo</b>	<b>Limites do desvio médio</b>	<b>Número de escolas</b>
<i>Negativo Acentuado</i>	Mínimo e 1º quartil	676
<i>Negativo</i>	1º e 2º quartis	677
<i>Positivo</i>	2º e 3º quartis	676
<i>Positivo Acentuado</i>	3º quartil e máximo	676

#### *A associação entre os resíduos e o cumprimento das metas*

A tabela a seguir indica o percentual de escolas que não conseguiu atingir suas respectivas metas do Ideb. Há colunas para os anos de 2007 e 2009 separadamente, e também para o caso do não-cumprimento ter se verificado em ambos os anos.

Tabela 7.22: Percentual das escolas que não atingiram a meta do Ideb, em associação ao tipo de resíduo de suas médias na Prova Brasil de 2005

<b>Resíduo</b>	<b>2007</b>	<b>2009</b>	<b>Ambos os anos</b>
	(%)	(%)	(%)
<i>Negativo</i>	14,2	9,1	3,0
<i>Acentuado</i>			
<i>Negativo</i>	22,1	14,3	7,2
<i>Positivo</i>	32,3	24,5	12,5
<i>Positivo</i>	53,7	45,5	34,5
<i>Acentuado</i>			

Observa-se, por exemplo, que, se, do ponto de vista geral, 30% das escolas haviam falhado em atingir suas metas em 2007, ocorre, por outro lado, acentuadas diferenças de desempenho observado quando essa informação é cruzada com o desvio sofrido pela média da escola em 2005, ano de estabelecimento das metas. Das escolas que tiveram um resíduo negativo acentuado, o percentual das que não atingiram a meta em 2007 foi consideravelmente baixo, quando comparado com o dos demais grupos: apenas cerca de uma em cada sete escolas desse grupo não alcançou suas metas. Entretanto, nesta mesma edição da Prova Brasil, o grupo das escolas cujo desempenho mais havia crescido em 2005, em relação às suas médias previstas, foi aquele em que se verificaram taxas altíssimas de não cumprimento das metas: mais da metade das escolas desse grupo não atingiram o Ideb para elas estipulado em 2007.

Algo semelhante, embora um pouco menos acentuado, em termos de percentuais de falhas em atingir o Ideb, se passa em 2009 e também no caso do não-cumprimento das metas em ambos os anos. Em todos esses casos, os dados parecem indicar evidências muito fortes de que uma parte considerável do sucesso das escolas em alcançar o Ideb *não* está dependendo de variáveis que estejam sob o controle da escola, mas sim de flutuações estatísticas. E estas flutuações, por sua vez, encontram-se em grande parte condicionadas ao tamanho das escolas, visto que, se elas já são consideráveis mesmo nas escolas de porte médio e grande, tornam-se ainda mais destacadas nas escolas pequenas.

### *Regressão*

Finalizando esta análise, é possível precisar ainda mais o impacto do resíduo das médias observadas sobre o resultado do Ideb por meio de uma regressão logística, onde a variável dependente é o logaritmo da chance de cada escola *não* alcançar a meta para o Ideb, e

a variável independente é, novamente, a média dos resíduos em ambas as disciplinas em 2005 por escola. A tabela a seguir resume os principais resultados dessa regressão.

Tabela 7.23: Efeito do resíduo das médias da escola sobre a chance dela não conseguir atingir seu Ideb.

ANO	COEF.	B	Sig.	Exp(B)
2007	INCLIN.	0,829144	,000	2,291357
	INTCPT	-0,93694	,000	0,391824
2009	INCLIN.	0,936902	,000	2,552063
	INTCPT	-1,39309	,000	0,248306
ambos	INCLIN.	1,191061	,000	3,29057
	INTCPT	-2,21628	,000	0,109014

Observa-se, pela tabela acima, que todos os coeficientes da regressão são significativos. Uma informação particularmente relevante encontra-se na última coluna – Exp(B) – e que, nas linhas das inclinações, significa a variação nas chances de uma escola *não* atingir o Ideb associada a um aumento de um desvio-padrão na variável do desvio entre as notas reais e as previstas em 2005 na Prova Brasil. Dessa forma, percebe-se que, para cada desvio-padrão a mais que a escola avança quanto à volatilidade de seus resultados médios, suas chances de *não* atingir a meta do Ideb em 2007 aumentaram cerca de 2,3 vezes. Esse efeito ainda subiu ligeiramente em 2009 – para aproximadamente 2,6 – e chegou a cerca de 3,3 vezes nos casos das escolas que não atingiram o Ideb em nenhuma das edições.

#### *Algumas conclusões acerca do desempenho das escolas públicas mineiras no 9º ano EF*

- 1) As escolas públicas de Minas Gerais apresentam uma considerável variedade de resultados nos testes de proficiência em grande escala aplicados atualmente (SIMAVE e Prova Brasil), tanto na disciplina de Língua Portuguesa quanto de Matemática.
- 2) As escolas vêm, paulatinamente, aumentando suas médias de proficiência ao longo das edições desses testes.
- 3) O desempenho médio das escolas estaduais é um pouco superior ao das escolas municipais, sendo esse resultado relativamente mais acentuado em Matemática. Apesar disso, entretanto, há também uma grande superposição dos resultados, o que torna essas duas redes razoavelmente semelhantes entre si, particularmente quanto à distribuição da proficiência de Língua Portuguesa.

- 4) Mesmo quando se controla a média das escolas pela sua evolução temporal e pela rede à qual elas pertencem, sobra ainda uma grande quantidade de variação do resultado que se associa a fatores não explicados (não incluídos no modelo).
- 5) A média dessa flutuação dos resultados das escolas alcança, em geral, dois dígitos nas escalas de proficiência de português e matemática utilizadas. Ao mesmo tempo, observa-se que essa volatilidade é maior nas escolas menores (com até 40 alunos avaliados por ano, em média), embora ela continue sendo significativa e aproximadamente igual para as escolas de porte médio e grande.
- 6) Um problema decorrente da volatilidade é que ela compromete a validade do acompanhamento anual do desempenho das escolas, quando este é realizado, ao menos parcialmente, com base nas médias de proficiência alcançadas pela escola nas últimas avaliações. Desse modo, fica difícil ou mesmo impossível, no caso das escolas públicas mineiras consideradas no presente estudo, associar as flutuações observadas em suas respectivas médias a fatores que estão, ao menos teoricamente, sob o seu controle, como, por exemplo, a competência de seus respectivos professores e diretores.
- 7) Outro problema decorre do fato de que essa volatilidade pode ter consequências consideráveis e indesejadas nas situações em que se fixam metas de desempenho para sistemas ou unidades educacionais com base nos resultados de um ano específico de uma avaliação em grande escala. Nesse sentido, os erros ou resíduos presentes podem provocar uma considerável subestimação ou superestimação do desempenho médio de uma escola. Por sua vez, tais erros podem ter consequências graves quando, com base no desempenho da escola num ano específico, se traçam metas para essas escolas. No caso da superestimação, as escolas se veem prejudicadas pela imposição de metas mais imediatas excessivamente elevadas; no caso da subestimação, ocorre o oposto: para as primeiras metas, os objetivos são relativamente fáceis de alcançar, não por causa de méritos intrínsecos à escola, mas sim devido ao desvio específico, aleatório e negativo que sua média sofreu por ocasião do estabelecimento das metas.
- 8) Atualmente, no Brasil, os sistemas de avaliação em grande escala correm o risco de terem a validade de suas medidas comprometidas por essa volatilidade dos resultados. E o que torna as coisas ainda piores no Brasil é que as avaliações são, via de regra, transversais, ou sejam, testam alunos diferentes em momentos diferentes, de modo que não há controles para a proficiência prévia dos estudantes. Além disso, embora em muitas dessas avaliações, ocorra

também a aplicação de questionários contextuais de alunos, escolas, professores e outros, percebe-se que muito raramente essas informações são utilizadas para fins práticos de estabelecimento de controles estatísticos no monitoramento e julgamento das avaliações.

9) Uma proposta, portanto, que se pretende abordar na continuação deste estudo, é a consideração de modelos que permitam recalculas as metas do Ideb de uma dada instituição educacional, como uma escola, com base na evolução dos seus respectivos resultados, tanto de desempenho quanto de rendimento, e não como se fez oficialmente, de se traçarem as metas com base apenas nos resultados de um ponto temporal único e específico (o ano de 2005).

## 8. UMA PROPOSTA DE RETIFICAÇÃO DAS METAS DO IDEB

### 8.1. INTRODUÇÃO

Conforme se viu no capítulo anterior, pode-se perceber pelo menos um problema consideravelmente sério na determinação das metas do Ideb. Trata-se do fato de que, no cálculo das metas, foram levados em conta somente valores pontuais – neste caso, referentes ao ano de 2005 – tanto do rendimento  $P$  das escolas como de seu desempenho em Língua Portuguesa e Matemática na Prova Brasil. Dessa forma, as metas traçadas correram o risco de, em não poucos casos, terem tido seu cálculo baseado em valores potencialmente anômalos para suas respectivas escolas. E, devido a esse efeito, poderia haver – e, de fato, há – casos de metas escolares apontando para um desempenho bastante diferente (para mais ou para menos) em relação ao que se esperaria dessas mesmas escolas, caso se tivesse como observar a evolução de seus indicadores ao longo de uma série temporal mais longa, preferencialmente que envolvesse medidas anteriores e posteriores a 2005.

Outro ponto de interesse para a presente proposta reside no fato de que, na determinação oficial, ou convencional, das metas do Ideb, em nenhum momento se considera diretamente o possível fator associado ao índice socioeconômico médio dos alunos matriculados na escola. Ora, a literatura sobre eficácia educacional, de modo regular e sistemático, aponta para a grande associação existente entre desempenho escolar e índice socioeconômico, tanto em contextos brasileiros como internacionais. Portanto, não parece, ao menos em tese, ser muito adequado, ou mesmo justo, um método de estabelecimento de metas do Ideb, ou, em geral, de qualquer outro índice de qualidade educacional, que não leve em conta a condição socioeconômica das escolas. Entretanto, o Ideb age precisamente dessa forma, e não parece haver muitos questionamentos – ao menos políticos – sobre isso...

Talvez uma alegação que se possa fazer em defesa do Ideb se baseie no seguinte raciocínio: As metas do Ideb incluem, no seu cálculo, os resultados de cada escola na Prova Brasil de 2005. Por sua vez, esses resultados, como, em geral, os de quaisquer outros testes de proficiência análogos, estão fortemente correlacionados com o índice socioeconômico das escolas (SOARES, 2013). Então, é muito provável que ocorram situações como a do tipo: uma escola mais pobre obteve, em geral, um pior desempenho na Prova Brasil de 2005, e o contrário se passou com uma escola mais rica. Portanto, a escola pobre, que teve um

desempenho mais baixo naquele ano, ficou com metas menores traçadas para o futuro, e o inverso de novo se passou com as mais ricas.

Entretanto, há que se observar que, embora seja inegável a associação entre desempenho acadêmico e índice socioeconômico dos alunos no ensino básico, ainda assim existe uma considerável parcela da variância dos resultados desses testes que não é explicada pelo índice socioeconômico, mas sim por outros fatores. Portanto, o que, na prática, esse tipo de raciocínio faz é usar o índice socioeconômico como uma espécie de *proxy* do desempenho, algo que, a nosso ver, embora possa ser razoável em algumas circunstâncias, parece ser um procedimento grosseiro demais para uma situação tão delicada quanto esta, de traçar metas institucionais das escolas ao longo de exames previstos para décadas a fio.

Assim sendo, em termos específicos, o objetivo central deste estudo é a proposição de uma nova metodologia de determinação das metas do Ideb para as escolas de Minas Gerais, que leve em conta os problemas acima levantados. E, de modo geral e análogo, tal princípio também poderia ter, pelo menos, duas outras grandes utilidades. Uma delas seria reajustar, ou, diríamos mesmo, corrigir, as metas do Ideb para outros estados que, analogamente ao caso de Minas Gerais e de seu SIMAVE, dispõem de sistemas próprios de avaliação que já vêm funcionando há vários anos, e detendo-se sobre as mesmas séries investidas pela Prova Brasil ou pelo SAEB. Ilustrações disso são os casos, por exemplo, do Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo, o SARESP, que vem sendo executado anualmente desde a década de 1990 – portanto, muito antes do surgimento do Ideb – e também do Sistema de Avaliação Permanente do Estado do Ceará, o SPAECE, entre vários outros (CAED, 2008).

Dessa forma, seria teoricamente possível, nesses casos, utilizar dados coletados nos respectivos estados, através dessas avaliações anteriores e posteriores a 2005, para que se recalculassem, de 2007 em diante, as metas de suas respectivas escolas. Além disso, também cabe dizer que o presente método propõe uma constante atualização das metas com base nos resultados mais recentes que se vão produzindo, seja oriundos dos exames nacionais da Prova Brasil, seja resultantes dos exames estaduais dos sistemas próprios de avaliação. Portanto, embora desejável, tampouco seria necessário que houvesse dados anteriores a 2005 para os estados em questão. E, indo um pouco além, também seria possível aplicar esse método mesmo no caso dos poucos estados brasileiros que não têm nem nunca tiveram um sistema

próprio de avaliação educacional. Isto porque, para fazer as atualizações, bastariam as próprias edições mais recentes da Prova Brasil.

