



UNIVERSIDADE DE BRASÍLIA

Instituto de Psicologia

**Avaliação Educacional Sistêmica na
Perspectiva dos Testes de Desempenho e
de seus Resultados:
Estudo do SAEB**

Margarida Maria Mariano Rodrigues

**BRASÍLIA
2007**

Universidade de Brasília – UnB
Instituto de Psicologia – IP
Departamento de Psicologia Social e do Trabalho – PST
Laboratório de Pesquisa em Avaliação e Medidas - LabPAM

**Avaliação Educacional Sistêmica na Perspectiva
dos Testes de Desempenho e de seus Resultados:
Estudo do SAEB**

MARGARIDA MARIA MARIANO RODRIGUES

ORIENTADOR:
PROF. DR. LUIZ PASQUALI

BRASÍLIA – DF
ABRIL/2007

**Avaliação Educacional Sistêmica na Perspectiva
dos Testes de Desempenho e de seus Resultados:
Estudo do SAEB**

TESE APRESENTADA AO CURSO DE DOUTORADO DO PROGRAMA DE PÓS-GRADUAÇÃO
DO INSTITUTO DE PSICOLOGIA DA UNIVERSIDADE DE BRASÍLIA COMO REQUISITO
À OBTENÇÃO DO GRAU DE DOUTORA EM PSICOLOGIA

ORIENTADOR:
PROF. DR. LUIZ PASQUALI

BRASÍLIA, 23 DE ABRIL DE 2007

**INSTITUTO DE PSICOLOGIA
UNIVERSIDADE DE BRASÍLIA**

ESTA TESE FOI AVALIADA PELA SEGUINTE COMISSÃO EXAMINADORA:

**PROF. DOCTEUR LUIZ PASQUALI – PRESIDENTE
PROFESSOR DO INSTITUTO DE PSICOLOGIA DA UNB**

**PROF^a. DR^a VERA LÚCIA MARQUES FIGUEIREDO
UNIVERSIDADE CATÓLICA DE PELOTAS – RS**

**PROF. DR. RUBEN KLEIN
LABORATÓRIO NACIONAL DE COMPUTAÇÃO CIENTÍFICA – MCT**

**PROF. DR. AMAURY PATRICK GREMAUD
DIRETOR DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA BRASILEIRA DO INEP**

**PROF. DR. JACOB ARIE LAROS
PROFESSOR DO INSTITUTO DE PSICOLOGIA DA UNB**

**PROF. DR. BARTHOLOMEU TORRES TRÓCCOLI
PROFESSOR DO INSTITUTO DE PSICOLOGIA DA UNB**

BRASÍLIA, 23 DE ABRIL DE 2007

À memória de meu pai que, por seus gestos e exemplo, me ensinou que é possível transformar o mundo por meio de pequenas ações.

Agradecimentos

O término deste trabalho representa o alcance da mais importante etapa de minha vida acadêmica e profissional. Reconheço que não poderia tê-lo levado a termo sem o precioso e constante apoio que recebi de meu marido, Benedito, e de meus filhos, Luís Fernando e Raquel, que abriram mão de inúmeros momentos de nossa convivência para que eu pudesse desenvolvê-lo.

Muitas pessoas estiveram comigo no processo de amadurecimento de minhas idéias. Minha trajetória de vida profissional e acadêmica me sensibilizou profundamente para questões relacionadas à educação. No momento de encerramento dessa importante fase, quando paramos para pensar nas pessoas que mais fortemente contribuíram para o desenvolvimento desta tese, torna-se impossível enumerá-las. Vêm à memória os alunos que tive em periferias miseráveis, que certamente poucas pessoas do meio acadêmico conhecem, os quais me puseram a par de uma realidade ímpar, de sofrimento, de falta de apoio, de desprezo. Também vem à memória o nosso rico meio acadêmico que nos dá a oportunidade de compreender os fenômenos que permeiam esses ambientes contraditórios e de propor ações que ajudem a superar essas diversidades.

De modo especial, agradeço a paciência, o incentivo, a sabedoria do Professor Pasquali que, além de orientador deste trabalho, vem iluminando minha jornada acadêmica há 12 anos.

Um agradecimento especial a quem tornou possível a realização desse trabalho. Aos diretores da DAEB que passaram por aquela diretoria desde 2001: na pessoa de Isa Locatelli, que foi a primeira diretora a abrir o banco de dados para a pesquisa; ao Carlos Henrique Araújo que apoiou o projeto e cedeu todos os dados para esse estudo; ao Amaury P. Gremaud, que disponibilizou todas as informações que eu precisei ao final do trabalho; à Coordenação do Banco Nacional de Itens que apoiou e incentivou essa pesquisa. Aos colegas da DAEB que, ao longo de uma jornada de trabalhos, mantiveram-se abertos para discussões que muitas vezes foram elucidativas. Agradeço, ainda, de modo especial à colega e amiga Elba M. L. Gomes, que foi minha interlocutora em meus momentos conflituos, ponderando minhas idéias.

Meu reconhecimento à amizade, ao carinho e às grandes contribuições dadas pela Professora Vera Figueiredo que muito me ajudou na organização desta tese.

A duas pessoas, em especial, que embora não tenham conhecimento da influência que exerceram sobre o desenvolvimento de minha tese, foram de fundamental importância: a Carmilva Souza Flores, que há cinco anos atrás me municiou com ricos materiais que me inspiraram a escolher esse tema, e ao Theodorus A. van Batenburg, que em sua rápida passagem pelo Brasil, em 2000, compartilhou comigo muito de sua experiência na área de avaliação educacional.

Aos meus familiares, de modo especial à minha mãe e aos meus irmãos, que muito colaboraram com discussões, reflexões e incentivos constantes à minha caminhada e ao meu crescimento.

Agradeço, ainda, às instituições que me deram oportunidade de trabalho na área, contribuindo fortemente para o aumento do meu conhecimento. Entre essas, à DAEB, ao CESPE, ao CAEd, à Missão Criança e ao Instituto Movens.

Nesse momento final, gostaria de deixar aqui registrados minha sincera gratidão e reconhecimento a todos que não foram citados, mas de uma forma ou de outra, contribuíram para o êxito desse trabalho.

SUMÁRIO

SUMÁRIO	VIII
LISTA DE TABELAS	XI
LISTA DE FIGURAS E QUADROS	XIV
APÊNDICE – LISTA DE TABELA E QUADRO	XV
ANEXO – LISTA DE QUADROS	XVII
RESUMO	XVIII
ABSTRACT	XIX

Introdução	1
-------------------------	----------

PARTE I – MARCO REFERENCIAL

CAPÍTULO 1

HISTÓRICO DA AVALIAÇÃO EDUCACIONAL

1. Histórico	8
Avaliação das Escolas Rurais do Nordeste - Edurural	16
São Paulo: Projeto Inovações no Ensino Básico - IEB	17
Minas Gerais	18
Ceará	18
O Sistema Nacional de Avaliação da Educação Básica – SAEB	19

CAPÍTULO 2

TESTES DE DESEMPENHO: CONSTITUIÇÃO, ANÁLISES E DESCRIÇÃO DOS RESULTADOS

2.1. Constituição de um teste para avaliação educacional	24
2.2. Análises dos testes de desempenho	30
2.2.1. Análise pedagógica dos itens	30
2.2.2. Análises psicométricas mais comuns utilizadas na avaliação de testes de desempenho	32
2.2.2.1. A Teoria Clássica dos Testes	32
2.2.2.2. A Teoria de Resposta ao Item	33
2.2.2.2.1. Itens dicotomicamente escorados	34

2.2.2.2.2. Itens politomicamente escorados	36
2.2.2.3. Análise da dificuldade dos itens.....	39
2.2.2.4. Análise da discriminação dos itens	40
2.2.2.4.1. Análise da discriminação por meio da TCT.....	40
a) Grupos critério	40
b) Correlação item-total	41
c) Correlação inter-itens	42
2.2.2.4.2. Análise da discriminação por meio da TRI.....	43
2.2.2.5 Análise da unidimensionalidade dos itens	44
2.2.2.5.1. O método da análise fatorial <i>full information</i>	45
2.2.2.5.2. O teste de decisão do número de fatores	46
2.2.2.6. Análise gráfica dos itens	48
2.2.2.7. Considerações sobre as análises.....	51
2.3. Descrição dos Resultados.....	54

PARTE II – METODOLOGIA, ANÁLISE E DISCUSSÃO DOS RESULTADOS

CAPÍTULO 3

MÉTODO	59
3.1. Material	60
3.2. Procedimentos metodológicos para análise dos dados.....	60

CAPÍTULO 4

PESQUISA DO PROCESSO DE AVALIAÇÃO BRASILEIRO SOB A PERSPECTIVA DO MODELO

NORTE-AMERICANO

4.1. Propósitos para o ensino e características do sistema educacional	63
4.1.1. Propósitos do ensino.....	63
4.1.2. Características do sistema educacional	69
4.2. Sistemas de avaliação.....	72
4.2.1. O que referenciam as avaliações nos dois países?	79
4.2.1.1. O que se avalia em Leitura nos dois países?	84
4.2.1.2. O que se avalia em Matemática nos dois países?	91
4.2.2. Como se procede às análises dos dados nos dois países?	98
4.2.3. Como os resultados são relatados nos dois países?	106

CAPÍTULO 5

ANÁLISE CRÍTICA DOS TESTES UTILIZADOS NA PERSPECTIVA DAS DESCRIÇÕES DAS ESCALAS

NO PERÍODO DE 1995 A 2003	113
5.1. SAEB – Edição 1995.....	115
5.1.1. Avaliação de Língua Portuguesa.....	115
5.1.2. Avaliação de Matemática.....	119
5.2. SAEB – Edição 1997.....	123
5.2.1. Avaliação de Língua Portuguesa.....	124
5.2.2. Avaliação de Matemática.....	128
5.3. SAEB – Edição 1999.....	132
5.3.1. Avaliação de Língua Portuguesa.....	132
5.3.2. Avaliação de Matemática.....	137
5.4. SAEB – Edição 2001.....	143
5.4.1. Avaliação de Língua Portuguesa.....	144
5.4.2. Avaliação de Matemática.....	150
5.5. SAEB – Edição 2003.....	160
5.1. Avaliação de Língua Portuguesa.....	160
5.2. Avaliação de Matemática.....	164

CAPÍTULO 6

ANÁLISE CRÍTICA DOS RELATOS DOS RESULTADOS DO SAEB	174
6.1. Edição 1995.....	175
6.2. Edição 1997.....	177
6.3. Edição 1999.....	179
6.4. Edição 2001.....	182
6.5. Edição 2003.....	187
6.6. Edição 2005.....	188

CAPÍTULO 7

PESQUISA DE POLÍTICAS PÚBLICAS IMPLEMENTADAS COM BASE NOS RESULTADOS NOS RESULTADOS DO SAEB	190
--	-----

PARTE III – CONSIDERAÇÕES FINAIS

CONSIDERAÇÕES FINAIS	199
REFERÊNCIAS BIBLIOGRÁFICAS	210
APÊNDICE	224
ANEXO	230

Lista de Tabelas

TABELA 1	Orientação para elaboração de itens abordando elementos literários e aspectos da leitura do NAEP	83
TABELA 2	Distribuição percentual dos itens por tópico Língua Portuguesa - SAEB.....	86
TABELA 3	Proporção de itens por contexto e grau	90
TABELA 4	Distribuição percentual dos itens por tema Matemática - SAEB.....	94
TABELA 5	Subtópico Operações com números relacionado com Propriedades dos números e operações.....	96
TABELA 6	Distribuição percentual dos itens por grau e área de conteúdo	97
TABELA 7	Exemplo de BIB	100
TABELA 8	Exemplo de BIPB.....	100
TABELA 9	Definição política dos níveis de desempenho do NAEP	107
TABELA 10	Níveis de desempenho em Leitura para o 4º grau	108
TABELA 11	Mapa de descrição dos itens de Matemática para o Grau 4 (NAEP 2005) ...	110
TABELA 12	Informações relacionadas aos testes analisados (SAEB)	115
TABELA 13	Estatísticas – Testes de Língua Portuguesa 4ª série – SAEB Edição 1995 .	117
TABELA 14	Descrição e percentual de alunos nos níveis com base no teste de Língua Portuguesa - 1995.....	117
TABELA 15	Níveis de Proficiência de Língua Portuguesa publicados e percentuais de alunos nos níveis – SAEB 1995.....	118
TABELA 16	Estatísticas – Testes de Matemática 4ª série – SAEB Edição 1995	120
TABELA 17	Descrição e percentual de alunos nos níveis com base no teste de Matemática - 1995	120
TABELA 18	Níveis de Proficiência de Matemática publicados e percentuais de alunos nos níveis – SAEB 1995	121
TABELA 19	Conteúdos da Matriz de Língua Portuguesa 1997	124
TABELA 20	Estatísticas – Testes de Língua Portuguesa 4ª série – SAEB Edição 1997 .	126
TABELA 21	Descrição e percentual de alunos nos níveis com base no teste de Língua Portuguesa - 1997.....	126
TABELA 22	Níveis de Proficiência de Língua Portuguesa publicados e percentuais de alunos nos níveis – SAEB 1997.....	127
TABELA 23	Estrutura da Matriz de Matemática 1997	128

TABELA 24	Estatísticas – Testes de Matemática 4ª série – SAEB Edição 1997	129
TABELA 25	Descrição e percentual de alunos nos níveis com base no teste de Matemática - 1997.....	130
TABELA 26	Níveis de Proficiência de Matemática publicados e percentuais de alunos nos níveis – SAEB 1997	131
TABELA 27	Estrutura da Matriz de Língua Portuguesa 1999	133
TABELA 28	Estatísticas – Testes de Língua Portuguesa 4ª série – SAEB Edição 1999 .	135
TABELA 29	Descrição e percentual de alunos nos níveis com base no teste de Língua Portuguesa - 1999.....	135
TABELA 30	Níveis de Proficiência de Língua Portuguesa publicados e percentuais de alunos nos níveis – SAEB 1999.....	137
TABELA 31	Estrutura da Matriz de Matemática 1999	138
TABELA 32	Estatísticas – Testes de Matemática 4ª série – SAEB Edição 1999	139
TABELA 33	Descrição e percentual de alunos nos níveis com base no teste de Matemática - 1999.....	139
TABELA 34	Níveis de Proficiência de Matemática publicados e percentuais de alunos nos níveis – SAEB 1999	141
TABELA 35	Estrutura da Matriz de Língua Portuguesa 2001	145
TABELA 36	Estatísticas – Testes de Língua Portuguesa 4ª série – SAEB Edição 2001 .	146
TABELA 37	Descrição e percentual de alunos nos níveis com base no teste de Língua Portuguesa - 1999.....	146
TABELA 38	Níveis de Proficiência de Língua Portuguesa publicados e percentuais de alunos nos níveis – SAEB 2001.....	149
TABELA 39	Estrutura da Matriz de Matemática 2001	149
TABELA 40	Estatísticas – Testes de Matemática 4ª série – SAEB Edição 2001	152
TABELA 41	Descrição e percentual de alunos nos níveis com base no teste de Matemática - 2001.....	152
TABELA 42	Níveis de Proficiência de Matemática publicados e percentuais de alunos nos níveis – SAEB 2001	155
TABELA 43	Construção de competências e desenvolvimento de habilidades de leitura de textos de gêneros variados em cada um dos estágios para 4ª série do EF....	157
TABELA 44	Construção de competências e desenvolvimento de habilidades de resolução de problemas em cada um dos estágios para 4ª série do EF	158

TABELA 45	Comparação entre a descrição do estágio crítico e as descrições do níveis de proficiência 125 ao 175	159
TABELA 46	Estatísticas – Testes de Língua Portuguesa 4ª série – SAEB Edição 2003 .	161
TABELA 47	Descrição e percentual de alunos nos níveis com base no teste de Língua Portuguesa - 2003	161
TABELA 48	Estatísticas – Testes de Matemática 4ª série – SAEB Edição 2003	165
TABELA 49	Descrição e percentual de alunos nos níveis com base no teste de Matemática - 2003	165
TABELA 50	Aproveitamento de itens por teste /edição do SAEB	170
TABELA 51	Síntese dos principais problemas apontados por edição pela análise pedagógica	171
TABELA 52	Síntese dos parâmetros obtidos das análises da qualidade psicométrica dos testes do SAEB	172
TABELA 53	Número de estabelecimentos de ensino da educação básica e ensino fundamental instalados na zona urbana e rural em 2005	203
TABELA 54	Número de estabelecimentos de educação do ensino fundamental com até trinta-alunos	203