Uma segunda grande utilidade que, a nosso ver, esse método poderia ter seria em relação ao estabelecimento de metas de outros indicadores de qualidade educacional que não apenas o Ideb. E há vários casos de estados brasileiros que propõem suas próprias metas de qualidade educacional. Um exemplo disso é o próprio estado de São Paulo, com o seu IDESP, o Índice de Desenvolvimento da Educação do Estado de São Paulo, que, analogamente ao Ideb, é formado por dois fatores referentes ao rendimento e ao desempenho de seus alunos. Cabe dizer que uma diferença notável entre o índice paulista e o nacional refere-se à mensuração do desempenho das escolas. Como se sabe, no caso do Ideb, tomam-se como referência as médias de Língua Portuguesa e Matemática das escolas obtidas na Prova Brasil. Já no caso do SARESP, o que se considera são os diferentes percentuais de alunos de uma dada escola situados em diferentes níveis da escala de proficiência adotada nos exames de São Paulo. Não obstante, tais escolhas metodológicas não chegam a comprometer a utilização do método aqui proposto, pois ele, a princípio, pode ser flexível a ponto de substituir as médias por outras medidas de desempenho, como os percentuais de alunos nos diferentes níveis da escala, à semelhança do que faz o IDESP e outros.

Em relação à metodologia ora proposta, e cujo detalhamento é fornecido adiante neste capítulo, pode-se, especificamente, elencar uma série de vantagens, a saber:

- 1) Para um grande número de escolas, a nova metodologia inibe a produção de metas inadequadas de desempenho, que parecem estar ocorrendo em duas situações extremas. De um lado, há casos de metas excessivamente elevadas para algumas escolas, o que dificulta inadequadamente, ou mesmo impossibilita, o seu respectivo cumprimento. De outro lado, há também casos de metas excessivamente baixas, que produzem um efeito contrário, qual seja, fazem com que o seu cumprimento seja inadequadamente fácil, fato que se traduz na possibilidade de haver escolas sendo consideradas como bem-sucedidas, ao mesmo tempo que seu progresso, conforme vem percebido através de uma série histórica mais estendida, não tem sido tão grande quanto esse sucesso parece indicar.
- 2) O uso desta nova metodologia faz com que, para fins de determinação das metas federais do Ideb, haja um aproveitamento de uma enorme quantidade de dados provenientes das avaliações estaduais, como é o caso, no presente exemplo de Minas

Gerais, do Sistema Mineiro de Avaliação da Educação Básica, o SIMAVE. Dessa forma, além de servirem para seus propósitos diretos e específicos – como fornecer um diagnóstico via de regra anual da situação da educação num determinado estado, e também de produzir subsídios para as políticas educacionais estaduais com base nessas avaliações, testes como o SIMAVE são também aqui considerados para aumentar a fidedignidade das metas traçadas pelo Ideb para cada escola da unidade federativa em questão.

- 3) A nova metodologia propõe um sistema “dinâmico” de determinação das metas, que se opõe ao sistema “estático” oficial e presentemente adotado. Dessa forma, a escola, a cada edição da Prova Brasil, depara-se com metas que estão mais condizentes, não apenas com o propósito final, de que todas atinjam um Ideb adequado a médio ou longo prazo, mas que também estejam ajustadas de acordo com as tendências mais recentes que se vêm observando dentro dessas mesmas escolas, tanto no que diz respeito ao fluxo como também ao desempenho do alunado.
- 4) Outra significativa vantagem, já inicialmente comentada, é que as metas assim recalculadas também passam a incorporar diretamente o índice socioeconômico médio das respectivas escolas. Dessa forma, por exemplo, quando uma escola atende alunos de origem socioeconômica mais privilegiada, ela vem a ter, segundo essa nova metodologia, metas maiores do que dela se esperaria no caso real, que não cria distinções entre escolas mais “ricas” e mais “pobres”. Inversamente, para as escolas que atendem alunos socioeconomicamente menos favorecidos, as metas de desempenho são ajustadas para valores menores do que os oficiais.

## 8.2. FÓRMULAS PARA O CÁLCULO DAS METAS AJUSTADAS DO IDEB

Para fins de proposição da metodologia apresentada no presente estudo, serão mostradas ambas as formas de cálculo das metas aqui consideradas, ou seja, a real e a recalculada segundo a nova metodologia proposta.

A meta real do Ideb é dada pela fórmula (INEP, 2012b):

*Meta real do Ideb:*

$$ideb_{it} = 10 / (1 + \exp(-(\ln(ideb_{i0}) / (10 - ideb_{i0})) + \gamma_i \cdot t))$$

onde:

$ideb_{it}$  : meta do Ideb da escola  $i$  para um período  $t$

$t$ : período ou tempo decorrido, em anos, desde o início da série em 2005

$ideb_{i0}$  : valor real inicial (em 2005) do Ideb da escola  $i$

$\gamma_i$ : coeficiente de crescimento temporal das metas da escola  $i$

A partir dessas considerações, a forma alternativa de cálculo das metas do Ideb passa a ser dada pela seguinte equação:

*Meta recalculada do Ideb:*

$$ideb_{it}^* = 10 / (1 + \exp(-(\ln(ideb_{i0}^* / (10 - ideb_{i0}^*)) + \gamma_i^* \cdot t)))$$

$ideb_{it}^*$  : meta recalculada do Ideb da escola  $i$  para um período  $t$

$t$ : período ou tempo decorrido, em anos, como antes

$ideb_{i0}^*$  : valor estimado inicial (em 2005) do Ideb da escola  $i$ , recalculado periodicamente

$\gamma_i^*$ : coeficiente de crescimento temporal das metas da escola  $i$ , recalculado periodicamente

Como se observa, a nova metodologia, a princípio, conserva intacta a forma para o cálculo das metas. Contudo, ocorrem agora duas mudanças substanciais, a saber:

Para uma dada unidade educacional – no presente caso, uma escola pública mineira –, o valor  $ideb_{i0}$  antes, representava o seu respectivo Ideb no ano de 2005, calculado tanto com base nos seus respectivos resultados médios de Língua Portuguesa e Matemática na Prova Brasil de 2005, como também nos valores dos indicadores de rendimento  $P$  tanto para a série ou ano escolar em questão – no presente caso, o 9º ano EF – como também para os anos imediatamente anteriores do mesmo ciclo de escolarização – como, para o presente caso, os valores de  $P$  também para os anos 6, 7 e 8 do Ensino Fundamental, no mesmo ano-base de 2005.

Entretanto, com a nova metodologia proposta, o Ideb “inicial”  $ideb_{i0}$ \* passa a ter um novo significado, bastante distinto do anterior. Embora ele continue se referindo a um valor do Ideb para 2005, agora não se trata mais de um resultado real e observado para este índice, porém de um valor estimado e recalculado de dois em dois anos, e que leva em conta uma série de alterações, ou, melhor dizendo, de atualizações. Estas últimas, por sua vez, são produzidas substituindo os valores reais de rendimento e desempenho em 2005 acima referidos pelos que se apresentam e discutem a seguir.

a) A partir de 2007 – ano para o qual se começaram a traçar as metas do Ideb – a “média” de Matemática  $MAT_{i,2005}$  da escola  $i$  em 2005, utilizada para se calcular a meta para 2007, não é mais simplesmente o resultado obtido por essa escola na Prova Brasil em 2005 nessa mesma disciplina. Ao invés disso, no caso das escolas públicas mineiras, ele agora passa a ser um valor estimado por uma regressão linear hierárquica longitudinal que tem a escola como unidade de medida e que, para fins de cálculo da meta para 2007, incorpora toda a série histórica disponível para a escola  $i$  até o ano de 2005, quando essa mesma meta foi traçada. Na prática, isso significa que cada uma das escolas mineiras assim consideradas teve estimado um valor para sua média de matemática em 2005 com base numa regressão sobre os resultados do SIMAVE de 2000 e 2003 e também com base nos resultados da Prova Brasil de 2005.

O modelo adotado é de dois níveis, sendo que o primeiro deles fornece a evolução temporal da média de cada escola, e o segundo nível modela os parâmetros de regressão – tanto o coeficiente linear quanto o angular – que aparecem no primeiro nível. Dessa forma, no segundo nível, ocorrem dois efeitos notáveis. O primeiro deles é que os coeficientes linear e angular passam a ter um componente aleatório, exatamente para permitir que as escolas possuam, neste modelo, diferentes médias iniciais de matemática e também diferentes taxas de sua respectiva evolução temporal. O segundo efeito é que ambos os parâmetros nesse segundo nível passam a ser modelados pelo índice socioeconômico médio da escola. Com isso, é possível “deduzir” a contribuição do índice socioeconômico do alunado, tanto em termos da média inicial de matemática da escola, como também em termos de sua respectiva taxa de crescimento. Numa linguagem matemática, este modelo se escreve como:

*Nível 1:*

$$MAT_{it} = \pi_{0i} + \pi_{1i} \cdot a_{it} + \varepsilon_{it}$$

onde:

$MAT_{it}$ : nota de matemática da escola  $i$  no tempo  $t$

$\pi_{0i}$ : média de matemática da escola  $i$  no tempo inicial

$\pi_{1i}$ : taxa de crescimento anual da média de matemática da escola  $i$

$a_{it}$ : instante  $t$  de testagem em matemática da escola  $i$

$\varepsilon_{it}$ : erro aleatório da média de matemática da escola  $i$  no tempo  $t$

*Nível 2:*

$$\pi_{0i} = \beta_{00} + \beta_{01} \cdot (ISE)_i + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} \cdot (ISE)_i + r_{1i}$$

onde, no lado direito da primeira das duas equações anteriores:

$\beta_{00}$ : média populacional de matemática no tempo inicial

$\beta_{01}$ : parâmetro populacional de associação da média de matemática com o índice socioeconômico médio das escolas

$ISE_i$ : índice socioeconômico médio da escola  $i$

$r_{0i}$ : erro aleatório da média de matemática da escola  $i$  no tempo inicial

E, no lado direito da segunda equação, também há:

$\beta_{10}$ : média populacional das taxas anuais de crescimento das médias de matemática das escolas

$\beta_{11}$ : parâmetro populacional de associação da taxa de crescimento da média de matemática com o índice socioeconômico médio das escolas

$r_{1i}$ : erro aleatório da taxa de crescimento da média de matemática da escola  $i$

Na equação do nível 1 acima, a variável temporal  $t$  mede a quantidade de anos decorridos desde a primeira avaliação de Matemática utilizada para este cálculo, e que, no caso específico aqui considerado, corresponde à edição do SIMAVE de 2000. Dessa forma, é possível estimar o valor previsto para a média de matemática da escola para qualquer instante  $t$ , com base em todas as avaliações da escola desde 2000 até o ano mais atualizado da série temporal, independentemente de se tratar do SIMAVE ou da Prova Brasil, visto que ambas essas avaliações, conforme já se disse, encontram-se na mesma escala de proficiência.

Um ponto que merece um destaque particular nessa questão é a determinação da média estimada de Matemática para a escola em 2005. Isto porque, conforme já se disse, esse resultado contribuiu diretamente para a determinação das metas da escola de 2007 em diante. Por sua vez, essa média escolar prevista de Matemática para 2005, com base nos resultados da regressão até 2005 – que aqui chamaremos de  $(MAT_{2005j})_t'$  – difere-se da média de Matemática na Prova Brasil de 2005 –  $(MAT_{2005j})_{PB}$  – segundo a equação:

$$(MAT_{2005j})_t' = (MAT_{2005j})_{PB} + \delta$$

onde o termo de erro pode assumir valores :

positivos, no caso em que o resultado na Prova Brasil de 2005 encontra-se subestimado, em relação ao que seria de se esperar para aquele mesmo ano, segundo a regressão dada;

negativos, onde aconteceria o contrário, ou seja, o resultado na Prova Brasil de 2005 encontrar-se-ia superestimado, em relação ao esperado pela mesma regressão;

nulos; onde o resultado na PB corresponderia exatamente ao valor previsto pela regressão.

Nas situações mais significativas de subestimação do resultado de Matemática, isso significa que, na Prova Brasil de 2005, a escola recebeu, nesta disciplina, uma nota consideravelmente menor do que se esperaria dela, levando-se em conta o seu desempenho médio durante um período de tempo maior, e também observando-se, a princípio, o índice socioeconômico médio da escola. Assim sendo, as metas do Ideb traçadas para essa escola, ao menos no tocante aos seus respectivos resultados de Matemática, ficarão subestimadas, ou, em outras palavras, consideravelmente mais fáceis do que deveriam ser.

Neste caso, do ponto de vista estritamente do resultado escolar, há para este estabelecimento, duas notícias, uma “boa” e outra “ruim”. A notícia ruim reside no fato de que, na Prova Brasil de 2005, a escola recebeu uma média relativamente baixa, o que talvez

tenha sido motivo de apreensão e/ou constrangimento, por exemplo, para os profissionais da educação com ela diretamente envolvidos, como professores, coordenadores, diretores, gerentes de Secretarias públicas de ensino, etc. Porém, por trás desse resultado ruim – ou, na verdade, como uma consequência dele –, há também uma notícia boa. Tal resultado modesto em 2005 fez com que as metas previstas para a escola, para qualquer ano (na “prática”, até para 2096!) foram consideravelmente mais baixas do que deveriam ser, deixando, portanto, essa escola, numa posição mais confortável quanto ao cumprimento dessas mesmas metas. Em casos assim, podem ocorrer situações nas quais o Ideb da escola cresceu o suficiente para cumprir metas relativamente fáceis de serem batidas, porém não aumentou o bastante para fazer a escola ter tido um desempenho satisfatório do ponto de vista geral de sua trajetória, ao longo dos anos avaliados. Na última seção do presente capítulo, serão apresentados alguns casos específicos de escolas que se comportaram dessa maneira.

Por outro lado, também podem acontecer – e, de fato, acontecem – casos opostos, ou seja, de superestimação das metas do Ideb. Neste caso, ocorre precisamente o contrário. Primeiro, a escola tem uma “boa” notícia: um elevado resultado na Prova Brasil, e também no Ideb de 2005. Entretanto, essa boa nova acaba se tornando um verdadeiro presente de grego: tais resultados “inflados” de 2005 produzem metas também infladas, fazendo com que estas últimas dificilmente consigam ser alcançadas, ainda que a escola demonstre estar significativamente crescendo de desempenho durante o período considerado. Também posteriormente serão apresentados casos específicos de escolas que vivenciam situações desse tipo.

Embora até o presente momento, por motivos de clareza de exposição, tenhamos nos detido no cálculo dos valores estimados para as médias de Matemática da escola ao longo da série temporal, um processo perfeitamente análogo pode ser – e, na realidade, é – aplicado ao cálculo das médias estimadas de português.

### 8.3. A ATUALIZAÇÃO DO MODELO DE DETERMINAÇÃO DAS METAS

No exemplo dado na seção anterior, tratou-se, especificamente, da estimação das médias de proficiência da escola com base nos resultados por ela obtidos até 2005, onde se começaram a traçar as metas para a edição seguinte da Prova Brasil, em 2007. E, para isso, como se viu, foram utilizados três resultados obtidos pela escola até aquele momento: os do

SIMAVE em 2000 e 2002/2003 (dependendo de se referir a português ou matemática, respectivamente), e a Prova Brasil de 2005. Considerando, porém, que se chega ao ano de 2007, faz-se então necessário recalculas as metas para 2009. Para este novo cálculo, o número de avaliações disponíveis duplica, de três para seis, sendo que as três novas avaliações correspondem às avaliações do SIMAVE de 2006 e 2007, e também a da Prova Brasil deste último ano.

Uma nova regressão linear hierárquica concebida nos mesmos moldes do caso anterior é então realizada. E, com base nela, é possível, como de hábito, obter, para cada ano desta série histórica, as diferenças entre os valores previstos e os reais em 2005 (ano de determinação das metas do Ideb), tanto para a proficiência em Matemática quanto em Língua Portuguesa. Dessa forma, o valor estimado para o desvio do resultado de Matemática na Prova Brasil de 2005 é, agora, em 2007, recalculado, com base nos dados de uma sequência maior e mais atualizada de edições dos testes censitários mineiros, e também da Prova Brasil. E, utilizando-se esses novos resultados, é possível estimar-se a meta “corrigida” da escola para 2009, e assim por diante.

#### *O tratamento do fator de rendimento $P$ na atualização das metas do Ideb*

Quanto ao outro fator que entra na equação do Ideb, correspondente ao indicador de rendimento  $P$ , o qual, conforme se viu, encontra-se diretamente associado às taxas escolares de aprovação, optou-se também neste caso por fazer uma regressão linear hierárquica que preveja a sua respectiva evolução, de um modo semelhante ao que se fez para as disciplinas de Língua Portuguesa e Matemática, definidoras do fator de desempenho  $N$ .

Entretanto, em termos metodológicos, cabe ressaltar a ocorrência de certas particularidades na determinação dos valores previstos para o indicador  $P$ . Uma delas decorre do fato que  $P$  e  $N$  são, de certo modo, de naturezas diferentes. O indicador  $P$  relaciona-se a uma medida – em última instância, à taxa de aprovação na escola – que corresponde, literalmente, a uma contagem de casos bem-sucedidos de matrículas (do ponto de vista da aprovação escolar) numa dada unidade de ensino, e a uma comparação entre este número e o total de matrículas. Embora possa haver uma certa variabilidade de resultados censitários, dependendo de algumas imprecisões no sistema de coleta ou processamento de informações, ainda assim, não se pode negar o fato de que tais medidas são, em termos de aferição de seu

valor “real”, muito mais “concretas” do que, digamos, a proficiência individual de um aluno em Matemática ou Português, ou a média de todos os alunos de uma dada escola nessas mesmas disciplinas. Ou, de forma algo um tanto quanto extrema, porém verdadeira, há que se admitir que possivelmente ocorre um considerável nível de variação nos resultados de uma mesma disciplina – de Matemática, Português, ou qualquer outra – de um mesmo aluno ou escola avaliada em dois dias diferentes, ainda que o conteúdo e o nível médio de dificuldade dos testes tenham se mantido aproximadamente iguais. Porém, caso se queira medir a taxa de aprovação de uma escola num determinado ano, uma vez que os dados para tal cálculo estejam já disponíveis, é muito provável que eles invariavelmente darão sempre o mesmo resultado, ainda que medidos em instantes ou dias diferentes.

Portanto, no caso da determinação do desempenho acadêmico  $N$ , essa maior imprecisão, ou flutuação, das medidas foi compensada pelo uso de um número maior de observações, fazendo com que, conforme já se disse, chegou-se mesmo ao caso de se empregarem duas avaliações censitárias realizadas na mesma escola e no mesmo ano: uma do SIMAVE e outra da Prova Brasil.

Já quanto ao indicador de rendimento  $P$ , em parte devido à sua característica própria acima apontada, e em parte também porque ele não se associa diretamente aos resultados das provas, que aqui são nosso principal interesse de estudo, para ele então se propôs um cálculo por regressão que levasse em conta somente os valores bianuais fornecidos pelo próprio INEP juntamente com os resultados do Ideb. Cabe enfatizar, no entanto, que tal procedimento corresponde a uma mera opção metodológica, que poderia ser, ao menos a princípio, trocada por outra similar, como, por exemplo, uma que também considerasse os dados anuais do censo escolar para essas mesmas escolas.

Porém, trabalhando com a hipótese de que a evolução temporal do indicador  $P$  para uma dada escola segue um padrão ao menos aproximadamente linear, então é possível estimar valores previstos para  $P$  para os diferentes anos da série histórica considerada. Não obstante, também é relevante ressaltar que tal método somente permite a ocorrência de diferenças entre os valores de 2005 real e previsto de  $P$ , de 2009 em diante. Isto porque, na primeira edição da série histórica de  $P$  aqui empregada, ou seja, em 2005, havia somente uma única observação, correspondente aos próprios valores de 2005. E, em 2007, passou a haver duas observações, fazendo com que tenha sido possível traçar uma reta entre esses dois valores, sendo que o erro – ou seja, a diferença entre o valor real e o previsto – é sempre zero nessa situação, porque

existem então apenas dois pontos na série temporal. Por outro lado, a partir de 2009, a série passa a se constituir de três observações, o que, novamente, permite o cálculo de uma reta de regressão para cada escola de modo a começar a haver a possibilidade da ocorrência de desvios entre essa reta e os casos reais observados.

### *O cálculo do valor estimado do Ideb em 2005*

Uma vez obtidas as estimativas referentes a 2005 para o desempenho em Língua Portuguesa e Matemática e também para o rendimento  $P$  de uma dada escola, o passo a seguir é aplicar esses resultados às fórmulas convencionais utilizadas para o cálculo do Ideb, e que são:

$$n_{ji}^{\alpha} = ((S_{ji}^{\alpha} - S_{inf}^{\alpha}) / (S_{sup}^{\alpha} - S_{inf}^{\alpha})) * 10$$

$$N_{ji} = (n_{ji}^{LP} + n_{ji}^{Mat}) / 2$$

$$Ideb = N_{ji} \cdot P_{ji}$$

No caso do 9º ano do Ensino Fundamental aqui considerado, como já se comentou anteriormente neste texto,  $S_{sup} = 400$  e  $S_{inf} = 100$ . Por sua vez, obtêm-se, pelo cálculo acima descrito, os valores estimados  $S_{ji}^{\alpha}$  para cada escola em 2005, os quais, juntamente com o valor estimado para o rendimento  $P_{ji}$ , permitem que se chegue ao valor recalculado para o Ideb da escola em 2005. E, finalmente, esse valor é então utilizado para se estimarem as metas futuras do Ideb.

#### 8.3.1. Os resultados da regressão

O uso do modelo de regressão aqui empregado para se obterem as médias estimadas de português e matemática das escolas foi então empregado por meio do pacote estatístico HLM. Em relação à aplicação desse modelo, cabe dizer que o mesmo abrangeu os resultados de 3591 escolas públicas mineiras, tanto da rede estadual quanto das municipais de ensino, as quais totalizaram um conjunto de 39.971 observações em Matemática e 39.912 em Língua Portuguesa, entre os anos de 2000 e 2011, utilizados para se refazerem as metas até 2013. Os resultados foram também calculados levando-se em conta a ponderação das escolas pelo seu respectivo número médio de alunos.

### *Estimativas de fidedignidade*

As duas tabelas a seguir apresentam as estimativas de fidedignidade para os parâmetros do intercepto e da inclinação da evolução temporal das médias de Matemática e Língua Portuguesa, respectivamente. Convencionalmente se considera que esses parâmetros – que teoricamente podem variar entre zero e um – são significativamente elevados quando assumem valores iguais ou superiores a 0,5 para os interceptos e iguais ou superiores a 0,1 para as inclinações. E, quando tal significância se verifica, isto também quer dizer que existe, entre as unidades de análise consideradas – neste caso, as escolas – uma variação considerável entre seus respectivos parâmetros.

Levando esses dados em conta, é possível perceber que as escolas em geral variam consideravelmente entre si, tanto no que diz respeito ao intercepto, ou seja, às suas respectivas médias de Matemática no início de suas respectivas séries históricas, como também no que toca à inclinação de suas respectivas retas de regressão longitudinal. Ou, dito ainda de outro modo, as escolas tendem a se diferir significativamente umas das outras quanto ao ponto de partida (ou seja, quanto às médias iniciais de Matemática) e também se diferem quanto à evolução de suas médias, havendo, portanto, escolas que progridem significativamente mais do que outras.

Um fato curioso é que, em Língua Portuguesa, as estimativas de fidedignidade tendem a ser consideravelmente mais modestas, indicando, assim, que os resultados das escolas nessas disciplinas tendem a se assemelhar mais, quando comparados com os resultados de Matemática. Muito provavelmente, isso se deve, ao menos em parte ao fato de que os resultados dos exames de linguagem – como, neste caso, de Língua Portuguesa – tendem a ser mais influenciados por outros fatores extraescolares, quando comparados com os resultados de Matemática, disciplina esta cuja formalização se associa mais estreitamente às práticas em sala de aula, conforme já se comentou anteriormente neste texto.

Tabela 8.1: Parâmetros de fidedignidade para os coeficientes da regressão em Matemática, por ano final da série histórica do SIMAVE/Prova Brasil considerada. Fonte: elaboração própria.

	2005	2007	2009	2011
<i>Intercepto <math>\pi_0</math></i>	0.607	0.497	0.560	0.658
<i>Inclinação <math>\pi_1</math></i>	0.167	0.359	0.463	0.547

Tabela 8.2: Parâmetros de fidedignidade para os coeficientes da regressão em Língua Portuguesa, por ano final da série histórica do SIMAVE/Prova Brasil considerada. Fonte: elaboração própria.

	2005	2007	2009	2011
<i>Intercepto <math>\pi_0</math></i>	0,354	0,316	0,364	0,503
<i>Inclinação <math>\pi_1</math></i>	0,005	0,219	0,29	0,423

### *Parâmetros de regressão*

As duas tabelas a seguir mostram os dados dos coeficientes de regressão, por ano final da série histórica considerada, e também para cada disciplina: Matemática e Língua Portuguesa. No primeiro caso, referente à Matemática, todos os coeficientes são altamente significativos, possuindo, em sua maioria, valores  $p$  menores do que 0,001. Na prática, isso quer dizer que todos eles são significativamente maiores do que zero. Embora tal fato possa parecer óbvio para o intercepto  $\beta_{00}$ , que se refere às médias populacionais iniciais (no ano de 2000) de Matemática, por outro lado, é significativo – porém tampouco surpreendente – observar que, de fato, essas médias iniciais estejam consideravelmente associadas ao aumento do índice socioeconômico médio do alunado. Esta última constatação pode ser percebida pelo fato de que, por exemplo, a cada desvio-padrão a mais que a escola apresenta no seu índice socioeconômico médio, em geral, essas médias esperadas de proficiência inicial em Matemática tendem a aumentar em dois dígitos na escala de proficiência considerada, que tem uma média de 250 pontos e um desvio-padrão de 50 na escala de Matemática adotada.

Por outro lado, as taxas de variação temporal das notas de Matemática também são estatisticamente significantes e positivas, indicando que, de fato, as escolas estão aumentando suas respectivas médias de Matemática ao longo do tempo. Entretanto, a associação dessa taxa temporal com o índice socioeconômico médio das escolas tem um comportamento variável. Isso porque ela começa com valores positivos e decrescentes e, continuando essa tendência, passa a assumir valores negativos. Certamente, este último comportamento está associado ao fato de que, com o tempo, tenderam a crescer mais em resultados aquelas escolas que tinham um menor índice socioeconômico médio, ao passo que as escolas de maior ISE, em geral, possuíam maiores resultados no início, e assim, tiveram menos teto para crescer.

Em Língua Portuguesa, conforme se percebe pela Tabela 8.4, o comportamento é semelhante ao de Matemática. Ocorrem flutuações nas estimativas para a média de português em 2000, dependendo da extensão da série histórica considerada. Ao mesmo tempo, observa-se uma previsível e considerável associação positiva (também de dois dígitos na escala de

proficiência adotada) entre o desempenho inicial e o índice socioeconômico médio das escolas.

Porém, quanto à taxa de crescimento dos resultados médios de português, observa-se que eles parecem ser mais modestos do que os já comentados em matemática, embora, naturalmente, seja preciso ressaltar que os números para as diferentes disciplinas encontram-se expressos em escalas diferentes.

Tabela 8.3: Parâmetros da regressão longitudinal em Matemática, por ano final da série histórica do SIMAVE/Prova Brasil considerada. Fonte: elaboração própria.

	2005	2007	2009	2011
<i>Média de profic.</i>				
<i>Intercepto <math>\beta_{00}</math></i>	228,6	228,6	227,4	226,7
<i>Inclinação (ISE) <math>\beta_{01}</math></i>	13,2	16	18,9	20
<i>Taxa de variação</i>				
<i>Intercepto <math>\beta_{10}</math></i>	3,3	2,7	3	3,2
<i>Inclinação (ISE) <math>\beta_{11}</math></i>	1,9	0,4	-0,3	-0,5

Tabela 8.4: Parâmetros da regressão longitudinal em Língua Portuguesa, por ano final da série histórica do SIMAVE/Prova Brasil considerada. Fonte: elaboração própria.