Lista de Figuras e Quadros

Lista de Figuras

Figura 1	ANÁLISE GRÁFICA DE UM BOM ITEM	50
Figura 2	ANÁLISE GRÁFICA DE UM ITEM RUIM	51

Apêndice

Tabela

Tabela I	PARÂMETROS PSICOMÉTRICOS DOS ITENS SELECIONADOS PARA DESCRIÇÃO DO TESTE LP/1997	226
----------	--	-----

Quadro

QUADRO 1 – Análise pedagógica de item.....	225
--	-----

Anexo

Lista de Quadros

QUADRO 1	Tópico I – Procedimentos de Leitura	231
QUADRO 2	Tópico II – Implicações do suporte, do gênero e/ou enunciador na compreensão do texto	231
QUADRO 3	Tópico III – Relação entre textos	231
QUADRO 4	Tópico IV – Coerência e coesão no processamento do texto.....	232
QUADRO 5	Tópico V – Relação entre recursos expressivos e efeitos de sentido	232
QUADRO 6	Tópico VI – Variação lingüística	232
QUADRO 7	Matrizes de Matemática referentes aos três ciclos	233
QUADRO 8	Contextos especificados para leitura	236
QUADRO 9	Aspectos da leitura descritos na Matriz do NAEP	236
QUADRO 10	Questões e estratégias de leitura por contexto e aspecto (tabela de especificação)	236
QUADRO 11	Amostra de itens do NAEP por Elementos de Textos Literários e Aspectos de Leitura	237
QUADRO 12	Amostra de itens do NAEP por Elementos de Textos Informativos e Aspectos da Leitura	237
QUADRO 13	Amostra de itens do NAEP por Elementos de Textos Práticos e Aspectos da Leitura	238
QUADRO 14	Sentido dos números	240
QUADRO 15	Estimação	241
QUADRO 16	Operações com números	241
QUADRO 17	Razão e Proporção	242
QUADRO 18	Propriedade dos números e operações	243
QUADRO 19	Atributos físicos da medida	243
QUADRO 20	Sistemas de medida	244
QUADRO 21	Dimensão e Forma	245
QUADRO 22	Transformação de formas e preservação das propriedades.....	246
QUADRO 23	Relações entre figuras geométricas	247
QUADRO 24	Posição e direção	247

QUADRO 25 Raciocínio Matemático	248
QUADRO 26 Representação dos dados	248
QUADRO 27 Características dos conjuntos de dados	249
QUADRO 28 Experimentos e amostras	250
QUADRO 29 Probabilidade.....	250
QUADRO 30 Padrões, relações e funções.....	251
QUADRO 31 Representações algébricas.....	252
QUADRO 32 Variáveis, expressões e operações	253
QUADRO 33 Equações e inequações	253

RESUMO

Este estudo está fundamentado em uma análise crítica descritiva do Sistema Nacional de Avaliação da Educação Básica (SAEB), realizada com base em pesquisa bibliográfica e documental. Teve por principal objetivo analisar a avaliação empreendida por esse sistema na perspectiva do instrumento que estima o desempenho dos alunos. Os dados do SAEB que foram objeto deste estudo foram os testes de Língua Portuguesa e Matemática, 4ª série, aplicados nas edições do SAEB no período de 1995 a 2003, as matrizes de referência, as escalas de proficiência e os relatórios publicados referentes a esse período. Os testes que avaliam o desempenho dos alunos são os principais instrumentos de informação que subsidiam os múltiplos propósitos dos sistemas de avaliação hoje em curso, e que orientam o alcance da superação dos problemas. A realização deste estudo implicou, em primeiro lugar, contextualizar a avaliação educacional por meio de sua trajetória histórica. Em seguida, pesquisar a constituição dos testes, os tipos de análises utilizadas na atualidade e os tipos de descrições mais comuns dos resultados dos testes. Em seguida, foi realizada pesquisa do processo de avaliação brasileiro na perspectiva do modelo norte-americano, cujo sistema se constitui como nossa referência. O passo seguinte foi proceder a uma análise crítica das avaliações do SAEB, no período de 1995 a 2003, considerando-se as habilidades aferidas e publicadas na forma de escalas e dos níveis de desempenho. Essa análise redundou a (re)descrição de todas as escalas, com a intenção de mostrar a redução apresentada na descrição das escalas publicadas. O procedimento seguinte foi a análise crítica dos relatórios de resultados publicados pelo SAEB e, por fim, pesquisou-se políticas implementadas com base nos resultados do SAEB, nas secretarias de educação de estado e no Ministério da Educação e Cultura (MEC). Concluiu-se que é premente a necessidade de revisão das matrizes de referência; que os testes de desempenho deverão ser constituídos com maior cuidado; que as escalas de proficiência devem descrever de modo mais amplo as habilidades demonstradas pelos alunos brasileiros; que as publicações de resultados devem ser dirigidas, com linguagem adequada a diferentes públicos (pesquisadores, elaboradores de políticas, educadores e comunidade em geral) e que o SAEB deveria orientar ações que pudessem ser transformadas em políticas pelos estados e por secretarias do MEC, além de manter diálogo com essas instituições. Contudo, ressalta-se o importante papel que o SAEB representa para o país na disseminação da cultura avaliativa, no avanço das análises psicométricas e no incentivo do desenvolvimento de instituições especializadas em avaliação.

Palavras-chave: avaliação em larga escala (SAEB); constituição de um teste de desempenho; relato de resultados; políticas públicas.

ABSTRACT

This study is based on a descriptive critical analysis of the National System for Basic Education Evaluation (Sistema Nacional de Avaliação da Educação Básica - SAEB) carried through on the basis of documentary bibliographical research. It took in consideration the tools applied to measure the achievement of fourth-grade students in Mathematics and Portuguese Language, during the period from 1995 to 2003, and the matrices of reference, the scales of proficiency and the reports published referring to this period. The tests that evaluate the performance of the pupils are the main instruments of information that today subsidize the multiple intentions of the systems of evaluation in course, and that they guide the reach of the overcoming of the problems. The accomplishment of this study implied, in first place, to contextualize the educational evaluation by means of its historical trajectory. This study also researched the modeling of the achievement tests, the types of analyses used in the present time and the types of more common descriptions of the results of the tests. After that, a research of the Brazilian process of evaluation in the perspective of the North American model was carried through, whose system constitutes as our reference. The following step was to proceed to a critical analysis of the evaluations of the SAEB, in the period above mentioned, considering the abilities surveyed and published in the form of scales and the levels of performance. As a result of the evaluation this work presents a new array of skills and achievement, in order to show the reduction attained in the description of the published scales. The following procedure was critical analysis of the reports of results published by SAEB. Also, it investigated which policies were based on the results of SAEB. Finally, it was considered the influence of SAEB reports on policymakers, the Ministry of Education (MEC) and Education Secretariats in the states. The conclusion reveals that it is urgent to review the reference matrices, that the performance tests will have to be constituted with well-taken care to attune the achievement tests; that the proficiency scales must describe in ampler way the abilities demonstrated for the Brazilian pupils; that the publications of results must be addressed with adequate language to different types of public (researchers, policymakers, teachers and members of the general public); and that SAEB would have to guide actions which could be transformed into policy for the states and secretariats of the MEC, beyond keeping dialogue with these institutions. The role that SAEB represents for the country is emphasized in relation with the dissemination of the assessment culture, the advance of psychometric analysis and in relation with the incentive of the development of institutions specialized in evaluation.

Key -words: large scale assessment; modeling of achievement tests; report of results; public policies.

INTRODUÇÃO

O processo de avaliação educacional é um processo complexo e abrangente e, por isso, definido de várias formas. Dietel, Herman & Knuth (1991) conceituam avaliação educacional como *qualquer método usado para melhor entender o conhecimento corrente que um estudante possui*, incluindo desde a simples avaliação realizada pelo julgamento subjetivo baseado numa única observação do desempenho do estudante até um teste padronizado complexo que exige técnicas de análise sofisticadas. Esses autores afirmam, ainda, que a idéia do conhecimento corrente implica que aquilo que um estudante sabe está sempre mudando e, dessa forma, julgamentos mais consistentes acerca do desempenho dos estudantes deveriam ser feitos por meio de comparações realizadas em um certo período de tempo.

A realização da avaliação no âmbito educacional pode alcançar diversos propósitos e ser realizada sob várias perspectivas. Os autores citados abordam, por exemplo, como propósitos de um orientador de políticas públicas, monitorar a qualidade da educação, ajustar níveis de escolarização, avaliar objetivos alcançados, formular políticas educacionais, direcionar recursos, incluindo pessoal e dinheiro. Como propósitos de um avaliador de programas, identificar forças e fraquezas em programas implementados, designar prioridades a determinado programa, planejar e melhorar programas. Como propósitos de professores e diretores, fazer uma reflexão da prática pedagógica adotada, diagnosticar o desempenho e prescrever fórmulas de melhoria, monitorar o progresso dos estudantes; avaliar e refinar o currículo implementado. E os pais podem e devem usufruir dos dados da avaliação para

estimar o sucesso ou o fracasso de seus filhos relacionados ao processo de ensino ou, ainda, como referência na prestação de contas da própria escola, entre outros.

Para realizar este estudo, tomou-se como referência o Sistema Nacional de Avaliação da Educação Básica – SAEB – que, segundo consta em sua apresentação, tenciona abranger todas as dimensões citadas por Dietel, Herman & Knuth (1991).

Para melhor compreender o que se pretende abordar neste estudo, é importante apontar as diferentes perspectivas sob as quais a avaliação se encontra relacionada, e que, muitas vezes, são causadoras de confusão e de mau entendimento do processo em si. Essas perspectivas se relacionam às diversas acepções que toma o termo **avaliação**, cujos processos podem ser registrados como mensuração, estimativa, classificação, prestação de contas (*accountability*), apreciação, entre outros.

A avaliação na perspectiva de **mensuração** advém da referência que se faz ao processo pelo qual os atributos ou dimensões de algum objeto físico são determinados (Kizlik, 2006). Wiersma & Jurs (1990), afirmam que, tecnicamente, mensuração é a designação de números para objetos ou eventos de acordo com regras que dão um significado quantitativo aos numerais. Popham (1975) diz que a aproximação da avaliação à mensuração decorre do fato de que educadores americanos desprenderam grande quantidade de tempo e projeção na medida da habilidade humana, gerando um falso conceito para grande parte da massa popular. Segundo ele, a mensuração na educação é simplesmente o ato de determinar o grau no qual um indivíduo possui certo atributo, ou é a designação de algum tipo de índice numérico ao desempenho medido de uma pessoa ou de um grupo, de maneira que se possa mais precisamente representar aquela situação com respeito ao atributo de interesse.

Estimativa é outro sentido comum atribuído ao termo avaliação que, segundo Popham (1975), pode ser uma forma atenuada de abordá-la causando menos furor na população que é objeto da avaliação. Assim, em vez de dizer ao professor que sua prática está sendo avaliada por meio de instrumentos, diz-se que tais práticas estão sendo estimadas. É importante ressaltar que tal termo está muito mais associado à mensuração, pois se relaciona à determinação, em termos probabilísticos, de parâmetros de uma população, o que significa dizer que se trata muito mais de uma medida sem valor de juízo.

Outra palavra que é associada analogamente à avaliação é **classificação**, que tem como origem a distribuição sistemática dos alunos em salas de aula em diversas categorias, classificação essa fundamentada no sistema de notas a eles atribuídas que, por sua vez, tem

como origem, principalmente, a estimativa de mérito (Popham, 1975), e que por esse motivo confunde-se com o termo avaliação. A diferença do significado está no enfoque, que, ao invés de estar pautado no fenômeno educacional, está nas unidades de análise individuais. Esse caráter é comum e se faz presente em diversas avaliações. Por exemplo, ao compararem-se médias de proficiência de regiões de um país, ou de unidades da federação, ou de escolas de um município, realiza-se um processo classificatório, desvinculando-os dos fatores intervenientes do sucesso ou insucesso.

A avaliação tem outra abordagem conotativa: a da **prestação de contas** ou *accountability*. Essa abordagem é uma das mais consideradas no âmbito da avaliação educacional de sistemas. O termo avaliação, nesse sentido, engloba uma concepção originada principalmente da imposição externa da sociedade ou do Estado de evidenciar resultados, ou de prestar contas sem, no entanto, estabelecer um juízo de valor que reverteria em ações efetivas de melhoria do objeto avaliado. Pelo fato de sua concepção estar relacionada a aspectos políticos, ao ser considerada como uma possibilidade de dar respostas, ou responsabilizar, ou cobrar uma dívida, o processo avaliativo acaba se tornando um instrumento gerador de indicadores estatísticos, os quais acabam se revertendo em justificativas superficiais de bons ou maus desempenhos sem, entretanto, se tornarem instrumentos de melhoria do sistema. Eraut (1991), chama a atenção para o sentido restrito de simples prestação de contas que tende a atribuir responsabilidades, em especial quando ocorre insucesso, exclusivamente, a determinadas categorias, deixando de considerar toda a gama de fatores que o determinam.

A origem desses diferentes sentidos atribuídos ao termo avaliação deriva, principalmente, das diversas realizações de avaliações no âmbito educacional ocorridas nos últimos 50 anos. Pela falta de experiência, estudos e publicações, as instituições responsáveis pelas avaliações foram se apropriando dessas diversas concepções. Aliado a esse motivo, houve o grande avanço nas medidas educacionais, com o desenvolvimento de modernas técnicas de análises psicométricas.

Entre os objetivos do SAEB, registrados em diversos documentos, constam: a oferta de subsídios à formulação, reformulação e monitoramento de políticas públicas e programas de intervenção ajustados às necessidades diagnosticadas nas áreas e etapas de ensino avaliadas; a identificação dos problemas e das diferenças regionais do ensino; a produção de informações sobre os fatores do contexto socioeconômico, cultural e escolar que influenciam o desempenho dos alunos; o ensejo aos agentes educacionais e à sociedade de uma visão clara

dos resultados dos processos de ensino e aprendizagem e das condições em que são desenvolvidos e o desenvolvimento de competência técnica e científica na área de avaliação educacional, ativando o intercâmbio entre instituições educacionais de ensino e pesquisa.

A proposta para este estudo foi motivada, principalmente, pelo interesse da pesquisadora em compreender o real funcionamento e alcance do sistema de avaliação nacional. Pois, ao tomar conhecimento de sua complexidade e sofisticação, observou-se um grande distanciamento entre o que é praticado e o que é proposto como missão. E fazendo-se uma reflexão sobre os diversos sentidos que são atribuídos ao termo avaliação educacional, e tomando-os no contexto sistêmico, considerou-se importante desvendar, por meio de pesquisa, o sentido que o sistema brasileiro atribui à avaliação. Pretendeu-se, ainda, analisar a avaliação na perspectiva do instrumento que avalia o desempenho dos alunos.

O desenvolvimento deste estudo não se apoiou simplesmente em concepções isoladas, mas, principalmente, em práticas exitosas adotadas por sistemas comprometidos que se empenham com a melhoria da educação no mundo.

Para alcançar esse intento, foi necessário, inicialmente, realizar pesquisa sobre o que pautam as avaliações que realmente estejam melhorando a qualidade educacional dos seus sistemas, além de investigar o processo de elaboração dos testes, das técnicas de análise dos dados e da interpretação e disseminação dos resultados.

Ao analisar sistemas de avaliação de vários países do mundo, constatou-se que a maioria deles toma como modelo o sistema norte-americano, cada um adequado às suas particularidades. No Brasil, o SAEB também adotou como referência o sistema norte-americano. Por esse motivo, um dos focos deste estudo foi a realização de uma pesquisa do processo de avaliação brasileiro sob a perspectiva do modelo norte-americano.

Sabe-se que os testes que avaliam o desempenho dos alunos são os principais instrumentos de informação que subsidiam os múltiplos propósitos dos sistemas de avaliação hoje em curso, e que orientam o alcance da superação dos problemas apontados. Considerando-se que um teste é uma forma padronizada de representar o comportamento e sendo ele concebido como o ponto central da avaliação, a sua construção implica o conhecimento detalhado da realidade a ser pesquisada e a formulação de objetivos bem definidos para a testagem.