	2005	2007	2009	2011
<i>Média de profic.</i>				
<i>Intercepto <math>\beta_{00}</math></i>	239,2	235,8	230,6	228,6
<i>Inclinação (ISE) <math>\beta_{01}</math></i>	17,6	16,4	17,8	17,9
<i>Taxa de variação</i>				
<i>Intercepto <math>\beta_{10}</math></i>	-2,0	-0,1	1,4	1,8
<i>Inclinação (ISE) <math>\beta_{11}</math></i>	-0,4	0,2	-0,2	-0,3

#### *Partição da variância*

Por fim, as duas tabelas a seguir fornecem a partição da variância dos resultados observados. Um ponto relevante a se mencionar é que essa variância é calculada levando-se em conta o controle pelo índice socioeconômico. Percebe-se que, tanto em Matemática quanto em Língua Portuguesa, uma considerável parte da variação dos resultados obtidos ao longo da série é explicada pelas médias iniciais. É como se, *grosso modo*, houvesse um certo paralelismo entre as trajetórias temporais das médias das escolas individuais, com tal fato se

verificando tanto em uma disciplina como em outra. Por outro lado, a parcela da variância devida às diferenças de taxas de crescimento das escolas é muito menos significativa, quando comparada com a anterior, o que, por sua vez, tampouco deixa de ser surpreendente em casos como esse. Por fim, a chamada variância não-explicada, associada ao erro aleatório das medidas, atinge valores consideráveis.

Tabela 8.5: Estimativa dos componentes da variância em Matemática, por ano final da série histórica do SIMAVE/Prova Brasil considerada. Fonte: elaboração própria.

	2005		2007		2009		2011	
	Var.	%	Var.	%	Var.	%	Var.	%
<i>Intercepto <math>r_0</math></i>	168,9	61,4	176,2	63,0	191,1	63,6	222,6	65,5
<i>Inclinação <math>r_1</math></i>	2,1	0,8	2,9	1,0	2,5	0,8	2	0,6
<i>Nível <math>1 e</math></i>	103,9	37,8	100,4	35,9	107	35,6	115,1	33,9
<b>Total</b>	<b>274,9</b>	<b>100</b>	<b>279,5</b>	<b>100</b>	<b>300,6</b>	<b>100</b>	<b>339,7</b>	<b>100</b>

Tabela 8.6: Estimativa dos componentes da variância em Língua Portuguesa, por ano final da série histórica do SIMAVE/Prova Brasil considerada. Fonte: elaboração própria.

	2005		2007		2009		2011	
	Var.	%	Var.	%	Var.	%	Var.	%
<i>Intercepto <math>r_0</math></i>	51,8	83,3	58,7	36,3	70,4	37,7	98,6	46,5
<i>Inclinação <math>r_1</math></i>	0	0,0	1,2	0,7	1,1	0,6	1,1	0,5
<i>Nível <math>1 e</math></i>	10,4	16,7	102	63,0	115,4	61,7	112,3	53,0
<b>Total</b>	<b>62,2</b>	<b>100</b>	<b>161,9</b>	<b>100</b>	<b>186,9</b>	<b>100</b>	<b>212</b>	<b>100</b>

### 8.3.1. A evolução temporal do parâmetro $P$

Passaremos agora a apresentar e discutir os resultados da regressão longitudinal multinível que se realizou com os dados de rendimento  $P$  das escolas. Conforme já se comentou anteriormente, essa regressão foi realizada com base nos dados coletados pelo Censo Educacional e referentes à mensuração bianual do Ideb das escolas entre os anos de 2005 e 2011. Em termos de estabelecimento das retas de regressão que preveem o valor de  $P$  para um determinado ano da série histórica de cada escola considerada, o conjunto de dados para os cálculos aqui considerados somente atinge as condições mínimas para ser empregado a partir de 2009, ano correspondente à terceira mensuração desse indicador, devido ao fato de que, antes disso, somente havia dois casos observáveis até 2007, e apenas um caso em 2005. Ou seja, para a série histórica englobando as duas primeiras medições do indicador  $P$  (2005 e 2007), não existe diferença entre os valores previstos e os reais obtidos pelas escolas, conforme já se comentou anteriormente.

O modelo adotado para o cálculo dos valores estimados de  $P$  é análogo àquele que já se apresentou para as estimativas das médias de Língua Portuguesa e Matemática. Ou seja, consiste de uma modelagem linear hierárquica e longitudinal de dois níveis, controlada pelo índice socioeconômico médio e ponderada pelo tamanho das escolas, segundo as equações:

*Nível 1:*

$$P_{ti} = \pi_{0i} + \pi_{1i} \cdot a_{ti} + \varepsilon_{ti}$$

onde:

$P_{ti}$ : indicador de rendimento  $P$  da escola  $i$  no tempo  $t$

$\pi_{0i}$ : valor estimado do indicador  $P$  da escola  $i$  no tempo inicial

$\pi_{1i}$ : taxa de crescimento anual do indicador  $P$  da escola  $i$

$a_{ti}$ : instante  $t$  do cálculo de  $P$  da escola  $i$

$\varepsilon_{ti}$ : erro aleatório do valor de  $P$  da escola  $i$  no tempo  $t$

*Nível 2:*

$$\pi_{0i} = \beta_{00} + \beta_{01} \cdot (ISE)_i + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} \cdot (ISE)_i + r_{1i}$$

onde, no lado direito da primeira das duas equações anteriores:

$\beta_{00}$ : média populacional do indicador  $P$  no tempo inicial

$\beta_{01}$ : parâmetro populacional de associação da média inicial de  $P$  com o índice socioeconômico médio da escola

$ISE_i$ : índice socioeconômico médio da escola  $i$

$r_{0i}$ : erro aleatório do valor de  $P$  da escola  $i$  no tempo inicial

E, no lado direito da segunda equação:

$\beta_{10}$ : média populacional das taxas anuais de crescimento do valor de  $P$  das escolas

$\beta_{11}$ : parâmetro populacional de associação da taxa de crescimento de  $P$  com o índice socioeconômico médio das escolas

$r_{1i}$ : erro aleatório da taxa de crescimento do valor de  $P$  da escola  $i$

### *Estimativas de fidedignidade*

Pelos valores da tabela abaixo, percebe-se que existe um grau considerável de variabilidade entre as escolas públicas mineiras, tanto no que diz aos valores iniciais (em 2005) de seus respectivos rendimentos  $P$ , como também no que se refere às taxas de evolução temporal desse mesmo indicador. Tudo isso se percebe pelo fato de que os coeficientes de fidedignidade desses dois parâmetros são consideravelmente elevados, em relação aos valores mínimos de 0,5 para o intercepto e de 0,1 para a inclinação, que geralmente se prescrevem em análises desse tipo. Apenas para a regressão com dados até 2011, o valor do coeficiente de fidedignidade para o intercepto ficou um pouco abaixo do mínimo previsto, porém cabe observar que essa distância é muito pouco acentuada.

Tabela 8.7: Parâmetros de fidedignidade para os coeficientes da regressão em  $P$ , por ano final da série histórica considerada. Fonte: elaboração própria.

	2009	2011
<i>Intercepto <math>\pi_0</math></i>	0,631	0,478
<i>Inclinação <math>\pi_1</math></i>	0,278	0,367

### *Parâmetros de regressão*

Em relação aos coeficientes da regressão, dados pela tabela a seguir, cabe observar que todos eles apresentam valores  $p$  menores do que 0,001, o que mostra a sua grande significância estatística.

Observa-se então que as estimativas de  $P$  inicial variam um pouco, dependendo da extensão da série histórica que se considere (até 2009 ou 2011), mas que, em geral, situam-se em torno de 70% de aprovação. Também se observa que esse valor inicial de  $P$  encontra-se positiva e significativamente associado ao índice socioeconômico das escolas, ou seja, aqueles estabelecimentos que atendem a clientela socioeconomicamente mais favorecidas tendem também a apresentar maiores níveis de aprovação.

Por sua vez, as duas linhas finais desta tabela descrevem o comportamento geral das taxas de variação temporal desse mesmo indicador. Observa-se que ambas as taxas médias correspondem a 0,013, valor este que equivale ao incremento médio anual de  $P$  observado nas escolas da amostra em questão. E, na última linha, observa-se um fato de certo modo curioso, porém nada surpreendente, de que, enquanto o valor inicial de  $P$  encontra-se positivamente associado ao ISE (conforme comentado no parágrafo anterior), porém a variação temporal de  $P$  está negativamente associada ao mesmo índice socioeconômico médio das escolas. Ou, em outras palavras, as escolas atendendo a clientela mais pobres apresentam menores valores iniciais de  $P$ , porém, ao mesmo tempo, também possuem maiores taxas de crescimento desse mesmo índice. Por outro lado, as escolas que atendem a alunos em condições socioeconômicas melhores, possuem maiores indicadores iniciais de rendimento; não obstante, também têm esses indicadores crescendo menos, até porque encontram-se mais próximas do teto de 1, ou de 100% de aprovação, correspondente a esta medida.

Tabela 8.8: Parâmetros da regressão longitudinal em  $P$ , por ano final da série histórica considerada. Fonte: elaboração própria.

	2009	2011
<i>Média de profic.</i>		
<i>Intercepto <math>\beta_{00}</math></i>	0,748	0,681
<i>Inclinação (ISE) <math>\beta_{01}</math></i>	0,069	0,119
<i>Taxa de variação</i>		
<i>Intercepto <math>\beta_{10}</math></i>	0,013	0,013
<i>Inclinação (ISE) <math>\beta_{11}</math></i>	-0,011	-0,01

### *Partição da variância*

Por fim, apresentam-se os dados da partição da variância dos resultados, cabendo novamente lembrar que essa variância é obtida já se controlando as medidas pelo índice socioeconômico médio das escolas. Os números mostram que a grande maioria da variância

observada é estatisticamente explicada pelos valores iniciais do rendimento  $P$ , valor este que chega à casa dos 80% para a série histórica com dados até 2011. Por sua vez, a variância não explicada pelo modelo é consideravelmente menor (30% com dados até 2009, e 17% com dados até 2011) e, finalmente, é ainda muito mais modesta a variância decorrente das disparidades entre as escolas quanto às suas respectivas taxas de variação temporal de  $P$ .

Tabela 8.9: Estimativa dos componentes da variância em  $P$ , por ano final da série histórica considerada. Fonte: elaboração própria.

	2009		2011	
	Var.	%	Var.	%
<i>Intercepto <math>r_0</math></i>	0,00888	68,0	0,01930	82,3
<i>Inclinação <math>r_1</math></i>	0,00025	1,9	0,00017	0,7
<i>Nível <math>1 e</math></i>	0,00392	30,0	0,00398	17,0
<b>Total</b>	<b>0,01305</b>	<b>100,0</b>	<b>0,02345</b>	<b>100,0</b>

#### 8.4. AS DIFERENÇAS ENTRE OS DOIS MÉTODOS

Na seção anterior, apresentaram-se os principais resultados das regressões temporais das médias escolares em ambas as disciplinas, controladas pelo índice socioeconômico médio e ponderadas pelo número médio de alunos nas escolas. Com base nisso, foram então traçadas as novas metas do Ideb para as escolas, segundo o método descrito anteriormente neste capítulo.

A seguir, passaremos a considerar uma consequência da utilização desse novo método, que consiste na detecção de possíveis casos de diferença ou, também ao contrário, de semelhança de resultados quanto ao cumprimento de metas, segundo as duas diferentes metodologias empregadas para o cálculo destas últimas. Nesse sentido, nas tabelas a seguir, são apresentadas essas comparações para todos os anos da série histórica para os quais já se produziram resultados comparáveis.

Tabela 8.10: Diferenças entre os resultados do cumprimento de determinação das metas, segundo o seu respectivo método de determinação. Fonte: elaboração própria.

DECISÃO SOBRE AS METAS	2007			2009			2011			2013		
	N	% vál.	% acm.									
CUMPRIU EM AMBAS AS MEDIDAS	1695	65,6	65,6	1749	69,2	69,2	1507	60,6	60,6	1072	43,8	43,8
NÃO CUMPRIU EM AMBAS AS MEDIDAS	789	30,5	96,2	565	22,3	91,5	610	24,5	85,1	800	32,7	76,5
CUMPRIU NA ALTERN., MAS NÃO NA REAL	1	0,0	96,2	13	0,5	92,0	25	1,0	86,1	35	1,4	77,9
CUMPRIU NA REAL, MAS NÃO NA ALTERN.	98	3,8	100,0	201	8,0	100,0	345	13,9	100,0	541	22,1	100,0
Total dos casos válidos	2583	100		2528	100		2487	100		2448	100	
Casos faltantes	1024			1079			1120			1159		
TOTAL	3607			3607			3607			3607		

Os dados da tabela acima podem também ser visualizados nos gráficos de barra a seguir, sendo um para cada ano das metas traçadas até 2013. Neste caso, os números no eixo horizontal dos gráficos obedecem à seguinte convenção, a respeito da concordância quanto ao cumprimento das metas nos dois métodos por parte da escola pública mineira:

1. Cumpriu em ambas as medidas.
2. Não cumpriu em ambas.
3. Cumpriu na medida alternativa, porém não na real.
4. Cumpriu na medida real, porém não na alternativa.

Observando esses dados, constata-se, por exemplo, que, no primeiro ano das metas (2007), houve um grau aparentemente considerável de concordância – 96,2% – entre os dois métodos, sendo que cerca de dois terços das escolas públicas mineiras consideradas alcançaram suas respectivas metas por ambos os métodos, e aproximadamente 30% delas não as alcançaram em nenhuma das duas metodologias.

Dos casos discrepantes, a quase totalidade correspondeu a situações nas quais as escolas atingiram suas respectivas metas oficiais, porém não atingiram as que foram aqui recalculadas. E houve somente um caso contrário, de uma escola não ter atingido sua meta real, ao mesmo tempo em que a atingiu segundo o critério alternativo.

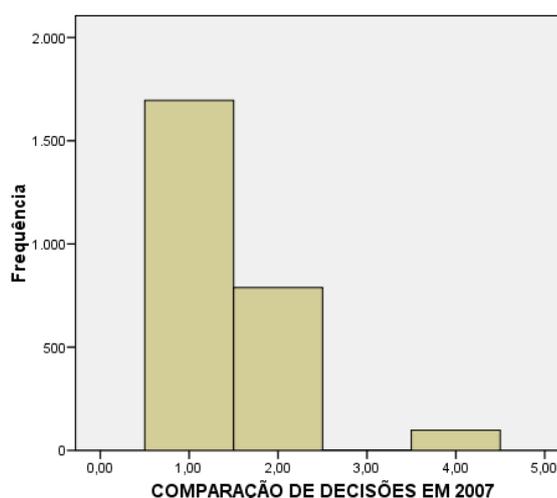


Gráfico 8.1: Diferenças entre os resultados do cumprimento de determinação das metas, segundo o seu respectivo método de determinação, no ano de 2007. Fonte: elaboração própria.

Contudo, observa-se que, com o avançar da série histórica, os níveis de concordância entre os dois métodos, apesar de ainda se conservarem relativamente elevados, vão diminuindo progressivamente, até que, em 2013, último ano da série avaliada, eles decresceram para cerca de três quartos dos casos. Ao mesmo tempo, e naturalmente, aumentam-se tanto os percentuais de casos em que a escola atinge a meta real, porém não a recalculada, quanto vice-versa, muito embora, da mesma forma que em 2007, não haja uma simetria entre esses dois casos. Isto porque continuam a ser maioria os casos das escolas supostamente beneficiadas pelo Ideb real, ou seja, estabelecimentos que cumprem as metas reais, porém não conseguem atingir as recalculadas.

Esse aumento geral da discordância entre os resultados das duas metodologias de cerca forma é esperável. Isto porque, no início da série histórica considerada, há menos tempo e menos dados para que as discordâncias se acentuem. Porém, à medida que os resultados vão se sucedendo e acumulando, podem aumentar as discrepâncias entre as duas trajetórias de metas. Tal fato, aliás, assinala um caso potencialmente preocupante, para os fins de avaliação que aqui estão sendo considerados, de que, com o tempo, progressivamente vão aumentando de forma considerável os casos de inadequação das metas do Ideb ao desenvolvimento de suas respectivas trajetórias reais.

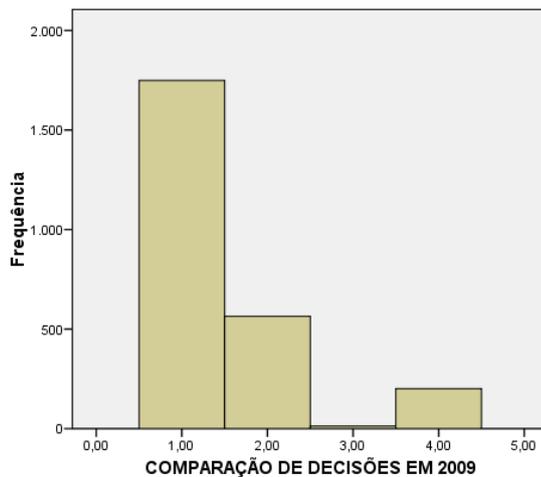


Gráfico 8.2: Diferenças entre os resultados do cumprimento de determinação das metas, segundo o seu respectivo método de determinação, no ano de 2009. Fonte: elaboração própria.

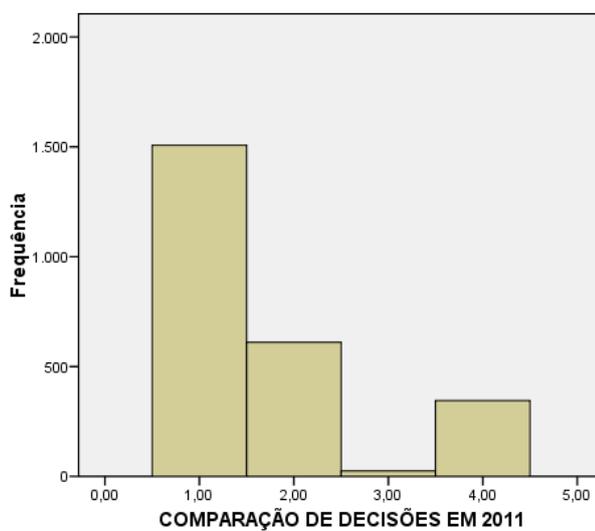


Gráfico 8.3: Diferenças entre os resultados do cumprimento de determinação das metas, segundo o seu respectivo método de determinação, no ano de 2011. Fonte: elaboração própria.

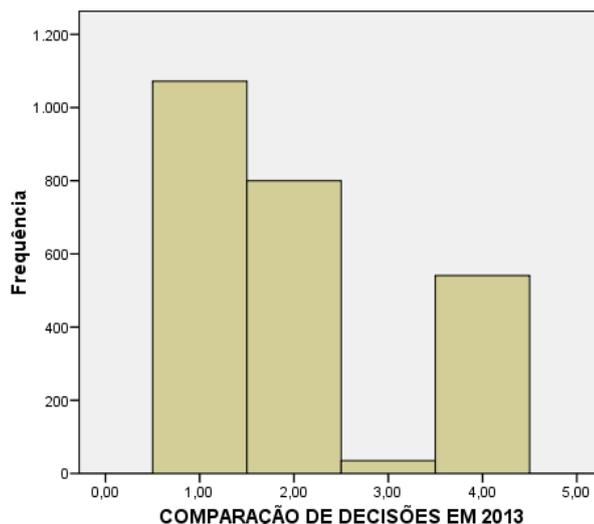


Gráfico 8.4: Diferenças entre os resultados do cumprimento de determinação das metas, segundo o seu respectivo método de determinação, no ano de 2013. Fonte: elaboração própria.

#### 8.5. CASOS ILUSTRATIVOS DE ESCOLAS AVALIADAS DIFERENTEMENTE QUANTO AO CUMPRIMENTO DAS METAS

A fim de ilustrar um pouco mais detalhadamente as implicações práticas da metodologia de determinação das metas do Ideb aqui proposta, apresentamos, a seguir, alguns exemplos de casos notáveis de escolas públicas mineiras, onde se percebe uma grande diferença entre os resultados de desempenho, em termos de cumprimento de metas, conforme auferidos pelo novo método proposto e pelo convencional.

*Ex. 1: A superestimação de metas numa escola grande, devido ao rendimento P excepcionalmente alto em 2005*

Começaremos com um exemplo no qual a escola em questão, ao longo de todas as quatro edições das metas, não as cumpriu nenhuma vez, muito embora, segundo as novas metas por nós recalculadas, teve um percentual total de cumprimento de 50%, valor que, embora não pareça corresponder a um desempenho brilhante, por outro lado, é muito mais aceitável que o desempenho nulo dos cálculos oficiais.

Trata-se da Escola Estadual Geraldo Teixeira da Costa, do município de Santa Luzia, pertencente à Superintendência Regional de Educação (SRE) Metropolitana, ou de Belo Horizonte C. Um ponto que já se salientou no presente trabalho é o fato de que as metas oficiais traçadas pelo MEC, e atual e oficialmente vigentes, têm a desvantagem de se basear em apenas um “instantâneo” da série histórica das diferentes unidades educacionais avaliadas,

qual seja, as medidas de desempenho e rendimento calculadas somente no ano de 2005. Isso sem também se levar em conta que tais metas oficiais, em seu cálculo, não consideraram explicitamente o índice socioeconômico dos alunos ou outra variável de contexto para serem calculadas.

Devido a esses problemas, conforme temos comentado, pode ocorrer, por exemplo, casos em que, precisamente no ano de 2005, as avaliações da Prova Brasil, digamos, numa escola, tenham tido resultados atipicamente elevados, tanto em Língua Portuguesa quanto em Matemática, visto que, em geral, existe uma correlação positiva e consideravelmente elevada entre o desempenho nessas duas disciplinas. Assim sendo, a escola em questão teria seu desempenho superestimado naquele ano crucial, o que faria com que, para os anos posteriores, fossem projetadas metas de desempenho artificialmente elevadas e, dessa forma, a escola, conforme se perceberia nos anos consecutivos, teria grande dificuldade em alcançar os números que lhes foram projetados.

Além disso, cabe mencionar que, além do fato de que o efeito supramencionado pode se verificar em qualquer escola, ainda que talvez com intensidades variáveis – por exemplo, pode ocorrer em escolas mais ricas ou mais pobres, e maiores ou menores –, esse efeito pode ter consequências ainda mais “nefastas” quando verificado em escolas de índice socioeconômico muito baixo. Neste caso, se alcançar as metas superestimadas (devido à acentuada flutuação positiva do desempenho em 2005) já se torna algo muito difícil para a escola, tal tarefa lhe fica ainda mais desafiadora quando essa instituição, ainda tem que melhorar, porém ao mesmo tempo atendendo a um alunado de nível socioeconômico relativamente baixo.

Porém, além disso, há outro fator que também precisa ser considerado em todo esse problema: o comportamento temporal das medidas de rendimento da escola, associados à evolução de suas respectivas metas de aprovação para um ciclo específico da escolaridade. Ou seja, interessa-se aqui pelo rendimento  $P$ , de uma forma análoga ao interesse pelo desempenho  $N$  (por sua vez, dependente de Português e Matemática), na equação:  $Ideb = N \times P$ .

Se  $P$  tiver sido superestimado em 2005 – levando-se em conta a forma como se comportou esse indicador na sua respectiva série histórica –, então poderá haver o mesmo problema aqui considerado: o de uma superestimação das metas do Ideb, com consequências negativas para que a escola consiga cumprir as metas que lhes foram traçadas.

Pois bem, observando o caso da Escola Estadual Geraldo Teixeira da Costa, constata-se, como se disse, que ela parece padecer exatamente do problema de ter tido suas metas

superestimadas, visto não ter conseguido cumprir nenhuma delas oficialmente. Só que, ao observarmos os motivos pelos quais possa ter ocorrido essa superestimação das expectativas para ela, percebemos que isso não se deveu a uma superestimação dos resultados na Prova Brasil. Na verdade, observa-se que, tanto em português, quanto em matemática, os resultados obtidos por essa escola ficaram até mesmo um pouco abaixo, embora muito próximos, do que lhes foi previsto pela série histórica total de avaliações, entre 2000 e 2013. E tampouco o índice socioeconômico dessa mesma escola é muito baixo. Muito pelo contrário, ele é um pouco superior à média estadual, situando-se acima desta 0,4 desvio-padrão de sua respectiva distribuição populacional.

Porém, ao se observar o valor do rendimento  $P$  dessa mesma escola no ano de 2005, constata-se uma acentuada superestimação desse indicador. Enquanto que o valor previsto para ele – obtido por regressão com base nos dados de  $P$  entre 2005 e 2013 – era de 0,52, por outro lado, o valor real obtido em 2005 correspondeu a 0,75, quase 50% maior, portanto, do que dele, *a posteriori*, se esperou. Este curioso comportamento se revela na evolução temporal dos valores de  $P$  para essa escola, conforme se observa no gráfico a seguir:

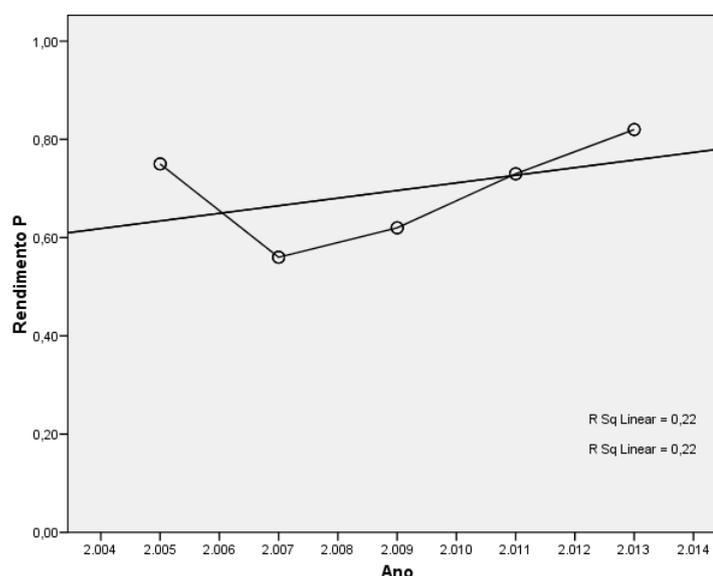


Gráfico 8.5: A evolução temporal do indicador de rendimento  $P$  da Escola Estadual Geraldo Teixeira da Costa. Fonte: elaboração própria.

Pelo gráfico acima, observa-se que, depois de 2005 – ano em que, conforme se disse, o rendimento  $P$  da escola foi de 0,75 –, os valores subsequentes foram, em geral, bastante menores do que seria de se esperar com base apenas no seu valor inicial, ficando na casa de

0,5 em 2007, 0,6 em 2009, voltando a 0,7 em 2011 e somente atingindo 0,8 em 2013. Portanto, em termos gerais, o crescimento de  $P$  ao longo de quase todo esse período foi relativamente modesto, fazendo com que, para 2005, seu valor previsto igualmente também tivesse que o ser. Porém, como não se levou em conta esse fato, mas sim o elevado valor “bruto” de  $P$  em 2005, então isso acarretou, para essa escola, valores consideravelmente altos para suas metas posteriores, os quais ela nunca conseguiu atingir.

Observando a evolução do Ideb dessa escola, dada no gráfico a seguir, percebe-se que ele sofre, no início de sua série histórica, uma vigorosa inflexão. Começa relativamente alto (em 2005), com um valor de 3,4, o qual faz com que, para 2007, sua projeção seja de 3,5.