O teste deve ser válido e fidedigno para assegurar que as informações fornecidas avaliem com precisão a realidade pesquisada, de forma a se poder tomar decisões indubitáveis

a seu respeito. Neste estudo são apontadas as principais análises que asseguram a qualidade pedagógica e psicométrica do teste que avalia o desempenho dos alunos.

Apesar de todos os avanços recentes na composição de um sistema de avaliação, existem ainda inúmeras questões de ordem técnica, social e política que demandam muito estudo. A realização dessa pesquisa envolveu o conhecimento das particularidades que compõem o sistema educacional brasileiro. Foram levantadas informações sobre os instrumentos que avaliam o desempenho, as escalas de proficiência e os níveis de desempenho produzidos com base nos resultados dos testes, os relatórios publicados e as políticas adotadas fundamentadas nesse conjunto de dados.

As questões norteadoras deste estudo foram:

- (1) Como deve ser constituído um teste que pretende avaliar o processo de construção do conhecimento dos alunos brasileiros e que informações ele pode oferecer ao sistema?
- (2) Como deve ser constituída e o que deve ser descrito em uma escala de proficiência?
- (3) Como devem ser apresentados os resultados para que sejam compreendidos por todos os agentes capazes de promoverem melhorias no sistema?
- (4) Que políticas têm sido implementadas com base nos resultados do SAEB?
- (5) Quais contribuições o SAEB pode fornecer para a adoção de políticas públicas que busquem melhorar a qualidade da educação básica brasileira?

Para tanto, a pesquisa delineada fundamentou-se basicamente nos dados fornecidos pelo SAEB. Foram investigados os testes com os resultados das análises psicométricas de Língua Portuguesa e de Matemática, de 4ª série, a partir da edição do SAEB 1995; as matrizes de referência de 1997, 1999 e 2001; as escalas de proficiência e os relatórios de resultados apresentados neste período.

O ano 1995 foi determinado pelo fato de ter tido início naquela edição, a nova metodologia adotada para análise dos resultados, a Teoria de Resposta ao Item – TRI. As disciplinas foram escolhidas em decorrência de, desde 2001, elas serem as únicas avaliadas, embora até 1999, houvesse a testagem de outras nove disciplinas. Considerando-se que a pesquisadora tem formação em Psicologia e Pedagogia, ela sentiu-se apta para discutir questões relacionadas à 4ª série. Além dos testes, foram estudadas as escalas de proficiência produzidas desde 1995 e os relatórios de resultados de todas as edições do SAEB. Foram,

ainda, investigadas as políticas implementadas com base nos resultados do SAEB, a partir de 1995.

Esta tese está estruturada em três partes. Na primeira parte, está o marco referencial, com dois capítulos. O primeiro, contextualiza a avaliação educacional por meio de sua trajetória histórica, e o segundo capítulo trata da constituição, dos tipos de análises e das descrições mais comuns dos resultados de testes de desempenho. A segunda parte é denominada “Metodologia, análise e discussão dos resultados” e é constituída de cinco capítulos. O primeiro dessa segunda parte (terceiro da tese), apresenta a metodologia da pesquisa; o quarto capítulo apresenta a pesquisa do processo de avaliação brasileiro sob a perspectiva do modelo norte-americano; o quinto capítulo aborda a análise crítica das avaliações do SAEB (1995 a 2003), em termos das habilidades aferidas e publicadas na forma de escalas e níveis de desempenho; o sexto capítulo apresenta a análise crítica dos relatórios de resultados do SAEB e o sétimo capítulo trata da pesquisa de políticas implementadas com base nos resultados do SAEB, realizada na Secretaria de Políticas Públicas para o Ensino Fundamental do Ministério da Educação e Cultura e nas Secretarias de Estado de Educação. Na última parte do trabalho, denominada “Considerações Finais”, são apresentadas as impressões sobre a realização da pesquisa e são sugeridas algumas respostas às questões que permeiam o trabalho.

1ª PARTE

MARCO REFERENCIAL

1. HISTÓRICO DA AVALIAÇÃO EDUCACIONAL

Embora a avaliação tenha um histórico antigo enquanto atividade voltada para o desempenho individual, nos últimos 30 anos é que ela ancorou-se na eficácia de sistemas educacionais de forma a adotar um caráter formal enquanto atividade cônica de seus propósitos e de seus objetivos. Atualmente, tem-se assumido mais fortemente a influência que a avaliação exerce sobre a educação. Decisões baseadas na própria avaliação informal, ou por meio de impressões ou percepções que as pessoas formam acerca de alternativas educacionais têm grande peso tanto na parte relativa à educação quanto na própria aprendizagem.

A testagem de aptidões tem um histórico que remonta à antiguidade. Os primeiros registros datam de 2000 anos a.C. com relatos do sistema de exames no serviço civil utilizado no império chinês. A literatura indica que ainda antes da era cristã, professores gregos, como Sócrates (470 a.C.), realizavam verificações do processo de aprendizagem de seus alunos por meio de avaliações verbalmente mediadas. Já no início da era cristã, há registros de que a China tenha sido a primeira nação a adotar a prática da medida de desempenho individual padronizada. Consta que seu primeiro exame nacional imperial, chamado “*Ke Ju*”, parece ter sido equivalente aos exames atuais de entrada para universidade, os vestibulares (Wang, 2006). Esse evento teve início no ano 605, na Dinastia Sui (anos 581 – 618). O seu propósito era selecionar altos funcionários para administração do império. A essa seleção, que tinha critérios rígidos e, segundo consta, de fato selecionava os candidatos melhores preparados, é atribuída uma grande prosperidade da nação, pois foram realizadas reformas e reajustes na administração, levando o império chinês a uma posição avançada nas áreas política, econômica e cultural no mundo.

No entanto, avaliações formais de programas educacionais só foram relatadas a partir da metade do século XIX. Até a metade do século XVIII, as escolhas educacionais se

pautavam pelas crenças políticas e religiosas e independente da competência demonstrada pelos professores, a culpa pelo fracasso da aprendizagem sempre recaía sobre o aluno (Travers, 1983).

A prática de coletar dados para fundamentar e apoiar decisões educacionais foi introduzida pelos americanos Henry Barnard e Horace Mann entre outros menos influentes, no período entre 1838 e 1850, nos Departamentos de Massachusetts e Connecticut e foi continuado no *United States Education Bureau*. Nesse período, Mann submeteu 12 relatórios anuais ao *Board of Education of the Commonwealth of Massachusetts*, os quais apresentavam registros de todos os problemas educacionais correntes com suporte empírico. Esse foi um marco na avaliação educacional formal (Worthen & Sanders, 1987).

Em 1845, o *Boston School Committee* foi o responsável pelo primeiro uso do teste impresso para avaliação de estudantes em larga escala. Prática que foi repetida em 1846, em decorrência do baixo desempenho constatado, em amostragem de estudantes. Em 1847 foi descontinuada porque não foi feito nenhum uso dos resultados. Assim, foi realizado o primeiro debate precursor dos dias atuais sobre o valor do uso da testagem com fins de avaliar a qualidade de um sistema educacional em larga escala.

Depois dessa primeira experiência, durante o período de 1895 a 1905, Josef Rice organizou um programa de avaliação semelhante para sistemas de educação em larga escala nos Estados Unidos. Rice foi um crítico dos sistemas educacionais daquela época e, por esse motivo, ele teve medo de que o seu julgamento pudesse ser rejeitado como sendo as reclamações de um observador decepcionado. Para precaver-se, ele começou a documentar suas afirmações de que o tempo de estudos era usado ineficientemente e sua crença foi ratificada por meio de pesquisas que foram mostrando grandes diferenças nos desempenhos entre escolas, o que demandou a determinação de exames padronizados.

A grande contribuição de Rice, entretanto, foi em relação à implementação de um painel de juizes para resolver questões de posições controversas. Esse processo foi proposto como um meio de eliminar subornos e desperdícios no governo e foi uma pequena e breve forma conhecida, que depois surgiria como uma defesa ou uma abordagem que daria suporte à avaliação.

No início do século XX, Thorndike foi aclamado o pai do movimento da testagem educacional. Ele ajudou a persuadir educadores de que valia a pena medir as mudanças do comportamento humano. Nesse mesmo período, mais especificamente em torno das duas

primeiras décadas, as tecnologias de medida para determinar habilidades humanas prosperaram e a testagem surgiu como um dos principais meios de avaliar escolas nos Estados Unidos. No período da I Guerra Mundial, entre 18 e 40 sistemas educacionais tinham seus ‘*bureaus*’ de pesquisa educacional trabalhando em avaliações de desempenho de estudantes em larga escala, na forma de *surveys*. Nessa época, esses *surveys* foram usados para vários propósitos, entre eles, diagnosticar fraquezas específicas do sistema; padronizar práticas curriculares; avaliar experimentos; avaliar o desempenho geral de um sistema, bem como tomar decisões acerca dos estudantes (Cronbach, 1996).

O movimento de testagem tornou-se mais confiável e reconhecido com o desenvolvimento dos testes Alfa (para alfabetizados) e Beta (para analfabetos) criados para testarem soldados na I Guerra Mundial. Esses testes disseminaram a idéia de que boas decisões poderiam ser tomadas tendo por base a disponibilidade de testes objetivos. Durante esse período, houve um predomínio dos testes referenciados a critério (testes construídos para avaliar objetivos atingidos). Com a necessidade do uso de testes para medir níveis de desempenho individuais, houve uma demanda de desenvolvimento de testes referenciados a norma (testes construídos para avaliar o desempenho em relação ao grupo) (Anastasi, 2000; Cronbach, 1996; Pasquali, 1997, 2003);.

Um episódio marcante dessas duas primeiras décadas, segundo Worthen & Sanders, foi a avaliação do sistema de escolas públicas de Gary, Indiana, realizada por Abraham Flexner, iniciada em 1915. Tratou-se de um meio inovador de apontar as necessidades educacionais da comunidade de Gary. Essa avaliação parece ter sido requisitada pelo superintendente de escolas William Wirt que se convenceu de que suas escolas eram as melhores do país. Para comprovar tal hipótese, seria necessário que um grupo externo realizasse essa avaliação. O estudo, que consumiu considerável quantia em dinheiro, foi completado em 1918. Os resultados indicaram que os estudantes alcançaram um desempenho acadêmico inferior em comparação aos outros estudantes de outros estados norte-americanos. Mas o episódio marcante dessa avaliação foi o fato de que alguns pesquisadores acreditaram que a pesquisa tenha sido altamente enviesada. Esse primeiro registro de uma avaliação externa de um programa educacional controverso é destacado em Worthen & Sanders. Eles fazem uma reflexão sobre os problemas técnicos e políticos inerentes a avaliações polêmicas, que produzem enviesamentos e conduzem à desconfiança do processo. Eles citam a meta-avaliação e o desenvolvimento de padronização para avaliação como meio de superar essa adversidade dos sistemas.

O movimento de testagem continuou evoluindo nos anos 20 e 30. Há relatos de que cerca de metade dos estados americanos já havia adotado, nesse período, alguma forma de testagem. Os testes de desempenho padronizados foram naturalmente se desenvolvendo para o uso em programas de larga escala. Além disso, testes de desempenho produzidos por professores foram sendo desenvolvidos e formaram uma base para os sistemas classificatórios escolares. Durante esse período, houve também um marcante desenvolvimento de testes de interesse e de personalidade. Essas novas ferramentas foram adotadas pelo setor militar e por grande parte das indústrias privadas americanas para fins de seleção e classificação de pessoal.

Nesse período, os termos medida e avaliação foram considerados quase sinônimos. O termo “avaliação” era freqüentemente utilizado para significar a designação de graus ou a sumarização do desempenho dos estudantes nos testes, características atuais de “medida”.

Worthen & Sanders citam como um fato marcante dos anos 30, o “movimento de educação progressista”, o qual foi experimentado e adotado por alguns distritos americanos, tendo por base os currículos escritos por John Dewey, que foram avaliados informalmente, demonstrando que havia um melhor desempenho dos estudantes. Porém, críticos do movimento continuaram afirmando que alunos do ensino secundário (*high school*) de escolas “progressistas” apresentavam desempenho inferior àqueles de escolas tradicionais. Essa controvérsia levou à implementação de um plano formal de avaliação prevista para ser realizada ao final do 8º ano (*Eight Year Study*). Este foi mais um marco na história da avaliação americana, pois esse estudo se mantém popular até os dias atuais.

Outro marco desse período, segundo os autores referidos anteriormente, foi a contratação de Ralph Tyler como diretor da equipe de estudos em avaliação educacional americana. A partir de então, foi impressa uma outra dimensão ao desenvolvimento de instrumentos e procedimentos para medir o progresso educacional fundamentados na abordagem desenvolvida por ele, baseada nos objetivos instrucionais. Essa abordagem foi adotada oficialmente em 1942, com a publicação de um manual de avaliação que influenciou fortemente, a avaliação educacional.

Em 1949, foram lançados nos Estados Unidos, sob a tutela de Ralph Tyler, os princípios básicos do currículo e do ensino (*Basic Principles of Curriculum and Instruction, 1949*) que imprimiram um outro olhar para a educação americana, que teve reflexos em nosso sistema educacional. Nessa obra, Tyler descreve os princípios básicos que deveriam orientar o

ensino, a aprendizagem e a avaliação. De modo conciso, esses princípios que foram adotados por muitos sistemas de ensino são os seguintes:

- (1) definição dos objetivos apropriados para o processo de ensino, ou seja, deveriam ser discutidos os propósitos educacionais que o sistema deveria alcançar;
- (2) introdução de experiências úteis de aprendizagem, isto é, deveriam ser selecionadas experiências de aprendizagem que fossem úteis ao alcance desses objetivos;
- (3) organização das experiências de aprendizagem para maximizar o seu efeito, isto é, deveriam ser preconizadas e organizadas as práticas de ensino que tornassem a instrução efetiva;
- (4) avaliação do processo de ensino com revisão das áreas que não alcançam eficácia, isto é, deveria ser avaliada, e se fosse o caso, revista, a capacidade de obtenção de bons resultados do sistema por meio das experiências de aprendizagem.

Ainda hoje, os avaliadores que empregam objetivos instrucionais como base para determinar se um currículo ou programa obteve sucesso, são chamados avaliadores tylerianos. Vários estudiosos seguiram essa abordagem, entre eles, Bloom *et al.* (1956); Taba (1962); Kratwohl, Bloom e Masia (1964); Metfessel e Michael (1967); Bloom, Hasting e Madaus (1971). O Sistema Norte Americano de Avaliação do Progresso Educacional (NAEP) foi concebido por Tyler, que seguiu a abordagem da avaliação no 8º ano (*Eight Year Study*). A maioria dos programas estaduais de avaliação americana seguem ainda essa concepção.

Como reflexo da adoção oficial da avaliação educacional sistêmica, houve um grande aumento das agências de acreditação escolares, regionais e nacionais, as quais alcançaram o seu auge nos anos 30. Esse movimento se concentrou principalmente sobre recursos e processos usados nas escolas. Por exemplo, algumas agências de acreditação desenvolveram guias e critérios para monitorar a adequação dos meios de aprendizagem, as qualificações do *staff* e a adequação dos programas, em vez de focar sobre o *status* educacional dos graduados. Essas agências funcionavam desde o final do século XIX e início do século XX e, desde então, exerceram forte influência na avaliação americana.

Os anos 40 até os 60 foram marcados pela consolidação do processo de avaliação com considerável desenvolvimento de técnicas de análise, construção de testes fundamentada em uma base tyleriana, e aplicação em larga escala. Em 1956, foi publicado por Bloom e outros, o livro que exerceu e exerce ainda forte influência nos processos de avaliação educacional: *Taxonomy of Educational Objectives: Handbook I: Cognitive Domain*. O referido livro tornou-se um documento importante, o qual explicita, com detalhes, uma

hierarquia de habilidades de pensamento aplicáveis a qualquer disciplina acadêmica. Ele tornou-se um referencial na testagem e no desenvolvimento, delineamento e avaliação de currículos.

Segundo Worthen & Sanders, outro episódio que gerou impacto na educação americana foi o lançamento, em 1957, do foguete Sputnik I pela União Soviética, o que resultou em grandes mudanças no enfoque educacional. Tal fato produziu um sentimento de inferioridade no povo americano e uma reação política quase imediata. Em 1958, foi decretada a Lei em Defesa da Educação, quando milhões de dólares foram destinados ao desenvolvimento de novos programas educacionais, especialmente, nas áreas de Matemática e Ciências. Subseqüentemente, fundos foram disponibilizados para avaliar esses esforços, e em 1964, sob a direção de Ralph Tyler, teve início o *National Assessment of Educational Progress - NAEP*. Os estudos de avaliação decorrentes desse período revelaram um empobrecimento metodológico e conceitual para atingir os objetivos que essas avaliações exigiam alcançar. Assim, em muitos casos, foram constatados delineamentos inadequados, dados inválidos, análises inacuradas e relatórios irrelevantes para questões de avaliações importantes que deveriam ter sido propostas. A maioria dos estudos se baseava em combinações idiossincráticas e aplicações de conceitos e técnicas de delineamento experimental, psicometria, desenvolvimento curricular e, em menor grau, na própria pesquisa. O trabalho teórico relacionado à avaliação educacional em si, era quase inexistente. Poucos estudiosos voltaram sua atenção a desenvolver projetos aplicáveis à educação. Nesse período, a escassez de informações e a má condução das pesquisas foram observadas por Cronbach (1963) que produziu um artigo (*Course improvement through evaluation*) criticando as avaliações por sua inutilidade e chamando a atenção de especialistas em avaliação, propondo novas direções. Reflexões sobre esse artigo levaram pesquisadores a lançarem uma concepção mais expandida de avaliação. Em seguida a essa publicação, a Lei de Direitos Civis de 1964 (*Civil Rights Act of 1964*) conduziu ao Estudo Coleman que focou sobre igualdade de oportunidades às minorias. Desse estudo, foi produzido o Relatório Coleman sobre a qualidade das escolas públicas nos Estados Unidos (Coleman, 1966), o qual gerou literatura considerável sobre o tema, tanto em economia como na área da educação.

Mas, ainda em 1964, quando o Ministério da Educação Americano tornou legal o sistema de avaliação, constatou-se que não havia pesquisadores que atendessem efetivamente à demanda da avaliação. Dessa forma, foi necessário que o Estado requisitasse que as escolas liberassem seus melhores professores da sala de aula para servirem como avaliadores. As

qualificações exigidas é que tivessem experiência e preparação como professores. Assim, após os primeiros processos, verificaram-se inúmeras falhas que levaram o Ministério a rever e repensar uma saída para tal situação. O principal problema recaía sobre a falta de estudiosos com conhecimento em avaliação educacional que revertessem tal quadro. Mas foram implementadas várias políticas de incentivo a estudos dessa natureza e até o ano 1973, essa situação foi revertida. Houve, ainda, uma expansão de estudos em todo o mundo. Dessa forma, estudiosos ingleses, australianos, israelenses, suecos, entre outros, empreenderam pesquisas e trouxeram à luz inúmeras outras abordagens sobre avaliação. Observa-se que em torno do ano 1974, até os dias de hoje, foi extremamente profícua a produção literária sobre avaliação (Talmage, 1982). Inúmeras pesquisas, estudos, artigos científicos, manuais, técnicas e, até uma comissão sobre padrões para programas de avaliação educacional (*Standards for Evaluations of Educational Programs*, 1975 – Estados Unidos) têm sido desenvolvidos. Além disso, grande esforço tem sido empreendido por parte dos Estados Unidos em disseminar a cultura da avaliação por todas as nações do mundo. Vários organismos internacionais têm assumido esse compromisso e projetos, técnicas e diversas outras avaliações têm sido divulgadas.

No Brasil, as primeiras experiências registradas relacionadas à avaliação da aprendizagem, fora do contexto da sala de aula, referem-se aos exames vestibulares, que surgiram em 1911, por uma lei implementada pelo então Ministro da Justiça e dos Negócios Interiores, Rivadávia da Cunha Corrêa. Embora o ingresso às faculdades continuasse garantido pelos exames de fim de curso, chamados de “exames preparatórios”, houve a determinação por lei de que o estudante que pretendesse ter acesso ao ensino superior deveria submeter-se a um exame. Essa determinação marca a origem do vestibular, tendo sido consolidado na década de 20, quando o número de candidatos ultrapassou o número de vagas oferecidas (Cunha, 1980). Em decorrência de a avaliação ter sido associada, durante todo esse período, até a década de 60, ao exame de seleção para entrada no curso superior, ela foi concebida simplesmente como medida do conhecimento. Pois, só os alunos que atingiam as melhores notas garantiam a classificação para o ingresso em tal instituição, já que as vagas eram limitadas. Mudanças só aconteceram em 1971, quando foram regulamentadas por lei e por várias portarias ministeriais, as condições do exame vestibular.

O país passou por um período de extrema falta de domínio adequado no campo da avaliação sem registrar nenhum avanço. *Poucos se dedicaram aos estudos sobre avaliação educacional, que praticamente desapareceu como área de estudo e formação em nossas*

universidades e cursos superiores, e a produção científica relevante sobre a questão foi pequena e esparsa por um longo período (Gatti, 2002).

O uso excessivo de testes objetivos associados aos vestibulares e aos materiais didáticos produzidos à época, contribuiu para um processo de crítica intensa e *desabonamento social e acadêmico. Essa realidade gera a desqualificação dos processos avaliativos, que se estende à própria avaliação educacional, descaracterizando o seu verdadeiro objetivo como área de estudos e pesquisa.* (Gatti, 2002).

Em relação à avaliação do rendimento escolar, há relatos de que na década de 70, várias pesquisas foram empreendidas no Brasil. Segundo Gomes Neto e Rosemberg (1995) essas pesquisas foram realizadas de forma dispersa — embora a maior parte se concentrasse nos Estados do Rio de Janeiro, Paraná, Rio Grande do Sul e São Paulo — não tendo havido sincronia e tendo seus resultados alcançado pouca divulgação. Ainda, segundo esses autores, no início da década de 80, o Ministério da Educação começou a incentivar a produção de estudos sobre avaliação do rendimento escolar, estimulado, principalmente, pelas discussões com as agências internacionais. A cooperação voltada para estimular a "igualdade de chances educacionais" retoma as linhas de ação anteriores, centradas na racionalização dos sistemas de ensino, na melhor distribuição dos insumos escolares e na seletividade do ensino: incentivadas pelo Banco Mundial, que é um organismo que ajuda governos em países em desenvolvimento a reduzir a pobreza por meio de empréstimos e experiência técnica para projetos em diversas áreas, entre elas, a da educação.

No que se refere à avaliação, o Banco sugere o estabelecimento de critérios gerenciais e de eficiência, por meio dos quais se busca alcançar a qualidade da educação. Embora não tenha sido possível realizar, no âmbito deste estudo, uma análise mais detalhada sobre as políticas educacionais brasileiras, é interessante notar como os documentos estratégicos do Ministério da Educação e Cultura (MEC) para o período 95/96 indicam a importância central dessas avaliações, com o incentivo, inclusive, do fortalecimento de instituições de avaliação fora do âmbito do MEC, visando atingir padrões internacionais (Brasil/MEC, *Planejamento político-estratégico*, 1995).

A partir da década de 80, além de algumas pesquisas de avaliação do rendimento escolar que timidamente se desenvolvem, fala-se na criação de uma nova 'cultura de avaliação', mas a falta de pessoas especializadas na área continua sendo um problema, apesar dos esforços de formação de algumas equipes. *Para minimizar o problema, formam-se grupos que, a partir da pequena massa crítica existente, desenvolvem, ao mesmo tempo, os processos*

avaliativos, no caso avaliação de sistemas (Gatti, 2002). Algumas secretarias estaduais de educação estabeleceram um sistema para avaliar a aprendizagem de seus alunos. Gomes Neto e Rosemberg (1995) citam em seu artigo “*Indicadores de qualidade do ensino e seu papel no sistema nacional de avaliação*”, pp. 16-19, alguns exemplos dessas iniciativas, os quais constituem a origem da consolidação do sistema de avaliação do País. Essas primeiras avaliações com características sistêmicas, são citadas a seguir.

Avaliação das Escolas Rurais do Nordeste: Edurural

Fez parte do primeiro Projeto de Educação Básica para o Nordeste, assinado em 1980, tendo a avaliação se estendido de 1981 a 1987. O objetivo dessa avaliação foi determinar os fatores que influenciavam o rendimento escolar e estudar o custo-eficácia dos insumos educacionais. Teve por característica ser uma avaliação de larga escala realizada em escolas rurais do Nordeste, da qual participaram aproximadamente 600 escolas, com a participação de 1.000 professores e 6.000 alunos e suas famílias. Os alunos foram submetidos a testes de Português e Matemática em 1981, 1983 e 1985.

Foi planejado um estudo longitudinal que previa obter dados de uma subamostra desses alunos. Em 1981, participou uma amostra de escolas da zona rural de três Estados: Piauí, Ceará e Pernambuco, das quais foram selecionadas as escolas que comporiam a subamostra. Os testes foram aplicados aos alunos de 2ª e 4ª séries a cada dois anos. Com isso, esperava-se que muitos alunos de 2ª série fossem reavaliados na 4ª série, dois anos depois. Entretanto, devido a fatores como o fechamento de muitas escolas, os altos índices de repetência e a evasão, conjugados com a migração da zona rural, poucos foram os alunos avaliados mais de uma vez: dos 4.000 alunos avaliados na 2ª série em 1981 e 1983, apenas cerca de 300 foram encontrados na 4ª série em 1983 e 1985, respectivamente. A pesquisa foi gerenciada pela Fundação Cearense de Pesquisa e Cultura (FCPC), responsável pela aplicação dos testes e entrevistas, com assistência técnica da Fundação Carlos Chagas (FCC), que elaborou e analisou os testes e ainda orientou tecnicamente a equipe de professores da Universidade Federal do Ceará envolvida na pesquisa. Essa pesquisa teve assistência técnica e financiamento do Banco Mundial.

Apesar de todo esse esforço, nenhuma ação pedagógica resultou da avaliação. Os relatórios foram encaminhados ao MEC que, na prática, os ignorou. Nenhum resultado retornou às escolas. Assim, alunos, pais, professores, diretores de escolas, além de vários técnicos das Secretarias Estaduais e dos órgãos municipais de educação ficaram sem nenhuma

informação sobre os resultados da avaliação a qual eles se submeteram ou da qual participaram.

Embora no Brasil a discussão dessa pesquisa tenha ficado restrita aos pesquisadores, fora do País seus resultados foram bastante difundidos, por se tratar de uma das primeiras avaliações escolares de larga escala aplicadas em zonas rurais de países em desenvolvimento. A análise dos efeitos das principais ações sobre o rendimento escolar dos alunos lançou novas luzes sobre a eficiência e eficácia dessas ações, com impactos sobre o desenho de projetos posteriormente financiados pelo Banco Mundial.

São Paulo: Projeto Inovações no Ensino Básico (IEB) – Projeto Escola-Padrão

Este projeto foi desenvolvido pela Secretaria de Educação do Estado de São Paulo em 1989, com financiamento do Banco Mundial. Foram planejadas ações na área de construção e reforma de prédios escolares, apoio à gestão escolar, capacitação de recursos humanos, aquisição e distribuição de livros didáticos e material escolar, desenvolvimento de projetos municipais de implantação de pré-escolas, etc. Para sua avaliação, foi contratada a Fundação Carlos Chagas, que pôde aproveitar a experiência adquirida no Edurural e corrigir alguns dos erros cometidos naquela avaliação.

A primeira avaliação foi realizada com uma amostra de escolas e, em cada uma delas, foram sorteadas duas classes do Ciclo Básico de Alfabetização Inicial (correspondente à 1ª série). Em 1992, todos os alunos dessas classes foram avaliados. Esses mesmos alunos foram avaliados novamente nos dois anos subseqüentes, independentemente da série ou da classe em que estivessem. Com esse procedimento foi possível realizar um estudo longitudinal, pois cerca de 2.000 alunos foram avaliados no mínimo duas vezes.

A Fundação Carlos Chagas foi encarregada de elaborar, aplicar e analisar os dados. O objetivo principal dessa avaliação foi realizar uma análise da eficácia das políticas adotadas pelo projeto, semelhante à avaliação Edurural.

Ao mesmo tempo, a Fundação Instituto de Pesquisas Econômicas (FIPE) da Universidade de São Paulo foi contratada para calcular o custo-eficácia dos diversos insumos analisados pela FCC. E isso completava o ciclo da pesquisa.

Minas Gerais

O procedimento de avaliação em Minas Gerais teve início em 1991 e foi conduzido pela Secretaria de Educação, como parte do plano de governo, que tinha como meta a melhoria da qualidade do ensino. O Banco Mundial apoiou a iniciativa e, no acordo assinado com o Estado, no início do referido ano, esse componente foi contemplado, garantindo a sua continuidade.

Para o estabelecimento desse sistema, foi criado na Secretaria um grupo de avaliação responsável por coordenar os trabalhos. A avaliação ocorreu ao final das seguintes séries: Ciclo Básico de Alfabetização (CBA), 5ª e 8ª séries do Ensino Fundamental, 2ª série do Ensino Médio e Magistério.

Todos os alunos, diretores de escolas, professores e pais de alunos das séries escolhidas, foram avaliados. Os resultados foram apresentados, discutidos e comparados com aqueles obtidos por outras escolas do mesmo município e do estado em várias reuniões com efetiva participação de todos os setores. Essa discussão crítica serviu para apoiar as ações da escola e da Secretaria no sentido de superar os problemas apontados pela avaliação. Esse modelo adotado de avaliação passou por melhorias relacionadas aos aspectos metodológicos, mas se mantém ativo e com as mesmas características até hoje.

Ceará

Em 1992, as Secretarias de Educação do Estado do Ceará e do Município de Fortaleza realizaram a avaliação de todos os alunos de 4ª e 8ª séries da rede pública de Fortaleza nas disciplinas Português e Matemática.

Nos dois anos seguintes, a Secretaria Estadual contratou novamente a mesma equipe para expandir essa avaliação a 13 cidades do interior. Nestas duas últimas avaliações, além da aplicação dos testes de aprendizagem, foram calculadas, para as escolas pesquisadas, taxas de sucesso ou aprovação no ensino fundamental, além de índices de ocupação e de manutenção. Assim, outras dimensões das escolas foram incluídas na avaliação.

As avaliações foram realizadas sob a coordenação do Parque de Desenvolvimento Científico e Tecnológico (CETREDE) e contaram com a supervisão da Secretaria de Educação do Estado, através do grupo de avaliação especialmente constituído. Na primeira avaliação realizada, o grupo contou ainda com o apoio do Instituto Nacional de Estudos e Pesquisas em Educação (INEP), que coordena o SAEB.

O Cetrede a partir do ano 1994 foi o responsável pelo processamento dos dados, da análise, elaboração dos testes e dos relatórios, sob a supervisão da Secretaria de Educação. Os relatórios incluíam uma análise geral dos resultados, bem como uma apresentação dos resultados alcançados em cada escola. Os relatórios traziam também as médias por disciplina e por série, o que permitia a comparação de cada escola com as respectivas médias estaduais. Os resultados eram apresentados para dirigentes e técnicos da própria Secretaria e para os diretores das escolas.

Como consequência dessas avaliações, a equipe pedagógica da Secretaria analisava as questões respondidas pelos alunos e, por meio dessa análise, detectava as principais dificuldades encontradas por eles. Essas informações eram utilizadas para redirecionar a proposta de capacitação dos docentes da rede escolar, de forma a incluir reforço em conteúdos e metodologias relacionados a essas dificuldades.

O Sistema Nacional de Avaliação da Educação Básica (SAEB)

A partir de 1988 iniciou-se no Brasil a implantação do SAEB. Essa proposta surgiu da discussão sobre a democratização do ensino durante o período de redemocratização do país, no início da Nova República, em 1985/1986. O contexto naquele momento foi marcado pelo fim do Regime Militar, pela eleição indireta de Tancredo Neves, seu falecimento e pela posse de José Sarney. Dessa forma, houve um estímulo para que se discutissem questões sobre educação de uma forma democrática e aberta. A discussão sobre essas questões educacionais já havia perdido o seu sentido pedagógico e assumido um caráter político. Algumas medidas já tinham sido implementadas em função do acesso ao ensino, mas poucas mudanças eram percebidas. Dessa forma, sentiu-se a necessidade de avaliar a qualidade do sistema educacional brasileiro para constatar quais mudanças de fato estavam ocorrendo. O objetivo principal dessa primeira proposta foi levantar informações úteis sobre o que estava sendo gerado no setor educacional, como, onde, quando e quem eram os responsáveis pelo produto obtido. Assim, surgiu o Sistema Nacional de Avaliação da Educação Básica (SAEB). O SAEB, inicialmente, foi uma atribuição exclusiva do MEC e, a partir de 1995, passou a ser coordenado pelo INEP (Gomes Neto & Rosemberg (1995); Castro, 1998, Franco, 2001; Barreto, 2001; Rodrigues, 2002; Locatelli, 2002).