Entretanto, ao chegar 2007, o Ideb da escola “afunda” para 2,5, ficando, portanto, muito abaixo de sua respectiva meta. A partir daí, o Ideb inicia uma relativamente lenta, porém consistente, recuperação, em boa parte devido à própria retomada das taxas de rendimento  $P$ . Dessa forma, a escola obtém 2,7 em 2009, sobe mais um ponto inteiro – indo para 3,7 – em 2011, e atinge 4,0 em 2013. Não obstante, apesar desse inegável movimento de recuperação, isso não foi suficiente para fazer a escola atingir nenhuma de suas metas, que, durante todo esse tempo, pairavam muito acima daquilo que ela, a escola, de fato conseguiu obter com seu crescimento.

Por outro lado, as metas recalculadas pelo método aqui apresentado levaram em conta essa dinâmica peculiar de evolução dos indicadores da escola, conseguindo, assim, captar, ao menos parte dessas tendências temporais de desempenho. E, assim, as metas para essa escola, calculadas por esse novo método, tornaram-se mais acessíveis neste caso, o que resultou no fato de que tal estabelecimento atingiu-as, precisamente, nas duas últimas avaliações consideradas (2011 e 2013).

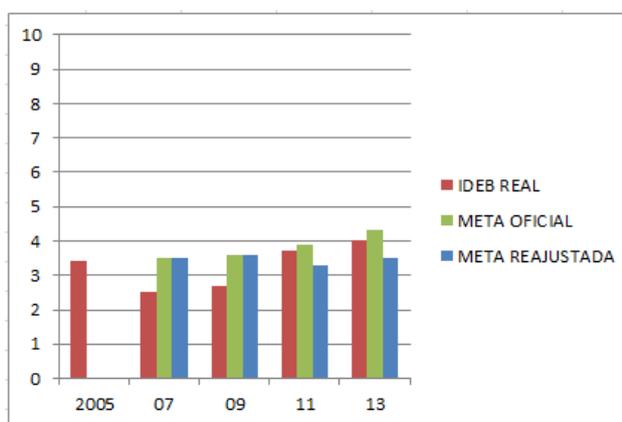


Gráfico 8.6: A evolução do Ideb e de suas duas metas (oficiais e reajustadas) da Escola Estadual Geraldo Teixeira da Costa. Fonte: elaboração própria.

*Ex. 2: Outro caso de superestimação de metas, porém numa escola pequena e devido ao desempenho excepcionalmente alto na Prova Brasil de 2005*

No caso inicial desta série de ilustrações, o estabelecimento em questão – a EE Geraldo Teixeira da Costa – é uma escola relativamente grande, que teve, ao longo da série histórica do SIMAVE – e, *grosso modo*, também da Prova Brasil – uma média de 261 alunos avaliados. Em casos assim, conforme já se demonstrou no presente trabalho, as médias das escolas em cada edição de um teste tendem a oscilar um pouco menos em torno de seus respectivos valores previstos pela tendência de crescimento, muito embora elas ainda continuem a apresentar uma oscilação nada desprezível, conforme já se demonstrou.

Entretanto, nos casos em que a escola é, de fato, muito pequena – contando, por exemplo, com somente poucas dezenas de alunos avaliados por série –, então essas flutuações podem, em geral, vir a ser muito maiores, fazendo com que os problemas acima especificados, acerca da determinação das metas do Ideb, fiquem também ainda mais acentuados.

Este é, precisamente, o caso, por exemplo, da Escola Estadual Prof. Biolkino de Andrade, no município de Divinésia, na SRE de Ubá. Trata-se de uma escola consideravelmente pequena, que, no 9º ano EF, apresentou, na série histórica do SIMAVE, uma média de somente 39 alunos avaliados por edição do teste. Quanto ao índice socioeconômico, observa-se que esse estabelecimento possui um ISE médio de -0,1, ou seja, quase igual à média populacional do estado, mostrando que, neste caso, essa variável específica não parece ter influenciado particularmente a disparidade de cumprimento de metas que nela se observa.

Porém, o comportamento anômalo se manifesta pelo fato de que, analogamente ao caso anterior, esta escola apresentou zero por cento de cumprimento de metas do Ideb segundo os dados oficiais, ao mesmo tempo em que também obteve 50% de cumprimento das metas retificadas pelo método ora proposto. Observando os dados históricos e previstos de desempenho dessa escola, percebe-se que a principal causa do problema aqui observado foi a superestimação de seus resultados de Matemática na Prova Brasil de 2005, que ficaram nada menos que 20 pontos acima do que lhe foi estimado tomando-se como base toda a série histórica Prova Brasil/SIMAVE, que temos disponível entre 2000 e 2013. Já a média real de Português nesse mesmo ano de 2005 na Prova Brasil, embora também tenha ficado superestimada, situou-se somente três pontos acima do seu respectivo valor previsto pela tendência de crescimento.

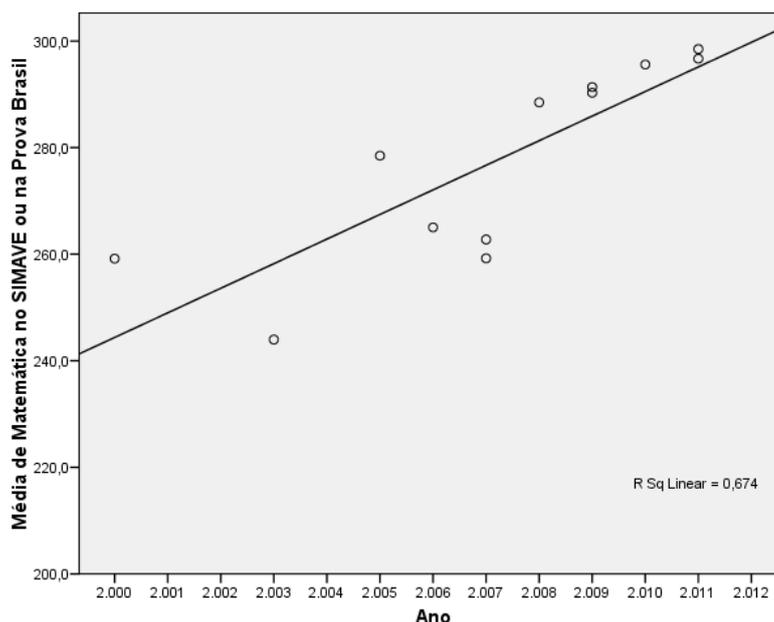


Gráfico 8.6: A evolução temporal da média de Matemática da Escola Estadual Prof. Biolkino de Andrade.

Além disso, para essa mesma escola, também houve uma superestimação do indicador de rendimento  $P$ , que, nos dados reais de 2005, ficou 0,13 ponto acima do valor previsto pela sua respectiva tendência de crescimento. Tudo isso provocou, ao mesmo tempo, a obtenção de valores artificialmente elevados, tanto para o desempenho  $N$  quanto para o rendimento  $P$  no cálculo das metas posteriores do Ideb, fazendo com que estas últimas se situassem em patamares que, na prática, tornaram-se inatingíveis pela escola nos anos subsequentes, entre 2007 e 2013.

Por outro lado, os resultados dessa mesma escola não foram tão ruins assim levando-se em conta a metodologia aqui apresentada. Por esses novos cálculos, percebe-se que, se o Ideb crescente dessa escola ao longo do período considerado não é suficiente para fazê-la cumprir as duas primeiras metas – de 2007 e de 2009 –, por outro lado, tanto em 2011 quanto em 2013, observa-se uma considerável recuperação dessa escola, pela qual se atingiram as metas revistas para ambas essas últimas edições do teste.

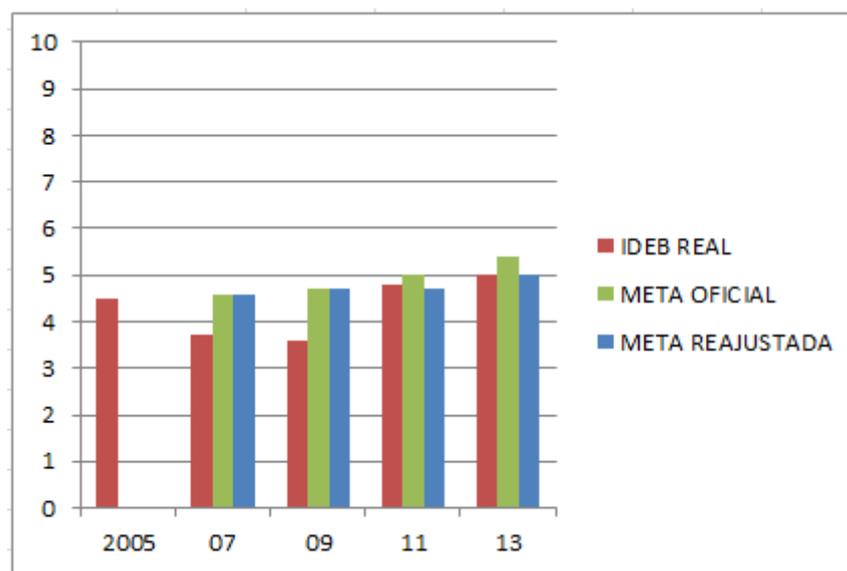


Gráfico 8.6: A evolução do Ideb e de suas duas metas (oficiais e reajustadas) da Escola Estadual Prof. Biolkino de Andrade. Fonte: elaboração própria.

*Ex. 3: Um caso de subestimação de metas numa escola pequena e devido ao desempenho excepcionalmente baixo nas duas disciplinas avaliadas em 2005*

Até aqui se comentaram dois casos de escolas que tiveram suas metas superestimadas pelos métodos oficiais, fazendo com que seus resultados ficassem muito aquém de seu real desempenho, conforme mensurado com base nas metas mais realistas por nós propostas. Entretanto, naturalmente, pode também ocorrer o contrário. Ou seja, pode haver casos de escolas que, em 2005, tiveram, ou subestimadas suas médias de Português e/ou Matemática, ou também tiveram subestimado seu valor para o rendimento  $P$ , ou ainda que possuem um índice socioeconômico de seu alunado relativamente elevado – algo que não foi considerado quando se traçaram as metas do Ideb – ou mesmo um conjunto de fatores reunindo tudo isso.

Em situações desse tipo, o cenário é previsível: a escola atinge todas as suas metas, ou quase, em boa parte devido ao fato de que estas últimas são consideravelmente baixas, quando comparadas com aquilo que se espera dessa mesma escola, ao se analisar, não apenas o seu comportamento em 2005, porém um período dilatado de tempo, correspondente à sua série histórica para os indicadores do Ideb.

Um caso típico dessa situação é a Escola Municipal União Comunitária, de Belo Horizonte, que se destaca, aliás, pelo seu pequeno tamanho: uma média de somente 32 alunos por edição do SIMAVE, algo que, conforme se comentou, tem o potencial de causar grandes diferenças entre as medidas de desempenho reais e previstas e, conseqüentemente, também

pode ocasionar grandes distorções acerca do cumprimento ou não das metas do Ideb, tanto do ponto de vista desfavorável quanto favorável, à escola.

Neste caso específico, por exemplo, observa-se que as médias obtidas na Prova Brasil resultaram numa considerável subestimação do desempenho real da escola, conforme observado por sua série histórica. Isso fez com que, para ela, fossem projetadas médias consideravelmente modestas do Ideb, que foram todas cumpridas entre 2007 e 2013.

Observando-se os indicadores dessa escola para 2005, nota-se que a principal razão disso reside no fato de que, para aquele ano, a média de Matemática da escola ficou bastante abaixo do previsto pela regressão com base na série histórica.

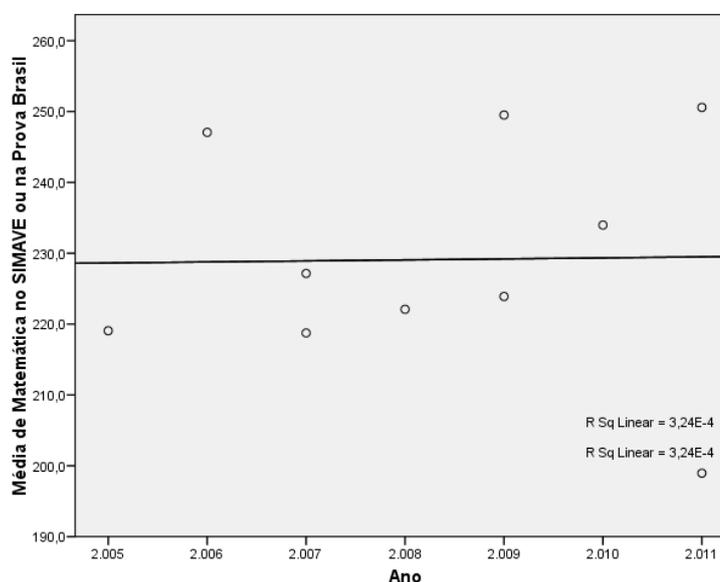


Gráfico 8.7: A evolução temporal da média de Matemática da Escola Municipal União Comunitária. Fonte: elaboração própria.

Como consequência disso, essa escola teve metas do Ideb consideravelmente modestas para as edições subsequentes da Prova Brasil. Entretanto, o seu desempenho real foi tão pouco destacado que, mesmo com uma meta extremamente modesta para 2007 (3,4), ela, naquele mesmo ano, ficou com 3,2, o que a levou, naturalmente, a não cumprir sua meta naquele ano. Entretanto, para as edições seguintes, as metas, continuando modestas, foram de 3,5 em 2009, 3,8 em 2011 e 4,2 em 2013. Tanto em 2009 quanto em 2011, a escola obteve valores reais do Ideb exatamente iguais a essas metas oficiais projetadas, ou seja, satisfez às expectativas oficiais, embora tenha ficado aquém de todas as metas recalculadas.