A proposta do SAEB foi fundamentada no Sistema de Avaliação do Ensino Público de 1º Grau (SAEP), que foi implementado na forma de programa e estruturado para avaliar o Programa de Educação Básica para o Nordeste, o qual visava a responder às demandas do

Banco Mundial no sentido de desenvolver uma sistemática de avaliação para o segmento da educação. Essa sistemática foi estendida para o resto do País em agosto de 1988, com os ajustes necessários. Para o gerenciamento do sistema, foi constituída uma equipe na antiga Secretaria Nacional de Ensino Básico (SENEB), e para a elaboração dos testes foi novamente contratada a Fundação Carlos Chagas. A aplicação piloto se deu no mesmo ano em dois Estados: Rio Grande do Norte e Paraná. Foram aplicados testes em alunos de 1^a, 3^a, 5^a e 7^a séries, matriculados em escolas públicas desses estados. As 1^a e 3^a séries foram avaliadas em Português e Matemática, enquanto as outras, além desses testes, foi aplicado um teste de Ciências e solicitado que os alunos fizessem uma redação (Gomes Neto e Rosemberg, 1995; Gatti, 2002, Freitas, 2004).

Após a reformulação dos instrumentos (testes e questionários de gestão, do professor e de custo-aluno), a SENEB assinou um convênio com o Programa das Nações Unidas para o Desenvolvimento (PNUD), em que alocou recursos para a aplicação do SAEB. Em 1990 foi realizada a primeira edição do SAEB. O MEC coordenou todo o processo, mas as aplicações dos instrumentos (testes e questionários) ficaram a cargo das secretarias estaduais de educação. As equipes estaduais, responsáveis pela aplicação, processamento e correção dos testes aplicados, receberam treinamento específico. Os resultados, porém, só foram disponibilizados em 1993.

O segundo ciclo do SAEB, que estava previsto para 1992, somente foi realizado em 1993. Os procedimentos foram quase os mesmos, ou seja, a coordenação ficou com o Ministério da Educação e a aplicação e correção, com as equipes estaduais. Os instrumentos dessa nova avaliação foram modificados e os alunos não foram avaliados em redação. Novamente, os resultados da avaliação só foram publicados dois anos depois.

No período da segunda edição do SAEB, alguns diretores e pesquisadores influentes do MEC/INEP tinham tomado conhecimento das novas tecnologias de análise, que incluíam a Teoria de Resposta ao Item (TRI) as quais vinham alcançando grande sucesso na área de avaliação educacional. Nessa ocasião, foi feito um convite, por parte do INEP, ao Professor Luiz Pasquali, do Instituto de Psicologia, da Universidade de Brasília, coordenador do primeiro Laboratório de Pesquisa em Avaliação e Medidas (LabPAM) surgido na América do Sul. O Prof. Pasquali acabara de participar de uma Conferência Internacional em Oxford (Inglaterra, de 27-30 jun, 1993), promovida pela *International Test Commission* que tratava, entre outras coisas, dos avanços nos modelos, métodos e práticas de medida. Nessa Conferência, encontravam-se representantes de 46 países e os maiores expoentes na pesquisa

com a Teoria de Resposta ao Item (TRI), como Frank B. Baker, Ronald K. Hambleton, José Muñiz Fernandez, W. J. van der Linden, entre outros. O convite foi para que o Prof. Pasquali procedesse à análise dos dados da avaliação realizada em 1990, utilizando a TRI. O software utilizado foi o ASCAL (*Assessment Systems Corporation*, 1989), cedido pelo Professor José Muñiz, da Universidade de Oviedo, Espanha.

Os resultados gerados dessa análise e apresentados em relatórios não alcançaram o impacto esperado. Havia poucas pessoas que conheciam essa teoria e não havia diálogo entre esses pesquisadores. Além da falta de interlocução, foram encontradas outras dificuldades como a falta de computadores potentes o suficiente para dar conta das análises. Desse esforço inicial, em 1994, o Professor Luiz Pasquali publicou artigo sobre a TRI e que se tornou um capítulo (cap. 7) do livro “Teoria e Métodos de Medida em Ciências do Comportamento (1996)”, organizado por ele e publicado pelo INEP.

No Brasil havia um grupo que vinha se destacando no empreendimento de uma avaliação que utilizasse a Teoria de Resposta ao Item. Este grupo, ligado à Fundação Cesgranrio, era composto pelos pesquisadores: Philip Fletcher, que em 1994 publicou o artigo “A Teoria de Resposta ao Item: medidas invariantes do desempenho escolar”; Ruben Klein, que realizou as análises da primeira avaliação nacional que utilizou a TRI como metodologia (SAEB/95); e Sérgio Costa Ribeiro, que neste período, junto com Ruben Klein, desmistificou o problema do fluxo relacionado à repetência escolar (Klein, R. & Ribeiro, S. C, 1991).

Em 1995 foi solicitado que a empresa que realizasse a avaliação da edição relativa a esse ano do SAEB, deveria introduzir análises fundamentadas na TRI. A Fundação Cesgranrio, em consórcio com a Fundação Carlos Chagas, que já vinha realizando outras avaliações sem, entretanto, utilizar essa nova tecnologia, realizaram essa avaliação que se tornou um marco para o sistema brasileiro. A Fundação Cesgranrio, nessa época estava desenvolvendo um projeto, com o apoio da Fundação Ford, de uma pesquisa de avaliação de alunos de 8ª série e 3ª série do 2º grau, do município do Rio de Janeiro, em 1995. Foram selecionadas duas áreas curriculares - Matemática e Língua Portuguesa (leitura) e duas séries finais de ciclo - 8ª do 1º grau e 3ª série do 2º grau - para serem avaliadas. Essa avaliação seguiu a mesma metodologia e os mesmos moldes das avaliações subsequentes do SAEB (Klein R. & Fontanive, 1995).

Com a adoção dessa nova metodologia, foi definida como finalidade primeira do SAEB obter resultados sobre a qualidade do ensino ao longo do tempo, identificar os fatores que contribuem para a ocorrência desses resultados e fornecer diagnósticos para políticas públicas de intervenção, visando à melhoria da qualidade da educação básica do Brasil. Uma

segunda finalidade definida foi acompanhar o monitoramento do avanço alcançado pelos programas e políticas governamentais em relação às metas educacionais.

Para atingir esses objetivos, outros instrumentos foram então desenvolvidos e empregados, além dos testes de avaliação do desempenho escolar. Adicionalmente, são utilizados questionários que permitem: (1) obter informações sobre as características da realidade socioeconômica e cultural e hábitos de estudo dos alunos; (2) avaliar o perfil e as práticas pedagógicas dos professores; (3) avaliar o perfil e as práticas de gestão escolar dos diretores e, (4) realizar o levantamento dos equipamentos disponíveis e características físicas e de conservação das escolas (Pestana, 1998).

É importante considerar que, segundo consta nos objetivos do SAEB, decisões sobre as políticas educacionais brasileiras são embasadas nas informações estatísticas e avaliativas da educação fornecidas, principalmente, por esse sistema. Ele se constitui no principal gerador de indicadores para o sistema de informações da qualidade da educação brasileira, devendo, portanto, primar pelos instrumentos utilizados para coletar essas informações, procurando demonstrar e comprovar sua validade, fidedignidade e objetividade.

O SAEB toma como referência, desde 1995, o sistema norte-americano de avaliação educacional – NAEP – *National Assessment of Educational Progress*, que também é o responsável pela avaliação da educação básica daquele país. Da mesma forma que o SAEB, ele é um sistema conduzido periodicamente, que fornece informações sobre o que os estudantes americanos sabem e podem fazer nas várias disciplinas preconizadas para o currículo comum. Observa-se, que o NAEP se pauta principalmente pelas informações estatísticas geradas dos dados coletados e na disseminação de seus resultados. São produzidos relatórios técnicos, narrativos e explicativos de todo o sistema avaliativo.

Nesse contexto, o SAEB preconiza que se empenha em fornecer subsídios à formulação de políticas e diretrizes adequadas à diversidade de situações presentes nos estados e regiões brasileiras, e se constitui uma referência nacional no que diz respeito ao desempenho escolar da população discente.

Pode-se afirmar que após a implementação do SAEB, a cultura avaliativa da educação, tendo por meta a busca de melhoria do sistema educacional básico se consolidou no Brasil. Entre os sistemas estaduais instalados, além dos citados anteriormente (Ceará, São Paulo, Minas Gerais) podem-se citar o Sistema de Avaliação Educacional de Pernambuco – SAEP, o Programa Nova Escola do Rio de Janeiro, Programa de Avaliação – AVA, do

Paraná, Programa “Educar para vencer”, da Bahia, e inúmeros outros. Hoje, no Brasil, há dezenove programas estaduais de avaliação em ação, e a maioria segue o modelo do SAEB.

Além desses programas implantados, o SAEB contribuiu para a formação e aperfeiçoamento de inúmeros pesquisadores e técnicos em avaliação educacional. Com a determinação de formar recursos humanos e realizar estudos especializados, a CAPES implementou no Brasil o Programa de Apoio à Avaliação Educacional, chamado PROAV, que por meio de proposta licitatória, selecionou cinco universidades brasileiras com essa função: a Universidade de Brasília, por meio do Instituto de Psicologia; a Universidade Federal de Juiz de Fora, por meio da Faculdade de Educação, a Universidade Federal de Minas Gerais, por meio do Instituto de Estatística, a Universidade Federal do Rio de Janeiro, por meio da Faculdade de Educação e a Pontifícia Universidade Católica do Rio de Janeiro, pela Faculdade de Educação. Por meio deste Programa, as referidas instituições geraram muitos estudos, dissertações de mestrados e teses de doutorado.

Além dessa expansão de estudos, surgiram várias instituições públicas e privadas que se consolidaram na especialização de avaliações educacionais em larga escala. Até 1999, as Fundações Cesgranrio e Carlos Chagas eram as únicas que detinham conhecimento e cacife para concorrer a processos licitatórios. Entre as instituições públicas que hoje se fazem presentes, podem se citar o Centro de Políticas Públicas e Avaliação da Educação (CAEd/UFJF); Centro de Pesquisa em Avaliação Educacional (CPAE/IP/UnB); o Centro de Seleção e de Seleção de Eventos (CESPE/UnB); Grupo de Avaliação e Medidas Educacionais (GAME/UFMG); Centro de Estudos em Avaliação Educacional IE/UFRJ; o Laboratório de Avaliação da Educação da PUC-Rio. Entre algumas instituições privadas, podem-se citar: o Instituto Movens; Núcleo de Avaliação Educacional e Psicológica (NAEP); AVALIA Assessoria educacional; Abaquer Consultores e Editores Associados, entre outras. Hoje, conta-se também com a Associação Brasileira de Avaliação Educacional (ABAVE), instituição criada com a finalidade de congregar pesquisadores e interessados em promover a melhoria da qualidade da educação brasileira.

Constata-se que a trajetória da avaliação educacional no Brasil é muito recente e encontra-se em franco desenvolvimento e amadurecimento. Nos capítulos seguintes serão tomados com mais profundidade os principais elementos que compõem esse rico sistema de avaliação.

2. TESTES DE DESEMPENHO:

CONSTITUIÇÃO, ANÁLISES E DESCRIÇÃO DOS RESULTADOS

2.1 CONSTITUIÇÃO DE UM TESTE PARA AVALIAÇÃO EDUCACIONAL

Testes de inteligência, de personalidade, inventários vocacionais e de interesse, exames de vestibular, provas de sala de aula e avaliações de desempenho em larga escala, como é o caso desse estudo, requerem o desempenho de algum tipo de comportamento – uma ação observável e mensurável que mede algum atributo, traço ou característica. Alguns testes podem ser usados para fazer uma predição acerca de algum resultado, como por exemplo, a estimativa do desempenho educacional (McIntire e Miller, 2000).

Para Anastasi (2000), um teste é *“essencialmente uma medida objetiva e padronizada de uma amostra de comportamentos”*. Para Cronbach (1996), *“é um procedimento sistemático para observar o comportamento e descrevê-lo com a ajuda de escalas numéricas ou categorias fixas”*. Pasquali (2001), afirma que *“é um conjunto de tarefas predefinidas, as quais o sujeito precisa executar numa situação geralmente artificializada ou sistematizada e seu comportamento, nessa situação, será observado, descrito e julgado, sendo que a descrição do mesmo é geralmente feita utilizando-se números”*. Tyler (1981) define teste como *“uma situação padronizada cujo intento é refletir a amostra de um determinado aspecto do comportamento individual”*. Com base nessas conceituações, percebemos que, em comum, todos os estudiosos apresentam o teste como uma forma padronizada ou sistemática de descrever objetivamente o comportamento.

Os itens que compõem o teste constituem "a representação comportamental do traço latente, eles são as tarefas, as ações empíricas por meio das quais o traço latente se manifesta" (Pasquali, 1997). A forma de o comportamento representar esses traços é por meio da resposta dada ao conteúdo representado no item. Pode-se, então, estabelecer, que a partir de uma amostra de comportamentos, há uma correspondência empírica entre o desempenho do examinando no teste e o comportamento predito. Assim, é importante observar que a mensuração em si não deve ser o objetivo da testagem, mas a correspondência entre a observação coletada e o objetivo da aplicação do teste.

Segundo McIntire e Miller (2000), a avaliação que adota essa perspectiva pressupõe três características em comum: (a) avalia-se uma amostra representativa de comportamentos que medem os conhecimentos alcançados ou predizem um resultado, ou seja, o teste deve envolver o conhecimento dos diferentes tipos de habilidades requeridas para o bom desempenho de um determinado comportamento; (b) as amostras do comportamento dos examinandos devem ser obtidas sob condições padronizadas, o que implica que o teste deva ser aplicado da mesma maneira, sob as mesmas condições, seguindo os mesmos padrões, para todos os examinandos e, (c) devem existir regras para a interpretação dos resultados, pois, elas são necessárias para assegurar que qualquer examinador que pretenda avaliar um mesmo conjunto de respostas que outro avaliador, possa proceder à interpretação da mesma forma.

A realização de um teste implica aplicar um instrumento ou procedimento para determinar a quantidade e a qualidade de uma variável possuída por um sujeito, devendo, portanto, obedecer a certos pressupostos. A **validade** é o pressuposto mais importante na análise de testes. Ela se refere à adequação, significação e utilização prática das inferências feitas a partir dos escores de testes. A validação de um teste é indicada pelo processo de acumulação de evidências que dão apoio a tais inferências (*Standards for Educational and Psychological Testing, 1985*). Dessa forma, validade é um conceito unitário, ou seja, o que temos são várias formas de evidenciá-la, incluindo análises de cunho qualitativo e quantitativo (Pasquali, 2006).

Segundo os *Standards for Educational and Psychological Testing* (1985), os vários meios de acumular evidências de validade estão agrupados em categorias ou fontes de evidências, a saber, as mais comuns são: (1) validade relacionada ao conteúdo; (2) validade relacionada ao critério; (3) validade relacionada ao construto. Essas categorias ou fontes de evidências podem ainda se subdividir em categorias mais refinadas não implicando que cada uma delas ou suas subdivisões sejam tipos distintos de validade. Elas apenas evidenciam

validade. É importante ressaltar que não é possível se fazer distinções precisas entre elas, pois elas se relacionam, como por exemplo, ao citarmos a validade de construto, estaremos relacionando-a a outra categoria como a validade de conteúdo. Quanto mais fontes de evidências melhor. Entretanto, a qualidade da evidência é de fundamental importância. Uma única fonte é preferível a várias linhas de evidência de qualidade duvidosa. Grande empenho deve ser investido para se alcançar a combinação de vários tipos de evidências que reflitam o valor de um teste para um propósito determinado.

Em testes educacionais, o primeiro passo para assegurar a validade é por meio da verificação da validade de conteúdo que pode ser entendida como um dos meios de investigar se o conteúdo relacionado ao teste aplicado corresponde àquilo que foi preconizado no período de escolarização considerado na avaliação; pode-se também utilizar a validade de critério que se baseia na escolha de critérios que assegurem que o teste seja válido para a população avaliada; e a validade de construto que se apóia na avaliação dos conceitos teóricos subjacentes ao desempenho.

Outro pressuposto importante é a **fidedignidade** da medida do comportamento do indivíduo evidenciado no teste. Esta é uma das propriedades de um conjunto de escores de um teste: a que descreve quão consistente é um teste, ou quão livre de erros suas medidas estão (Frisbie, 1988). Algum erro está sempre envolvido em qualquer tipo de medida (Nunnally & Bernstein, 1994). No entanto, espera-se que os escores obtidos da medida de um comportamento sejam consistentes ao longo do tempo, ou em conjuntos de itens equivalentes, ou sob outras condições variáveis de exame. Assim, pode-se dizer que um teste apresenta um grau X de fidedignidade à medida que se avalia a extensão em que as diferenças individuais dos escores no teste são atribuíveis a diferenças verdadeiras e a extensão em que elas são atribuíveis a erros casuais (Anastasi & Urbina, 2000). A questão central, portanto, está na definição da variância de erro. Há inúmeras formas de calcular a precisão de um teste (Nunnally & Bernstein, 1994).

É comum ela ser avaliada por quatro meios distintos: pela estabilidade temporal, por conjuntos de itens equivalentes, pela consistência interna dos itens e pelas diferenças entre avaliadores. Entre os métodos relacionados à estabilidade temporal pode-se citar o *teste-reteste*. Por meio de conjuntos equivalentes, pode-se citar o método das *formas alternadas*. Entre os que indicam a consistência interna dos itens podem-se citar os métodos de correlação entre as *duas metades dos testes*, o *coeficiente alfa de Cronbach* e seus casos particulares [*Kuder-Richardson (KR20 e 21)*, *Rulon*, *Flanagan e Guttman*]. O método do avaliador indica

a variância entre os escores atribuídos pelos diferentes avaliadores (Anastasi & Urbina, 2000; Crocker & Algina, 1986; Nunnally & Bernstein, 1994; Pasquali, 1997, 2003).

Outro pressuposto a ser observado é a **objetividade** do teste que, para o caso de testes de aptidão ou cognitivos, é a garantia de que os itens propostos irão cobrir objetivamente comportamentos passíveis de avaliação, permitindo tão-somente uma resposta certa ou errada (Pasquali, 1999). Assim, o respondente deve, por meio do item que expressa um comportamento, executar objetivamente a tarefa proposta, ou seja, demonstrar se conhece ou não a solução.

A verificação desses pressupostos é importante porque permite que se saiba se estamos trilhando o caminho certo e se, de fato, estamos medindo o que tencionamos medir. As qualidades de fidedignidade, de validade e de objetividade são essenciais para se avaliar um instrumento porque estas compõem a primeira linha de defesa adotada contra a formação de conclusões inacuradas.

A elaboração de um instrumento, que tem por finalidade avaliar o desempenho de alunos, prescinde do conhecimento de tudo o que perpassa o sistema educacional. Para alcançar efetividade, ele deverá corresponder ao alcance das seguintes demandas:

- 1) explicitar e tornar públicas as expectativas de aprendizagem;
- 2) ajustar critérios apropriados e padrões altos para a avaliação da qualidade da aprendizagem;
- 3) reunir, analisar e interpretar evidências, sistematicamente, para determinar quão bem o desempenho nivela-se às expectativas de aprendizagem;
- 4) usar as informações resultantes para documentar, explicar e melhorar o desempenho.

Segundo Angelo (1995), quando a avaliação é delineada eficazmente, ela chama a atenção dos agentes que promovem mudanças, leva à reflexão dos pressupostos que a embasaram e cria uma cultura acadêmica compartilhada dedicada a assegurar e melhorar a qualidade da educação.

A elaboração de um teste para avaliação do desempenho de alunos deve estar fundamentada no conhecimento dos propósitos do sistema de ensino e suas práticas, na compreensão do processo de ensino e aprendizagem dos estudantes e deve refletir a intenção da avaliação, devendo ser elaborado de tal modo que permita dar um retorno a todos os

agentes envolvidos no processo (Linn, 2001). Do mesmo modo, os critérios precisam ser compreensíveis, explícitos e públicos, pois assim, todos os envolvidos no processo, desde os próprios estudantes até os pais e todos os que compõem a equipe escolar saberão das intenções da avaliação. E, da mesma forma, deve-se oferecer um retorno sobre o processo, mesmo que este não traga resultados pontuais referentes aos alunos avaliados ou à escola. Esse retorno provoca reflexões que por si só já se constitui um passo dado na busca da melhoria da qualidade do processo de ensino.

Anterior, ainda, ao processo de construção do instrumento, é importante definir o tipo de análise a ser utilizada na pós-testagem. Cada análise requer o cumprimento de alguns pressupostos, sem os quais, os resultados poderão estar comprometidos.

Se as técnicas a serem utilizadas nas análises se relacionam à Teoria Clássica dos Testes (TCT), a composição de itens para testes de desempenho deve ser feita com base em uma seleção de itens realizada de acordo com seu conteúdo, com as características de dificuldade e com a discriminação dos itens. Itens com alto poder de discriminação e com o nível apropriado de dificuldade são geralmente os mais desejáveis para o propósito de qualquer teste e pela distribuição antecipada da habilidade do grupo inteiro, dada pelo escore total, a quem o teste é direcionado (Nunnally e Bernstein, 1994).

Índices clássicos não são invariantes sobre populações que diferem na habilidade. Portanto, o sucesso das técnicas clássicas de seleção dos itens depende de quão proximamente o grupo utilizado para determinar os índices do item se emparelham com a população como um todo para a qual o teste é proposto. Quando esse emparelhamento é pobre, os índices obtidos do item não serão apropriados para esta população. Em algumas situações práticas, o grupo para o qual os índices do item são obtidos e o grupo para o qual o teste é direcionado são muito diferentes (Hambleton e Jones, 1993).

Para desenvolver um banco de itens, as características dos itens armazenados devem ser determinadas. Na prática, esses itens, freqüentemente chamados itens "experimentais", são administrados na forma de testes em grupos e, assim, seus índices são obtidos. Múltiplas formas do teste são criadas, cada uma contendo itens experimentais variados e formas diferentes são administradas para grupos diversos de examinandos, por meio de delineamentos amostrais de itens chamados blocos incompletos balanceados (BIB), ou parcialmente balanceados (BIPB). Geralmente não é possível assegurar que essas diferentes formas sejam administradas para grupos equivalentes; portanto, os índices dos itens

experimentais que foram aplicados para grupos diferentes de examinandos não poderão ser comparáveis (Hambleton, Swaminathan & Rogers, 1991).

Os referidos autores alertam ainda que além do problema de índices não-invariantes, o principal problema dos procedimentos clássicos para a construção de testes é que, mesmo quando um banco de itens bem elaborados está disponível, eles não poderão ser selecionados para produzir um teste que coloque uma especificação fixa em termos de precisão da medida. A contribuição de um item para a fidedignidade de um teste não depende das características de um item só, mas também da relação entre um item e os outros itens do teste. Assim, não é possível isolar a contribuição de um item para a fidedignidade e, portanto, para o erro padrão de medida de um teste.

A Teoria de Resposta ao Item (TRI) oferece um método mais poderoso de seleção de itens que a Teoria Clássica de Testes. Obedecendo-se aos pressupostos básicos para análise dos itens que utiliza essa teoria, os parâmetros dos itens devem apresentar-se invariantes, sobrepondo aos problemas dos índices clássicos descritos anteriormente. Além disso, a dificuldade do item e a habilidade do examinando são medidas na mesma métrica, tornando possível selecionar itens que sejam mais úteis em certas regiões da escala de habilidade. Por vezes é citada como a maior vantagem da TRI, o fato de que ela permite a seleção de itens baseada na quantidade de informação que os itens possuem na contribuição para a quantidade total de informação necessária no teste para reunir as especificações de um teste. Desde que a informação seja relacionada à precisão da medida, é possível escolher itens para produzir um teste que tenha a medida de precisão desejada em qualquer nível de habilidade, por exemplo, em um ponto de corte do escore (Hambleton, Swaminathan & Rogers, 1991).

Adotando-se como método de seleção a TCT ou a TRI, a construção dos itens constitui-se um importante indicador da qualidade dos itens. Partindo-se do pressuposto de que a avaliação deva ter um caráter orientador e integral e deva ser realizada em um processo contínuo, de maneira sistemática e planejada, e que as qualidades de fidedignidade, de validade e de objetividade tenham sido asseguradas, a construção de testes educacionais exige que o construtor de itens domine a tecnologia da construção de testes.

A elaboração do instrumento deve seguir algumas orientações técnicas (Haladyna, 1999; Hopkins, Stanley & Hopkins, 1990; Crocker & Algina, 1986; Guia de elaboração de itens do SAEB, 2004; Manual de elaboração de itens do Programa Nova Escola, RJ, 2006). Na análise do item o primeiro ponto a ser observado na busca da garantia de sua qualidade é o cumprimento das orientações técnicas para sua construção.

A garantia de uma avaliação eficaz passa pelo planejamento e conhecimento dos tipos de instrumentos disponíveis, das diferentes naturezas dos dados, dos processos usados para selecionar instrumentos apropriados e das habilidades dos instrumentos para medir, avaliar e descrever os resultados observáveis de aprendizagem.

2.2 ANÁLISES DOS TESTES DE DESEMPENHO

2.2.1 ANÁLISE PEDAGÓGICA DOS ITENS

Os itens devem ser analisados com rigor em seus dois campos de análises possíveis: o pedagógico e o psicométrico. Um dos campos não deve ser mais privilegiado do que o outro. Ao contrário, eles devem se complementar (Rodrigues, 2002). A análise pedagógica é realizada considerando-se os aspectos de conteúdo e forma dos itens. Ela é realizada por meio da avaliação da validade de conteúdo e dos procedimentos efetivos da elaboração dos itens. Ambas as análises, pedagógica e psicométrica, visam avaliar a validade dos itens e dos testes.

A análise pedagógica de um **teste** implica observar: (1) a construção do item; (2) a distribuição e a porcentagem de conteúdos abrangidos em relação a todo o conteúdo avaliado; (3) o nível de dificuldade dos itens que compõem os temas abrangidos; (4) a distribuição das competências exigidas para a resolução dos itens do teste; (5) o levantamento da distribuição das alternativas corretas (A, B, C, D, ou E) nos blocos que compõem os testes.

A análise pedagógica de um **item** pressupõe revelar o seu processo de resolução, com base na qualidade da forma abordada do conteúdo, explicando exatamente o que se pode dizer sobre o conhecimento do aluno. Essa análise é fundamental, pois será com base nela que se poderá dizer o que os alunos sabem ou não (Rodrigues, 2002, 2006). Um item de múltipla escolha de um teste é constituído de um texto, se for um teste de língua portuguesa, um enunciado e das alternativas. O processo de análise pedagógica do item deve ser realizado seguindo as observações enumeradas a seguir e considerando-se o objetivo do teste.

(1) Em relação ao texto, deve-se analisar:

- (a) a adequação ao nível avaliado;

- (b) o tema abordado [que não se configure em viés cultural (discriminação, preconceito – gêneros, etnias, profissões, crenças, religiões etc., apologia a condutas em desacordo com preceitos educativos e legais, temas controversos)];
- (c) erros de digitação ou de transcrição.

(2) Em relação ao enunciado, deve-se observar:

- (a) a forma da redação apresentada;
- (b) a clareza e objetividade;
- (c) a adequação da linguagem ao aluno da série a que se destina o item;
- (d) a utilização de expressões como: Assinale a alternativa correta, Qual das alternativas..., A alternativa que indica...;
- (e) a utilização de termos como: sempre, nunca, todo, totalmente, absolutamente, completamente, somente;
- (f) a consistência da redação em relação ao modo gramaticalmente correto, sem dicas verbais.

(3) Em relação às alternativas, deve-se observar:

- (a) se possui, indubitavelmente, somente uma alternativa correta;
- (b) se há coerência entre as alternativas e o enunciado;
- (c) se a redação está correta, do ponto de vista gramatical;
- (d) se não há atração a alguma alternativa, em decorrência da forma (tamanho, formato);
- (e) se todas as alternativas são claras e objetivas em sua linguagem.

(4) Em relação aos conteúdos, avaliar:

- (a) se o texto não se constitui um elemento impertinente na composição do item;
- (b) se o enunciado se relaciona com a competência a ser avaliada;
- (c) se o enunciado apresenta-se contextualizado, considerando-se o texto, ou, no caso de matemática, se ele apresenta uma situação de aplicação do conteúdo avaliado;
- (d) se a situação apresentada é factível e admissível;

- (e) se o problema expresso pelo enunciado testa a capacidade de raciocínio e não de memorização;
- (f) se há uma compreensão universal dos conceitos, fatos, terminologias etc. utilizados;
- (g) se nenhuma alternativa se configura como uma pegadinha;
- (h) se as alternativas apresentam-se coerentes com o conteúdo abordado pelo enunciado.

Por fim, devem ser levantadas as possibilidades das operações mentais empreendidas pelos alunos para fazer a opção a cada alternativa. Isso significa dizer que se trata de um exercício no qual o professor se coloca na posição do aluno que sabe e daquele que não sabe e tenta raciocinar como ele, descrevendo esse processo.

2.2.2 ANÁLISES PSICOMÉTRICAS MAIS COMUNS UTILIZADAS NA AVALIAÇÃO DE TESTES DE DESEMPENHO

A análise empírica dos itens é realizada com dados coletados de uma amostra representativa de sujeitos da população para a qual o teste está sendo construído, utilizando-se análises estatísticas. Essa análise pode ser feita utilizando-se a Teoria Clássica dos Testes e a Teoria de Resposta ao Item. As análises, embora utilizem técnicas estatísticas diferentes, fornecem dados que devem ser congruentes e se confirmarem.

2.2.2.1 A Teoria Clássica dos Testes

O modelo clássico da psicometria (Pasquali, 1997; Crocker & Algina, 1986) que fundamenta a Teoria Clássica dos Testes (TCT) está apoiado no seguinte paradigma: o escore empírico ou bruto do sujeito é constituído de dois componentes, a saber: (1) o escore real ou verdadeiro (V) do sujeito no traço latente e, (2) o erro de medida (E). O erro sempre presente em qualquer medida empírica, resulta no modelo fundamental da psicometria, de que o escore bruto de um examinando é a soma do escore verdadeiro e do erro ($T = V + E$). Esse modelo implica alguns postulados básicos: (a) o escore esperado é o escore verdadeiro. Esse postulado decorre do conceito de esperança matemática do escore empírico que significa que se o sujeito responde infinitas vezes ao mesmo teste ele terá infinitos diferentes escores empíricos e a média desses infinitos será o escore verdadeiro; (b) não há correlação entre o

escore verdadeiro e o erro, pois a correlação entre o escore verdadeiro e o erro é zero; portanto, não há nenhuma razão para se supor que escores verdadeiros maiores terão erros positivos e escores verdadeiros menores terão erros negativos ou vice-versa; e (c) os erros em testes paralelos não são correlacionados.

Ainda segundo Pasquali (1997) e Crocker & Algina (1986), o modelo da TCT é baseado em dados empíricos coletados de um conjunto de itens agrupados inicialmente de maneira intuitiva. O teste pode ser construído por meio da seleção de uma amostra de itens coletados de um universo que parece medir um dado construto. Essa maneira de construir instrumentos psicométricos está fundamentada na idéia de que existe para cada construto um conjunto indefinido de itens do qual uma amostra é extraída para construir o teste. A definição dos itens que compõem o teste é feita por meio da validade aparente, ou seja, escolhem-se os itens que parecem estar medindo a mesma coisa. Na TCT, os parâmetros do item e da habilidade são dependentes da amostra e do teste.

A validade na TCT consiste na verificação da hipótese de que o teste é capaz de prever um critério externo, no qual este é representado por comportamentos. Assim, a demonstração da validade na TCT é uma questão de legitimação do instrumento em relação ao erro de estimação, ou seja, é a verificação da magnitude do escore verdadeiro concebido como representante legítimo do traço latente. Essa verificação pode ser feita utilizando-se o modelo da análise fatorial.