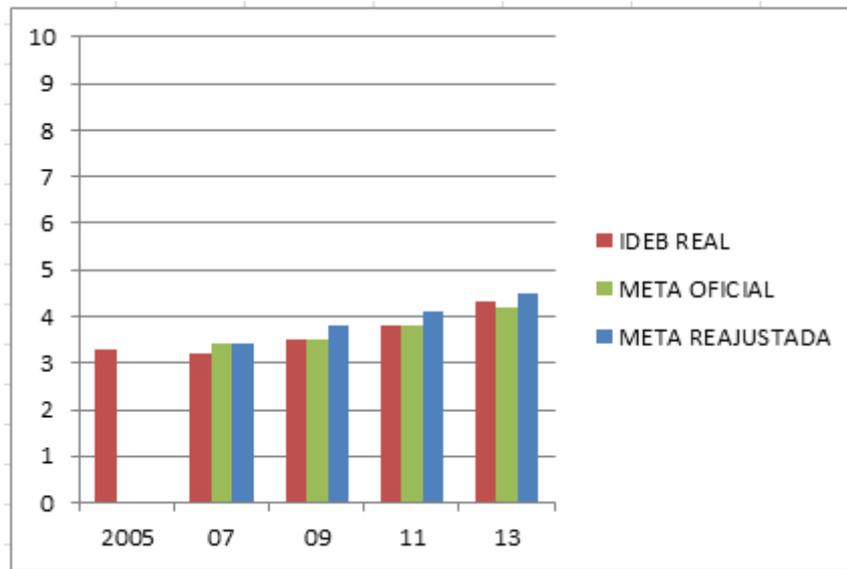


Gráfico 8.7: A evolução do Ideb e de suas duas metas (oficiais e reajustadas) da Escola Municipal União Comunitária. Fonte: elaboração própria.

## CONCLUSÕES

A título de encerramento, apresentamos, nos parágrafos seguintes, uma sequência de conclusões referentes ao presente texto.

Ao longo deste trabalho, foram considerados diversos tópicos inter-relacionados, que começaram com uma questão mais ampla acerca dos sistemas de avaliação de desempenho, os quais se vêm consistindo num mecanismo de uso amplamente difundido pelos governos em todo o mundo contemporâneo, inclusive no Brasil.

Passou-se então ao caso específico da avaliação do desempenho educacional, levando-se inicialmente em conta a questão de se definir e delimitar, de uma forma operacional, aquilo que se pode considerar como tema de estudo da qualidade das escolas.

Tanto a literatura especializada quanto o senso comum concordam que, à escola, compete uma série de desenvolvimentos que se devem promover nos alunos que por ela passam, e que incluem fatores como o exercício da cidadania, a formação de indivíduos social e economicamente produtivos, o desenvolvimento da consciência crítica e da capacidade de expressão, o domínio de habilidades acadêmicas, etc.

No entanto, diante da necessidade concreta de se avaliarem os sistemas educacionais de uma forma exequível, rápida, fidedigna e padronizada, uma opção que vem sendo amplamente adotada em diversos lugares é a adoção de testes de proficiência de disciplinas específicas, elaborados com base em matrizes de proficiência bem delimitadas e com seus resultados expressos em escalas igualmente bem definidas de desempenho.

Um avanço significativo na proposição e na implementação desse tipo de procedimento avaliativo tem ocorrido nos Estados Unidos, país que é responsável pela implementação de sistemas nacionais de avaliação de várias décadas de existência, além de também ter sido palco de grandes movimentos de reforma educacional ao longo de sua história.

No início do século 21, intensificou-se nos Estados Unidos o movimento de testagem em larga escala de suas escolas do ensino básico. E a isso, também se acrescentou a vinculação, a tais resultados, de conseqüências mais altas às escolas e aos profissionais educacionais avaliados.

Um grande responsável por essa intensificação do movimento de prestação de contas norte-americano foi o *No Child Left Behind Act*, que passou a exigir que todas as escolas e estados pleiteando subvenções federais para a educação passassem a prestar contas do desempenho de todos os seus alunos, levando-se também em conta as possíveis diferenças de desempenho acadêmico associadas ao pertencimento dos alunos a diferentes grupos sociais, econômicos, étnicos, bem como a grupos delimitados por outros fatores, como o linguístico e também a capacitação física e motora dos alunos.

No Brasil, tem-se verificado uma marcada influência das iniciativas norte-americanas, que se tem traduzido num movimento que já dura algumas décadas, envolvendo o desenvolvimento dos sistemas de avaliação tanto em nível nacional como estadual ou municipal.

Em parte, esse movimento de testagem no Brasil se associa ao fato de que o país vem experimentando, também em décadas mais recentes, uma certa mudança de foco em termos de temas de grandes debates nacionais sobre a educação. Até por volta da entrada do último quartel do século 20, o país ainda convivia com elevadas taxas de analfabetismo em sua população, além de altos índices de repetência e abandono, inclusive no ensino fundamental. Portanto, o foco de maior destaque naquele período era a expansão e capacidade de inclusão dos sistemas de ensino existentes.

De certo modo, este último objetivo se alcançou razoavelmente bem por volta do final do século 20. E, uma vez conseguido isso, acentuou-se a discussão pela qualidade do ensino, que, por sua vez, contribuiu para o fortalecimento do debate acerca da avaliação educacional como um mecanismo necessário para se alcançar tais níveis de qualidade, tema que persiste na agenda educacional dos estados e do país até os dias de hoje.

De modo geral, é possível, observar, no Brasil, três momentos distintos no que diz respeito aos seus sistemas de avaliação educacional. Primeiramente, houve a ocorrência de um certo “balizamento” das políticas e técnicas de avaliação em grande escala, levada a cabo pelo estabelecimento de sistemas avaliativos nacionais, por parte do Ministério da Educação. E, como uma consequência disso, surgiu o SAEB, adotou-se a Teoria da Resposta ao Item na elaboração e interpretação dos instrumentos e resultados da testagem, estabeleceram-se as escalas de proficiência em diferentes disciplinas, com uma ênfase em Matemática e Língua Portuguesa, etc.

O segundo período assistiu ao desenvolvimento dos sistemas de avaliação estaduais, sendo que São Paulo e Minas Gerais foram alguns casos de estados precursores neste sentido.

Um ponto que cabe ser observado em relação a esses sistemas de mensuração do desempenho educacional é a conveniência, ou mesmo a necessidade – na prática, poucas vezes observada – de se avaliar a qualidade da educação não somente com base nos resultados de uma única edição do teste, mas sim a partir da evolução de sua série histórica.

Uma possibilidade para se fazer isto, aqui aplicada ao SIMAVE, é elaborar índices de status e progresso da proficiência escolar, com base em uma sequência de resultados obtidos ao longo de toda a série histórica da avaliação em questão. Embora, ainda assim, haja muitos casos de escolas com desempenho ambíguo – ora maior, ora menor – também é possível detectar casos consistentes de desempenho, tanto para mais quanto para menos, o que pode ser especialmente útil, em particular para os gestores e pesquisadores educacionais.

Por sua vez, o terceiro período de desenvolvimento da experiência brasileira com a avaliação educacional em larga escala, correspondeu à adoção de um sistema nacionalmente unificado de acompanhamento do desempenho, que se deu através do Ideb e de suas respectivas metas. Além disto, igualmente tem se intensificado, nesse mesmo período, o desenvolvimento dos sistemas de prestação de contas, e também de criação de índices de qualidade educacional próprios dos estados da federação.

Nesse sentido, o Índice de Desenvolvimento da Educação Brasileira, o Ideb, tem assumido um papel de grande relevo no Brasil, e seu emprego vem se generalizando para diversos fins, como a formulação de políticas públicas educacionais, o planejamento escolar, a informação e prestação de contas, etc.

Entretanto, há certos aspectos do Ideb que merecem ser considerados com certa cautela, e isso se torna ainda mais premente no contexto atual, devido à generalizada e crescente utilização desse índice.

Um ponto que merece um cuidado especial é o fato de que as metas oficiais do Ideb foram calculadas apenas com base em dados escolares centrados no ano de 2005.

Um aspecto decorrente desse fato é que as metas do Ideb encontram-se “congeladas” no tempo, ou seja, elas não mais se alteraram, desde quando foram propostas em 2005.

Entretanto, as medidas utilizadas para se calcular as metas não existem somente para 2005. Por exemplo, há dados de aprovação nas escolas públicas mineiras desde bem antes de 2005. E, além disso, no caso de Minas Gerais, por exemplo, também tem havido, desde o ano 2000, embora com certas interrupções iniciais, testes censitários de Língua Portuguesa e Matemática.

Portanto, quando se consideram as medidas geradoras do Ideb segundo uma perspectiva mais ampla, envolvendo períodos anteriores e posteriores a 2005, percebe-se que os resultados daquele ano específico, em geral, representam certas flutuações verificadas em torno das trajetórias de evolução traçadas temporalmente para essas mesmas medidas.

Um problema significativo de não se levar em conta essa flutuação é o risco que se corre de que, no ano de 2005, os resultados da escola tenham fugido consideravelmente àquilo que deles se esperava, seja para mais ou para menos. Portanto, na fórmula do cálculo das metas do Ideb, há um grau considerável de instabilidade decorrente dessas flutuações.

No caso tanto de Língua Portuguesa quanto de Matemática no 9º ano do Ensino Fundamental, as escolas tiveram suas respectivas médias fluando em torno de seus valores esperados por distâncias médias que chegam a alcançar dois dígitos nas respectivas escalas de proficiência, sendo que, em ambas, para esta série específica, a média nacional valeu 250 pontos e o desvio-padrão, 50, segundo os resultados básicos do SAEB de 1995.

Outro ponto de relevo é que, ainda que essas flutuações diminuam com o tamanho da escola (estimado pelo número médio de alunos matriculados na série em questão), ainda assim, elas são consideráveis mesmo nas escolas maiores, ficando, em média, na casa dos dois dígitos das respectivas escalas de proficiência.

Além da variação da proficiência, existe também a evolução temporal do fator de rendimento  $P$ , associado às taxas de aprovação, evasão e abandono das escolas, as quais também se comportam com oscilações em torno de uma tendência de crescimento ao longo do período considerado neste estudo.

Uma das ideias propostas por este trabalho é chamar a atenção para a possibilidade de se levar em conta toda essa instabilidade na determinação das metas do Ideb.

Além disso, aqui também se propõe uma metodologia de determinação das metas do Ideb que leve em conta o índice socioeconômico médio dos alunos das escolas, algo que tampouco foi observado quando elas foram fixadas em 2005.

Acompanhando as metas do Ideb, é possível perceber que há vários casos de escolas que não vêm conseguindo cumprir suas metas, apesar de virem experimentando um razoável crescimento do seu respectivo Ideb. E também é possível perceber que muitas dessas escolas tiveram, em 2005 – ano de fixação das metas –, um desempenho consideravelmente elevado, quando se consideram os valores previstos pela regressão acima mencionada. Isso fez com que suas metas tenham ficado relativamente elevadas, o que, ao menos em parte, explica o seu fraco desempenho no cumprimento das mesmas.

Também se observam casos inversos, de escolas que vêm obtendo um certo sucesso quanto ao cumprimento de metas, muito embora isso se deva, também ao menos em parte, ao fato de que elas tiveram um desempenho excepcionalmente ruim em 2005, quando comparado com a média dos valores de suas respectivas séries históricas. Dessa forma, essas escolas foram “beneficiadas” por metas mais baixas, o que ajuda a explicar o seu relativo sucesso, mesmo sem virem apresentando um desempenho muito elevado.

Portanto, propõe-se, aqui, um método que leva em conta essas flutuações de desempenho, a fim de se reajustarem as metas depois de uma ou mais edições de testes de proficiência envolvendo a população considerada.

Outra vantagem desse método é que ele promove uma atualização das metas com base nas tendências mais recentes de desempenho apresentadas pelas escolas. Dessa forma, torna-se mais dinâmica a relação entre o desempenho das escolas e o estabelecimento de suas respectivas metas para o Ideb.

Outra vantagem da utilização desse método é o fato dele possibilitar um aumento da fidedignidade das medidas, por meio do uso de um número muito maior de observações, oriundas, no caso de Minas Gerais, dos testes anuais do seu Sistema Mineiro de Avaliação da Educação Básica, o SIMAVE, além dos resultados acumulados das edições bianuais da Prova Brasil.

Para obter as novas metas, a metodologia aqui empregada envolve a realização de uma análise linear hierárquica longitudinal de dois níveis, tendo a escola e a evolução temporal de suas respectivas médias como níveis de análise.

Para as estimativas dos parâmetros de desempenho e rendimento, adotou-se um modelo linear de crescimento, onde tanto a proficiência inicial quanto a sua respectiva taxa de variação para cada escola é controlado por seu respectivo índice socioeconômico médio. Além disso, os cálculos também levaram em conta a ponderação das escolas pelo seu número médio de alunos matriculados na série em questão.

Com base, então, nesses valores previstos, e controlados pelo índice socioeconômico das escolas, traçaram-se novas metas do Ideb, utilizando-se, basicamente, as mesmas fórmulas de cálculo adotadas oficialmente pelo MEC.

Numa comparação entre o desempenho das escolas segundo o cumprimento das duas diferentes metas – a oficial e a por nós proposta –, constatou-se um acentuado grau de concordância dos resultados, a qual, para o primeiro ano das metas (2007) correspondeu a cerca de dois terços dos casos.

Entretanto, em meio ao universo das escolas públicas mineiras, um número considerável de estabelecimentos apresentou um resultado discrepante em termos do cumprimento das duas metas. E essas diferenças vêm aumentando com o tempo, até 2013, ano que correspondeu ao resultado mais recente da série histórica aqui utilizada.

Um problema de gravidade considerável que se percebe no caso do Ideb, com base no que se tem apontado neste trabalho, é o fato de que, não poucas escolas vêm recebendo uma avaliação baseada nas metas desse índice que simplesmente não se coaduna com a trajetória de desenvolvimento de suas medidas de desempenho e rendimento num prazo mais dilatado.