2.2.2.2 Teoria de Resposta ao Item

A Teoria de Resposta ao Item (TRI) representou um grande avanço na análise dos dados coletados dos testes de desempenho de uma avaliação educacional. Sua função real em uma avaliação é *sumarizar informações complexas, reduzindo um grande conjunto de dados em poucas e resumidas estatísticas tratáveis e interpretáveis. Por exemplo, dos escores individuais alcançados em um teste cognitivo, pode-se estimar – e reproduzir em algum grau – as respostas reais dos indivíduos aos itens. Indivíduos com altos escores tendem a acertar mais itens, enquanto indivíduos com baixos escores tendem a acertar menos. Se diferentes itens são randomicamente designados a examinandos, então se pode também estimar como um examinando teria desempenhado sobre itens que não tenham sido administrados a ele. O escore é, portanto, um sumário simples de como os indivíduos desempenham sobre muitos itens na avaliação* (Beaton & Johnson, 1992, p. 167).

Em testes educacionais, normalmente são abordados itens de múltipla escolha e de respostas construídas. A TRI tem modelos diferentes para cada um dos tipos de respostas. A seguir, são apresentadas algumas propriedades dos modelos para cada tipo de resposta.

2.2.2.2.1 Itens dicotomicamente escoreados

A idéia básica da TRI se apóia em dois postulados básicos: (a) o desempenho de um examinando em um teste pode ser predito ou explicado por um conjunto de fatores chamados traços latentes ou habilidades; e (b) o relacionamento entre o desempenho de um examinando no item e o conjunto de traços subjacentes ao desempenho no item pode ser descrito como uma função monotonicamente crescente, chamada função característica do item ou curva característica do item. Essa função especifica que à medida que o nível do traço aumenta a probabilidade de uma resposta correta ao item aumenta. Portanto, examinandos com valores mais altos no traço têm probabilidades mais altas de responderem corretamente ao item do que estudantes com valores mais baixos no traço, independentemente do grupo a que pertencem (Hambleton, Swaminathan e Rogers, 1991).

Existem muitos modelos possíveis de resposta ao item que se diferem na forma matemática da função característica do item e/ou no número de parâmetros especificados no modelo. Todos os modelos de TRI contêm um ou mais parâmetros descrevendo o item e um ou mais parâmetros descrevendo o examinando.

Os modelos de TRI, ao contrário do modelo da TCT, são modelos falseáveis (van der Linden e Hambleton, 1997), pois um dado modelo de TRI pode ou não ser apropriado para um conjunto particular de dados de um teste, isto é, o modelo pode não predizer ou explicar adequadamente os dados. Em qualquer aplicação da TRI, é essencial avaliar a adequação do modelo aos dados. Quando um modelo de TRI é adequado aos dados do teste de interesse, várias características desejáveis são obtidas. As estimativas da habilidade dos examinandos não são dependentes do teste e os índices não são dependentes do grupo. Estimativas de habilidade obtidas de diferentes conjuntos de itens serão as mesmas (exceto por erros de medida) e as estimativas dos parâmetros do item em diferentes grupos de examinandos serão também as mesmas (exceto por erros amostrais). Na TRI, os parâmetros do item e da habilidade, obedecidos os pressupostos básicos, são invariantes e essa propriedade é obtida pela iteração da informação acerca do processo de estimação das habilidades dentro do processo de estimação dos parâmetros do item.

Segundo Hambleton, Swaminathan e Rogers (1991), os modelos matemáticos empregados na TRI pressupõem que a probabilidade de um examinando responder a um dado item corretamente depende da habilidade ou habilidades do examinando e das características do item. A TRI inclui um conjunto de pressupostos acerca dos dados para os quais o modelo será aplicado. O pressuposto mais comum da TRI é o de que somente uma habilidade esteja sendo medida pelos itens que compõem o teste. Esse é o pressuposto da unidimensionalidade. Relacionado a esse conceito, está outro pressuposto, que é o da independência local. Esse pressuposto significa que as respostas aos itens dependem do nível de habilidade que está sendo medido e não de outros níveis de habilidades. Assim, hipoteticamente, as respostas dos examinandos para qualquer par de itens serão estatisticamente independentes. Outro pressuposto para todos os modelos da TRI é o de que a função característica do item deve refletir o relacionamento verdadeiro entre variáveis não-observáveis (habilidades) e variáveis observáveis (respostas aos itens). função matemática que relaciona a probabilidade de sucesso (dada uma resposta correta) em um determinado item segundo a habilidade medida pelo teste e segundo as características do item. A escolha do número de parâmetros a serem usados envolve pressupostos acerca dos dados que podem ser verificados mais tarde pelo exame de quão bem o modelo explica os resultados observados pelo teste. Os três modelos de TRI mais populares são os modelos logísticos de um, dois e três parâmetros e, historicamente, o modelo da ogiva normal.

O modelo logístico de um parâmetro é bastante usado, posto que é considerado o mais estável dos três modelos da TRI, e é mais comumente chamado de Rasch, em homenagem a quem primeiro publicou este modelo, o dinamarquês Georg Rasch. A função característica do item para o modelo logístico de um parâmetro é dada pela equação:

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \quad i = 1, 2, \dots, n$$

onde,

$P_i(\theta)$ é a probabilidade de um examinando com habilidade θ responder corretamente um item i .

b_i é o parâmetro de dificuldade do item

n é o número de itens do teste

e é um número transcendental, cujo valor aproximado é 2,718

O modelo logístico de dois parâmetros foi desenvolvido por Birnbaum em 1968 e é dado pela equação:

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} \quad i = 1, 2, \dots, n$$

onde, onde $P_i(\theta)$, b_i e n , e são definidos como no modelo anterior e

$D = 1,7$, sendo este um fator introduzido para tornar a função logística tão próxima quanto possível da função ogiva normal.

a_i = parâmetro de discriminação do item.

O modelo logístico de três parâmetros é dado pela expressão matemática:

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} \quad i = 1, 2, \dots, n$$

onde $P_i(\theta)$, b_i , n , e , D e a_i são definidos como nos modelos anteriores e

c_i = parâmetro de pseudo-chance.

O parâmetro c_i representa a possibilidade de uma assíntota maior que zero para a curva característica do item e representa a probabilidade de examinandos com baixa habilidade responderem corretamente ao item.

2.2.2.2.2 Itens politomicamente escorados

Grande parte das avaliações internacionais já consolidadas e reverenciadas, como por exemplo, a do sistema norte-americano - NAEP, apresenta testes que contêm, em sua maior parte, itens com formatos de respostas construídas. As informações que esses testes fornecem em relação à evolução da construção do conhecimento dos alunos são muito mais consistentes do que aquelas fornecidas apenas por itens com o formato de múltipla escolha.

Esse tipo de testagem exige outros tipos de análises e de recomendações técnicas para a construção dos itens. As análises utilizadas para os itens de múltipla escolha são baseadas nas modernas técnicas que abordam a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI). Da mesma forma, os itens de respostas construídas devem utilizar essas técnicas.

O diferencial entre a análise dos itens de resposta construída e os de múltipla escolha é que aqueles consistem em respostas classificadas em duas ou mais categorias predefinidas,

ao passo que, para as respostas de múltipla escolha, as opções são certo ou errado. Entretanto, o item de múltipla escolha também pode apresentar a possibilidade de trazer alternativas às quais se possa atribuir créditos parciais, ou seja, elas podem representar respostas parcialmente verdadeiras, que podem acompanhar a construção do conhecimento do aluno, configurando-se também um formato interessante na interpretação dos resultados.

Os itens de respostas construídas são de dois tipos: de respostas curtas e de respostas extensas. As respostas curtas são predefinidas como certas ou erradas, e as extensas devem ser analisadas por especialistas da área que poderão avaliá-las como: respostas completamente corretas, respostas parcialmente corretas ou respostas totalmente erradas, podendo atingir várias categorias, e diagnosticarão concepções e modos de raciocinar certos ou errados dos alunos.

O alto custo da escolha por esse formato de item (resposta construída) nos testes é compensado pela série de informações adicionais fornecidas quando comparadas às respostas aos itens de múltipla escolha. A ancoragem multicategórica é mais informativa porque permite apontar as dificuldades correspondentes aos limites entre as diversas categorias atribuídas na correção do item, informações essas que são difíceis de serem extraídas de questões dicotômicas.

O processo de correção desses itens exige a elaboração de um guia de ancoragem para análise dos itens, com as categorias necessárias e a produção de descrições detalhadas para cada um deles. Esses itens são submetidos à avaliação de especialistas que têm por base o guia. Feita a correção, a base de dados é montada, geralmente, utilizando programa desenvolvido para esse fim e a análise dos itens é realizada pelo programa PARSCALE, que utiliza o modelo de Crédito Parcial Generalizado de Masters ou o Modelo de Resposta Graduada de Samejima.

O modelo de crédito parcial foi desenvolvido por Andrich em 1978 e reformulado por Masters em 1980 (Muraki e Bock, 1997) para a análise de itens que não apresentam somente as categorias certo e errado ou verdadeiro e falso. O item é pontuado parcialmente, e o respondente ganha créditos à medida que sua resposta se aproxima da completa.

A equação fundamental do modelo de crédito parcial generalizado é baseada na probabilidade que uma pessoa com escore θ_k em uma escala k terá para dar uma resposta x_j , de um enésimo item j que é escoreado em categorias m_j , de escores ordenados *i-ésimo*:

$$P(x_j = i | \theta_k, a_j, b_j, d_j, l, \dots, d_j, m_j - 1) = \frac{\exp(\sum_{v=0}^i D a_j (\theta_k - b_j + d_j, v))}{\sum_{g=0}^{m_j-1} \exp(\sum_{v=0}^g 1.7 a_j (\theta_k - b_j + d_j, v))} \equiv P_{ji}(\theta_k)$$

onde,

m_j é o número de respostas ao item j ;

x_j é a resposta ao item j , com possibilidades 0, 1, ..., $m_j - 1$;

$D = 1,7$, sendo este um fator introduzido para tornar a função logística tão próxima quanto possível da função ogiva normal.

a_j é o parâmetro de discriminação;

b_j é o parâmetro de localização, caracterizando a dificuldade geral; e

d_j é o parâmetro b da categoria;

As indeterminações nos parâmetros do modelo acima são resolvidas pela fixação de $d_j = 0$ e de $\sum_{i=1}^{m_j-1} d_{j,i} = 0$. Muraki (1992) pontua que $b_j - d_{ji}$ é o ponto na escala θ_k no qual as curvas de $P_{j,i-1}(\theta_k)$ e $P_{ji}(\theta_k)$ intersectam e então caracterizam o ponto sob a escala θ_k no qual a resposta para o item j tem a probabilidade igual de falhar na categoria de resposta $i-1$ e na resposta de categoria .

Tanto o modelo de três parâmetros como o modelo de crédito parcial generalizado indicam que têm uma indeterminação linear da escala θ , isto é, os parâmetros do item são estimados em uma métrica diferente. Para o propósito de relatar estimativas dos parâmetros do item e outras estimativas intermediárias, a indeterminação aparente pode ser resolvida por meio da escolha arbitrária de uma origem e do tamanho da unidade em uma dada escala. Na maioria dos casos, uma distribuição com média zero e desvio padrão um é empregada na escala padronizada provisória. Resultados finais para cada disciplina avaliada são linearmente transformados da escala θ para uma escala de 0 a 500 ou de 0 a 300.

O modelo criado por Samejima (1997) pressupõe que codificando as respostas do sujeito entre 0 e $m-1$ (sendo m o número de respostas distintas), a probabilidade de ocorrer a escolha de uma categoria maior ou igual a k , em função do nível de habilidade θ , é dada pela função:

$$P_{jk}(\theta) = \frac{1}{1 + \exp[-D a_j (\theta - b_j + c_k)]} - \frac{1}{1 + \exp[-D a_j (\theta - b_j + c_{k+1})]}$$

onde $D = 1,7$

Esta função é a mesma que se utiliza para o modelo logístico de dois parâmetros. Por tal motivo se afirma que tanto o modelo de resposta graduada como o modelo de crédito parcial são uma generalização deste modelo ao caso politômico, em especial, os que utilizam o modelo da escala Likert (Embretson & Reise, 2000).

2.2.2.3 Análise da dificuldade dos itens

Cada item em um teste está contribuindo com uma parcela significativa para o resultado final na avaliação. A análise de itens possibilita identificar aqueles itens que desempenham bem e aqueles que são problemáticos em relação à produção da informação desejada. Um teste que tenha sido desenvolvido para comparar os escores dos indivíduos, resultará em uma grande extensão de escores.

Um parâmetro importante a ser analisado é a dificuldade do item. A dificuldade do item pode ser definida pela TCT, como a porcentagem de examinandos que respondem corretamente a um dado item. O cálculo da dificuldade do item, ou o valor p é feito pela divisão do número de pessoas que acertaram o item pelo número total de pessoas que responderam ao item. Geralmente, itens com um índice de dificuldade 0,5 produzem distribuições de escores no teste com maior variação. Os itens com dificuldades fora da faixa média devem ter um número reduzido e limitado. Diversos autores como Bloom (1971), Vianna (1987), McIntire e Miller (2000), Anastasi e Urbina (2000) consideram que a média da dificuldade dos itens deva ficar em torno de 0,50.

Quanto mais fácil o item, maior a probabilidade de as pessoas o acertarem e, quanto mais difícil, menor a chance de acerto. Os itens, geralmente, são arranjados em ordem de dificuldade, de modo que os examinandos comecem com itens relativamente mais fáceis prosseguindo para itens de dificuldade mais alta. Esse arranjo, geralmente, deixa os testandos com mais confiança. No caso dos testes atuais do SAEB, a orientação é que se inicie com itens mais fáceis, mas que no prosseguimento haja uma alternância entre itens fáceis, medianos e difíceis para que o aluno não se desanime quando se deparar com itens mais difíceis.

Os modelos de TRI permitem que a partir de informações indiretas sobre uma determinada característica não observável do sujeito (traço latente – θ), possa-se obter uma estimativa dessa característica (dificuldade) para cada indivíduo da amostra. O parâmetro b para um item é o ponto na escala da habilidade onde a probabilidade de uma resposta

correta, considerando-se o modelo de um ou dois parâmetros, é 0,5. Ao considerar-se o modelo de três parâmetros, essa probabilidade é somada ao valor do parâmetro c dividido por dois. Esse é, portanto, um parâmetro de localização, indicado pela posição da Curva Característica do Item (CCI) em relação à escala de habilidade. Quando os valores de habilidade de um grupo são transformados para a escala de média 0 e desvio padrão 1, os valores de b normalmente variam de -2 a $+2$, sendo que os valores de b próximos de -2 correspondem aos itens que são mais fáceis e, ao contrário, valores de b próximos de $+2$, correspondem aos itens que são mais difíceis para o grupo de examinandos.

2.2.2.4 Análise da discriminação dos itens

Discriminação refere-se ao poder de um item em diferenciar sujeitos com magnitudes de traços diferentes do qual o item constitui a representação comportamental (Pasquali, 1997). Quanto mais próximas forem as magnitudes do traço que o item puder diferenciar, mais discriminativo ele será. Estatisticamente, um dos conceitos de discriminação que tem por base a TCT, é dado pela correlação dos escores dos sujeitos no item com seus escores no teste total. De acordo com Marshall e Hales (1972) em Wilson, Wood e Gibbons (1991), existem mais de 60 índices propostos para medir o poder de discriminação de um item.

Dessa conceituação de discriminação surgem duas formas mais comuns de calcular estatisticamente índices de discriminação, (1) por meio da TCT: (a) dos grupos critérios; (b) da correlação do item com o total dos itens e, (c) da correlação inter-itens; e, (2) utilizando a TRI.

2.2.2.4.1 Análise da discriminação por meio da TCT

a) Grupos critério

O índice de discriminação dado pelos grupos critérios é estabelecido pela diferença entre grupos verificada em algum comportamento estabelecido como referência aos objetivos do teste. Tipicamente é escolhido o escore total no próprio teste para determinar grupos extremos de sujeitos: grupo superior e grupo inferior. Em amostras grandes, selecionam-se os 27% superiores e os 27% inferiores para comporem os dois grupos (Kelley, 1939). Com base nesses grupos critério pode-se calcular o índice de discriminação por meio da estatística **D** e do teste **t** de Student. O índice **D** consiste simplesmente na diferença de desempenho no teste

para um item representada pela porcentagem de acertos no grupo superior (27%) e no grupo inferior (27%). O índice **D** deveria ser positivo e quanto maior for mais discriminativo será considerado o item.