E assim, surge a possibilidade de ocorrência de vários efeitos perversos, associados tanto à avaliação propriamente dita, como também ao uso que se faz dela. Por exemplo, uma escola pode estar sendo considerada como tendo um desempenho satisfatório, enquanto que talvez isso não seja o caso. E tal julgamento pode implicar tanto numa possível compensação “injusta” para a escola, como também pode contribuir para que, equivocadamente, gestores, pesquisadores ou outros mais passem a vê-la como um modelo de sucesso, enquanto que tal fato não propriamente corresponde à realidade. E, ainda segundo essa visão, procurarão estudar elementos da escola – como aspectos de sua gestão, de suas práticas de ensino, de sua organização, etc. – que podem vir a dar sinais imprecipientes de uma possível associação com o suposto sucesso escolar.

Naturalmente, o outro extremo também pode ocorrer, ou seja, o de escolas cujo desempenho em relação ao cumprimento das metas vem sendo, digamos, objeto de grande preocupação de gestores, professores e outros, enquanto que há casos em que elas não parecem estar tão mal assim.

Portanto, esse feixe de conclusões se encerra com a recomendação de que os fatores apontados no presente estudo, bem como a metodologia por ele apresentada, possam ser objeto de consideração das instâncias educacionais responsáveis pela determinação e também pela análise dos resultados do Ideb.

E, dada a possibilidade de adaptação do modelo proposto a outros contextos estaduais brasileiros, fica igualmente a sugestão de que ele possa ser utilizado em outros contextos estaduais, que não apenas o mineiro, por meio de um “casamento” dos dados da Prova Brasil com os provenientes dos sistemas de avaliação estadual, nos moldes do SAEB, que se vêm implementando no país ao longo das últimas décadas.

## REFERÊNCIAS BIBLIOGRÁFICAS

ALDENDERFER, M.; BLASHFIELD, R. *Cluster analysis*. Beverly Hills: SAGE Publications, 1987.

ALDRICH, J.; NELSON, F. *Linear probability, logit, and probit models*. Beverly Hills: SAGE Publications, 1993.

ANDRADE, E. "School accountability" no Brasil: experiências e dificuldades.. *Revista de Economia Política*, v. 28, n. 3 (111), p. 443–453, 2008.

ARELLANO, D. et al. *Sistemas de evaluación del desempeño para organizaciones públicas: ¿Cómo construirlos efectivamente?* México: CIDE, 2012.

BINKLEY, M. et al. *Defining 21<sup>st</sup> century skills*. Artigo apresentado no Learning and Technology World Forum, Londres, 2010.

BOHAN, C.H., & NULL, J.W. Gender and the evolution of normal school education: A historical analysis of institutions. *Educational Foundations: The Journal of the Social Foundations of Education*, 21(3&4), 3–26, 2007.

BOLFARINE, H., BUSSAB, W. *Elementos de amostragem*. São Paulo: Editora Blucher, 2005.

BRANNICK, M. *Item response theory*. Disponível no URL: <<http://luna.cas.usf.edu/~mbrannic/>>. (Acessado em 19 de abril de 2012).

BRINKLEY, A. *American History: A Survey*, Vol. I. New York: McGraw Hill, 2009.

BROOKE, N. (Org.). *Marcos históricos na reforma da educação*. Belo Horizonte: Fino Traço, 2012.

\_\_\_\_\_. (Org.). *Pesquisa em eficácia escolar: origem e trajetórias*. Belo Horizonte: Editora UFMG, 2008.

BRYK, A. S.; RAUDENBUSH, S. W. *Hierarchical linear models: applications and data analysis methods*. Newbury Park: SAGE, 2002.

CAED. *Boletim de resultados: SAERJ 2008*. Juiz de Fora: CAEd, 2009.

\_\_\_\_\_. *Boletim de resultados: Ceará, avaliação da alfabetização; SPAECE-Alfa 2007*. Juiz de Fora: CAEd, 2008.

CARLSON, D. *Focusing state educational accountability systems: four methods of judging school quality and progress*. Disponível no URL: <<http://nciea.org/publications/Dale020402.pdf>> (Acessado em 19 de novembro de 2014).

CHUNG, C.; REIMERS, F. *GEI<sup>2</sup> Singapore workshop: 21CC & US*. Apresentação inédita no Encontro Internacional sobre Competências do Século 21 de Singapura, 2014.

COMER, J.P. (2004). *Leave no child behind: preparing today's youth for tomorrow's world*. Connecticut: Yale University Press, 2004.

CROMWELL, J. et al. *Univariate tests for time series models*. Thousand Oaks: SAGE Publications, 1994.

DEMARIS, A. *Logit modeling: practical applications*. Beverly Hills: SAGE Publications, 1992.

DORAN, H. *From saint to sinner and back again: the confounding effect of linking error on gains estimated from value-added models*. American Institutes for Research. Disponível no URL: [http://marces.org/conference/value\\_added/Doran%27s%20Value-Added%20Slides.pdf](http://marces.org/conference/value_added/Doran%27s%20Value-Added%20Slides.pdf) (Acesso em 23 de novembro de 2014).

\_\_\_\_\_. *Value-added models and adequate yearly progress: combining growth and adequacy in a standards-based environment*. Artigo apresentado na *Annual CCSSO Large-Scale Assessment Conference*, Boston, 2004.

FRANCO, A. M. *Uma análise de rankings de escolas brasileiras com dados do SAEB*. American Institutes for Research. Disponível no URL: <http://core.kmi.open.ac.uk/download/pdf/6357394.pdf> (Acessado em 11 de maio de 2014).

GARDNER, D. et al. *A nation at risk: the imperative for educational reform*. Disponível no URL: [http://datacenter.spps.org/uploads/sotw\\_a\\_nation\\_at\\_risk\\_1983.pdf](http://datacenter.spps.org/uploads/sotw_a_nation_at_risk_1983.pdf) >. (Acesso em 30 de janeiro de 2015)

GLENN, N. *Cohort analysis*. Thousand Oaks: SAGE Publications, 2005.

HAGENAARS, J. *Loglinear models with latent variables*. Beverly Hills: SAGE Publications, 1993.

HAMBLETON, R. et al. *Fundamentals of item response theory*. Newbury Park: SAGE Publications, 1991.

HANUSHEK, E.; RAYMOND, M. *Improving educational quality: how best to evaluate our schools?* Disponível no URL: <http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%2BRaymond%202003%20Educ21stCent.pdf>. (Acessado em 17 de maio de 2014).

HILL, R.; DePASCALE, C. *Reliability of No Child Left Behind accountability designs*. The National Center for the Improvement of Educational Assessment, 2003.

HOGAN, T. *Introdução à prática de testes psicológicos*. Rio de Janeiro: LTC Editora, 2006.

INEP. *Características do SAEB*. Disponível no URL: <http://portal.inep.gov.br/caracteristicas-saeb> >. (Acessado em 20 de abril de 2012a).

\_\_\_\_\_. *Nota técnica: índice de desenvolvimento da educação básica – Ideb*. Disponível no URL: [http://www.inep.gov.br/download/Ideb/Nota\\_Tecnica\\_n1\\_concepcaoIdeb.pdf](http://www.inep.gov.br/download/Ideb/Nota_Tecnica_n1_concepcaoIdeb.pdf)>. (Acessado em 18 de setembro de 2010).

\_\_\_\_\_. *Nota técnica: metodologia utilizada para o estabelecimento das metas intermediárias*

para a trajetória do Ideb no Brasil, Estados, Municípios e Escolas. Disponível no URL: [http://portalideb.inep.gov.br/index.php?option=com\\_content&view=article&id=40&Itemid=58](http://portalideb.inep.gov.br/index.php?option=com_content&view=article&id=40&Itemid=58)>. (Acessado em 18 de março de 2012b).

\_\_\_\_\_. *Nota técnica: nota metodológica sobre a compatibilização de desempenhos do PISA com a escala do SAEB*. Disponível no URL: <[http://portalideb.inep.gov.br/index.php?option=com\\_content&view=article&id=40&Itemid=58](http://portalideb.inep.gov.br/index.php?option=com_content&view=article&id=40&Itemid=58)>. (Acessado em 18 de março de 2012c).

KANE, T.; STAIGER, D. The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, v. 16, n. 4, p. 91–114, 2002.

\_\_\_\_\_. *Volatility in school test scores: implications for test-based accountability systems*. 2002. Disponível no URL: <<http://www.jstor.org/discover/10.2307/20067246?sid=21105466072533&uid=3737664&uid=4&uid=2>>. (Acessado em 18 de abril de 2013).

KORETZ, D. *Measuring up: what educational testing really tells us*. Cambridge: Harvard University Press, 2008.

LEVINE, D. et al. *Estatística: teoria e aplicações: usando Microsoft Excel em português*. Rio de Janeiro: LTC Editora, 2000.

LINN, R. *Stability of school building accountability scores and gains*. Los Angeles: California University, 2002.

MELLO, M. Sistemas de avaliação educacional: possibilidades e desafios. In BAUER, A; GATTI, B. (Org.) *Vinte e cinco anos de avaliações de sistemas educacionais no Brasil*. Florianópolis: Insular Livros, 2013.

MENARD, S. *Longitudinal research*. Beverly Hills: SAGE Publications, 2002.

MINISTÉRIO DA EDUCAÇÃO. *SAEB – 2005, Primeiros resultados: médias de desempenho do SAEB 2005 em perspectiva comparada*. Disponível no URL: <[http://www.inep.gov.br/download/saeb/2005/SAEB1995\\_2005.pdf](http://www.inep.gov.br/download/saeb/2005/SAEB1995_2005.pdf)>. (Acessado em 18 de março de 2008).

MOONEY, C; DUVAL, R. *Bootstrapping: a non-parametric approach to statistical inference*. Newbury Park: SAGE Publications, 1984.

MOORE, D. et al. *A prática da estatística empresarial: como usar dados para tomar decisões*. Rio de Janeiro: LTC Editora, 2006.

O'CONNELL, A. A.; MCCOACH, D. B. (ed.) *Multilevel modeling of educational data*. Charlotte: Information Age Publishing, 2008.

ORGANIZAÇÃO PARA A COOPERAÇÃO E O DESENVOLVIMENTO ECONÔMICOS – OCDE. *OECD Programme for International Student Assessment (PISA)*. Disponível no URL: <http://www.oecd.org/pisa>. (Acessado em 18 de agosto de 2009).

PELLEGRINO, J; HILTON, M. *Education for life and work: developing transferable knowledge and skills in the 21<sup>st</sup> century*. Washington, D. C.: The National Academies Press, 2012.

RAUDENBUSH, S. What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics*, v. 29, n. 1, p. 121–129, 2004.

RAUDENBUSH, S.; WILLMS, J. The estimation of school effects. *Journal of Educational and Behavioral Statistics*, v. 20, n. 4, p. 307–335, 1995.

RAVITCH, D. *Vida e morte do grande sistema escolar Americano: como os testes padronizados e o modelo de Mercado ameaçam a educação*. Porto Alegre: Sulina, 2011.

REIMERS, F. *Informed dialogue: using research to shape education policy around the world*. Cambridge: Harvard University Press, 2008.

ROGOSA, D. *Analyses of AB1114 schools*. Stanford: Stanford University, 2002a. Disponível no URL < <http://statweb.stanford.edu/~rag/api/ab1114report.pdf> > Acesso em 12 de março de 2014a.

\_\_\_\_\_. *Plan and preview for API accuracy reports*. Stanford: Stanford University, 2002b. Disponível no URL < <http://statweb.stanford.edu/~rag/api/apiacc.pdf> > Acesso em 13 de maio de 2014b.

\_\_\_\_\_. *Confusions about consistency in improvement*. Disponível no URL: <http://www-stat.stanford.edu/~rag/api/consist.pdf>. (Acessado em 18 de junho de 2012).

\_\_\_\_\_. *Irrelevance of reliability coefficients to accountability systems: statistical disconnect in Kane-Staiger “volatility in school tests”*. Disponível no URL: <http://statweb.stanford.edu/~rag/api/ksresst.pdf>. (Acessado em 18 de junho de 2014).

SÃO PAULO. *Relatório pedagógico: 2010 SARESP – Matemática*. São Paulo: SARESP, 2010.

SCHWARTZMAN, S. *Education-oriented social programs in Brazil: the impact of Bolsa Escola*. Disponível no URL: [http://www.schwartzman.org.br/simon/pdf/bolsa\\_escola\\_eng.pdf](http://www.schwartzman.org.br/simon/pdf/bolsa_escola_eng.pdf). (Acessado em 11 de agosto de 2009).

SHAW, I. et al. (Org.) *The SAGE handbook of evaluation*. Thousand Oaks: SAGE Publications, 2006.

SINGER, J. D.; WILLETT, J. B. *Applied longitudinal data analysis: modeling change and event occurrence*. New York: Oxford University Press, 2003.

SOARES, F.; XAVIER, F. Pressupostos educacionais e estatísticos do Ideb. *Educação e Sociedade*, v. 34, n. 124, p. 903–923, 2013.

TEDDLIE, C., REYNOLDS, D. (Org.) *The international handbook of school effectiveness research*. New York: Falmer Press, 2000.

US DEPARTMENT OF LABOR. *What work requires of schools: a SCANS report for America 2000*. Disponível no URL: < <http://wdr.doleta.gov/SCANS/whatwork/whatwork.pdf> >. Acesso em 2 de fevereiro de 2015.

US SECRETARY OF EDUCATION. *No child left behind: a desktop reference 2002*. Disponível no URL < <http://www2.ed.gov/admins/lead/account/nclbreference/reference.pdf> >. Acesso em 2 de fevereiro de 2015.

WAY, W. *Precision and volatility in school accountability systems*. Educational Testing Service, 2006.

WEISS, C. *Evaluation: methods for studying programs and policies*. Upper Saddle River: Prentice Hall, 1997.