O teste **t** consiste na análise da diferença entre as médias obtidas no teste pelo grupo superior e pelo grupo inferior. Nesse caso, é necessário o cálculo das médias e das variâncias que pressupõem que os escores constituam uma variável contínua e que ambos os grupos possuam variância. O nível de significância do teste **t** pode ser verificado com exatidão em tabelas estatísticas próprias. A fórmula para o cálculo do teste **t** é a seguinte:

$$t = \frac{\overline{X}_s - \overline{X}_i}{\sqrt{\frac{S_s^2}{n_s} + \frac{S_i^2}{n_i}}} \quad \text{com graus de liberdade, gl} = n_s + n_i - 2$$

onde

\overline{X}_s e \overline{X}_i = médias do grupo inferior e superior

S_s^2 e S_i^2 = variâncias dos dois grupos

n_s e n_i = número de sujeitos nos dois grupos

b) Correlação item-total

O maior número de índices de discriminação decorre dessa forma de cálculo. As mais comuns são a correlação ponto bisserial e a correlação bisserial.

A correlação **ponto bisserial** é uma correlação produto-momento entre duas variáveis, quando uma das variáveis é binária e a outra é o teste completo, de distribuição intervalar. A fórmula é:

$$r_{pb} = \frac{\overline{X}_A - \overline{X}_T}{S_T} \sqrt{\frac{p}{q}}$$

onde,

\overline{X}_A = média no teste dos sujeitos que acertam o item

\overline{X}_T = média total do teste

S_T = desvio padrão do teste

p = proporção de sujeitos que acertam o item

$$q = 1 - p$$

Observe que a correlação ponto bisserial é também uma função da facilidade do item sendo, portanto, influenciada por ela.

A **correlação bisserial** é uma medida na qual se assume que a variável latente subjacente ao desempenho no item é normalmente distribuída, tornando possível derivar uma fórmula para a correlação entre essa variável e um critério de distribuição contínua como os escores de um teste. Ou seja, ela é uma medida de associação entre o desempenho no item e o desempenho no teste. A correlação bisserial é menos influenciada pela dificuldade do item e tende a ser invariante de uma situação de testagem para outra (Wilson, Wood e Gibbons, 1991). Sua fórmula é:

$$r_b = \frac{\overline{X}_A - \overline{X}_T}{S_T} x \frac{pq}{y}, \text{ onde}$$

\overline{X}_A = média no teste dos sujeitos que acertam o item

\overline{X}_T = média total do teste

S_T = desvio padrão do teste

p = proporção de sujeitos que acertam o item

$$q = 1 - p$$

y = é a ordenada na curva normal no ponto de divisão dos segmentos que contêm as proporções **p** e **q** dos casos.

c) **Correlação inter-itens**

A **correlação phi** (Φ) é utilizada quando as duas variáveis a se correlacionarem são originalmente dicotômicas. A fórmula é a seguinte:

$$\Phi = \frac{ad - bc}{\sqrt{(a + b)(a + c)(b + d)(c + d)}}$$

a, b, c e d são as possíveis combinações de erro e acerto dos dois itens a se correlacionarem;

a = acerto e acerto

b = acerto e erro

c = erro e acerto

d = erro e erro

A **correlação tetracórica** é utilizada quando duas variáveis contínuas, não observadas e normalmente distribuídas, relacionadas linearmente, foram artificialmente reduzidas a duas categorias (dicotomizadas), o mais próximo possível à mediana.

A fórmula é:

$$r_t = \frac{ad - bc}{\sqrt{yy'} N^2},$$

onde, *a*, *b*, *c* e *d* representam as frequências da tabela 2 x 2.

N = n° de sujeitos

yy': são os valores das ordenadas nos pontos p e p'

2.2.2.4.2 Análise da discriminação por meio da TRI

A TRI introduziu técnicas nessa área que, embora sejam complexas e complicadas, são fundamentais para a análise da qualidade dos itens. Na TRI, a discriminação do item para modelos de dois ou três parâmetros é chamado parâmetro a_i . O parâmetro a_i é proporcional à inclinação (*slope*) da Curva Característica do Item (CCI) no ponto b_i , no momento da inflexão, ou seja, quando a curva CCI corta a linha que corresponde à probabilidade de 0,5 de resposta correta na escala da habilidade, para o modelo de dois parâmetros, ou 0,5 mais valor do parâmetro $c/2$, para o modelo de três parâmetros (Hambleton, Swaminathan & Rogers, 1991; Pasquali e Primi, 2003, Pasquali, 2007). Itens com inclinações mais altas são mais úteis para separar examinandos dentro de diferentes níveis de examinandos do que aqueles itens com inclinações mais baixas. Teoricamente, o parâmetro de discriminação do item pode assumir valores na escala de 0 a $+\infty$. Embora não tenham sido encontradas na literatura informações acerca da indicação de critérios para determinação do índice de qualidade do parâmetro a , por meio desse estudo, observou-se que itens com parâmetros

inferiores a 0,60 e superiores a 4,00 discriminam pouco e não devem ser considerados na análise.

2.2.2.5 Análise da unidimensionalidade dos itens

A unidimensionalidade de um conjunto de itens é um pressuposto fundamental tanto da TCT quanto da TRI. A unidimensionalidade está definida como a existência de um único traço latente subjacente ao conjunto de itens. Parece desejável, que um teste de habilidade matemática, por exemplo, não esteja também medindo um outro construto como habilidade verbal. Em um estudo realizado sobre a unidimensionalidade dos itens dos testes do SAEB 1997 e 1999, pelos pesquisadores Laros, Pasquali e Rodrigues (2000), é chamada a atenção para as conseqüências negativas que a violação deste pressuposto incorre.

Entre essas conseqüências, pode-se citar a que se relaciona à validade de construto do teste que conduz a uma diminuição de sua validade, dificultando a interpretação de seus escores. Outra conseqüência negativa diz respeito ao fenômeno chamado *Differential Item Functioning – DIF* (função diferencial do item) para diferentes culturas ou grupos étnicos. Se estes grupos não são análogos nas habilidades que eles utilizam para solucionar os itens, o potencial para o viés do item existirá. As questões de viés do item e validade de construto são conceitos inter-relacionados. Se em um teste falta validade de construto, o teste conterá itens que estarão medindo outras habilidades que não aquelas que se propôs medir e, portanto, o potencial para viés do item também existirá. Esse viés pode surgir se grupos de interesse diferem em sua distribuição subjacente dessas habilidades alheias (Ackerman, 1992). A terceira conseqüência negativa de violações da unidimensionalidade é o impacto negativo sobre a equalização (*equating*) das várias formas de uma prova ou teste. A equalização propriamente dita é muito difícil com testes multidimensionais, mesmo quando um modelo de resposta ao item multidimensional é utilizado (Dorans e Kingston, 1985). E o quarto efeito negativo de violações da unidimensionalidade está relacionado à estimação da proficiência do aluno. Em testagem educacional, em que o propósito é medir proficiências em tópicos amostrados de um domínio conceitual e unidimensional, a ocorrência da multidimensionalidade é problemática.

Portanto, a verificação do pressuposto de unidimensionalidade dos instrumentos de avaliação educacional é uma questão de fundamental importância para testes de avaliação de larga escala e deve ser um dos primeiros passos nos procedimentos de análises. O método

mais popular de avaliação da unidimensionalidade é por meio da análise fatorial. Entretanto, Hambleton e Swaminathan (1985) chamam a atenção para o fato de que sendo a análise fatorial um procedimento linear, pode não produzir uma única dimensão quando existe uma considerável não-linearidade nos dados. Apesar disso, eles insistem que ela deva ser rotineiramente realizada. Eles sugerem que a análise fatorial tenha por base a correlação tetracórica.

Um dos modelos para verificar o pressuposto da unidimensionalidade é a *Full-Information Factor Analysis*, principalmente porque essa análise adota o modelo multifatorial de Thurstone e se livra de uma série de dificuldades que a análise fatorial convencional encontra. Este foi o modelo adotado na pesquisa empreendida pelos pesquisadores Laros *et al* (2000) na análise da unidimensionalidade dos itens dos testes do SAEB. A seguir, a discussão sobre esse método de análise.

2.2.2.5.1 O método da análise fatorial *full information*

A *Full-Information Factor Analysis* - FIFA (Bock, Gibbons e Muraki, 1988) é um método de análise fatorial de itens introduzido por Bock e Aitkin (1981) baseado na TRI que não requer o cálculo de coeficientes de correlação inter-itens. A abordagem de Bock e Aitkin é chamada análise fatorial “*full information*” (de informações completas) porque usa como dados as frequências de todos os vetores distintos de resposta ao item, em contraste aos métodos de informação limitada (ver Cristofferson, 1975 e Muthén, 1978) baseados nas frequências conjuntas de ordem reduzida dos escores dos itens.

O método de análise fatorial de itens *FIFA* é baseado no modelo multifatorial de Thurstone, implementado pela estimativa *Marginal Maximum Likelihood* (MML) e o algoritmo *Expectation-Maximization* (EM). Este método permite que se trabalhe com os efeitos do *acerto ao acaso* (probabilidade de acerto ao acaso) dos itens de múltipla escolha, com itens omitidos e itens não apresentados.

O modelo de Bock e Aitkin assume que os dados tenham sido obtidos de uma amostra de pessoas delineadas de uma distribuição normal e que os fatores não são correlacionados, mas estes pressupostos podem ser relaxados para permitir fatores correlacionados e uma distribuição não-normal. Por meio do uso da estimativa de máxima verossimilhança, estima-se a dificuldade e as cargas fatoriais dos itens. Na parte E do algoritmo EM usado com a estimativa máxima verossimilhança, os seguintes dois valores são

estimados: (1) o número de pessoas esperadas para responder corretamente ao item, considerados os dados amostrais e, (2) o número de pessoas esperadas com habilidade X , a qual é normalizada para o tamanho da amostra. Na parte M do algoritmo EM uma análise probit-múltipla ou logit múltiplo melhora a estimação da dificuldade dos itens e das cargas fatoriais obtidas no passo E.

O princípio básico subjacente ao algoritmo EM é muito antigo e intuitivo. Dado um problema que é difícil solucionar, reformula-se esse problema de modo que se fossem observados os *missing data* (dados ausentes) a solução seria fácil de se obter. Especificamente, reformula-se o problema de modo que uma boa estimativa do parâmetro θ seja fácil de encontrar se além dos valores observados, os valores *missing* também fossem conhecidos. Conseqüentemente, insere-se um conjunto de valores para os *missing data* e soluciona-se o problema, ou seja, encontra-se uma estimativa para o parâmetro θ . Utilizando-se esta estimativa, encontram-se melhores valores para os *missing data* e, então, repete-se o processo de estimação para encontrar uma melhor estimativa para o parâmetro θ . Esse processo de iteração é continuado até que as estimativas do θ convirjam. Resumidamente, a idéia básica do algoritmo EM é a seguinte: inserir *missing data*, estimar parâmetros, reestimar *missing data* até que nenhuma mudança real ocorra na estimação dos parâmetros.

2.2.2.5.2 Teste de decisão do número de fatores

No estudo realizado por Laros *et al* (2000) sobre a unidimensionalidade dos itens, foi abordado o teste de decisão do número de fatores baseado no modelo relacionado à alternativa multinomial geral (Wilson, Wood e Gibbons, 1991). Neste modelo, se o tamanho da amostra é suficientemente grande que todos os 2^n possíveis padrões de resposta têm valores esperados maiores que 1 ou 2, a aproximação qui-quadrado para o teste da *deviance*, ou o teste da razão de verossimilhança é igual a $-2 \log likelihood$ e indica o grau da falta de ajuste entre o modelo e os dados.

$$G^2 = 2 \sum r_l \ln \left(\frac{r_l}{NP_l} \right)$$

onde r_l é a freqüência de padrões de resposta l , P_l é a probabilidade de observar o padrão l , calculado com base nas estimativas de máxima verossimilhança dos parâmetros dos itens e

$N = \sum r_i$, que é a soma das frequências de padrões de resposta que é igual ao número de casos N . Os graus de liberdade do teste de adequação do modelo são:

$$2^n(m + 1) + m(m - 1) / 2,$$

onde n é o número de itens e m é o número de fatores.

Quando o número de padrões é maior que o tamanho da amostra, o que acontece frequentemente, algumas das frequências esperadas podem ser próximas de zero. Neste caso, o teste de *goodness-of-fit* torna-se inacurado e pode não ser confiável. Os padrões de resposta aumentam rapidamente. Por exemplo,

- com um item: temos $2^1 = 2$ padrões possíveis, isto é, 0 e 1;
- com dois itens: temos $2^2 = 4$ padrões possíveis de resposta, isto é, 0-0, 0-1, 1-0, 1-1;
- com três itens: temos $2^3 = 8$ padrões;
- com 150 itens ou mais (como é o caso do SAEB): temos $2^{150} =$ um número com 45 dígitos (astronômico!).

Os problemas com o teste de adequação do modelo aumentam quando a matriz de correlação contém *missing data*. Kaplan (1985) mostrou que a estatística qui-quadrado desse teste é extremamente sensível aos *missing data*, que são o resultado do uso do delineamento *Balanced Incomplete Block* - Blocos Balanceados Incompletos (BBI). Kaplan concluiu que se deve tomar muito cuidado com a interpretação do teste de *goodness-of-fit* do número de fatores quando o delineamento BBI é utilizado. Entretanto, Haberman (1977) demonstrou que em grandes amostras, a diferença dos qui-quadrados entre dois modelos é distribuída independentemente como um qui-quadrado com graus de liberdade igual à diferença entre os graus de liberdade do primeiro e segundo modelo, mesmo que a tabela de frequência seja esparsa. Então a contribuição do último fator adicionado ao modelo é significativa se a mudança correspondente do qui-quadrado for estatisticamente significativa.

No manual do TESTFACT (Wilson, Wood e Gibbons, 1991), os autores alertam para a interpretação cuidadosa da diferença dos qui-quadrados entre dois modelos. Especificamente em pesquisas de larga escala em locais diferentes, efeitos grupais podem aumentar os valores da diferença dos qui-quadrados entre dois modelos. Para ser cauteloso em relação ao número de fatores em tais estudos, os autores aconselham dividir o qui-quadrado por 2 ou 3 antes de avaliar a significância.

2.2.2.6 Análise Gráfica dos Itens

Esta é uma técnica introduzida no Brasil por T. A. van Batenburg em 2000 e desenvolvida por J. A. Laros, que é baseada no pressuposto de que os construtores de itens devem conhecer muito bem o conteúdo ensinado e seus objetivos. Eles precisam de habilidades específicas para construir um bom item de múltipla escolha. Construir um item bom é uma tarefa complexa, o item deve ter uma – e somente uma – resposta correta, sem haver nenhuma discussão a esse respeito. As alternativas devem ser atrativas, mas não confusas. Algumas vezes é importante não somente avaliar a dimensão de reconhecimento do que foi ensinado, mas também a dimensão de aplicação e de compreensão.

Procedimentos de estimação de modelos e de TRI têm se tornado muito sofisticados e eles são de domínio de psicometristas. Existe uma grande fenda entre psicometristas, construtores e usuários de testes. Usuários práticos freqüentemente não entendem TRI. Psicometristas freqüentemente negligenciam o contexto do uso prático dos testes. Elaboradores de itens dão o melhor de si para construírem bons itens e são difíceis de serem convencidos de que alguns de seus itens não são bons por razões psicométricas. Van Batenburg e Laros (2001) em seus estudos, tentaram reduzir a fenda entre psicometristas, usuários práticos e construtores de testes oferecendo gráficos fáceis de interpretar e que permitem uma compreensão dos dados dos testes, podendo ser usados para a seleção de itens.

Os dois pressupostos, considerando-se avaliação de desempenho e que representam a propriedade estatística básica de qualquer teoria de medida, são que, um aluno que dá uma resposta certa em um item de múltipla escolha, sabe mais que um aluno que dá a resposta errada e que um aluno que tem mais itens certos sabe mais que um aluno com menos itens certos. Destes dois pressupostos, segue a definição de um bom item, segundo van Batenburg e Laros (2001):

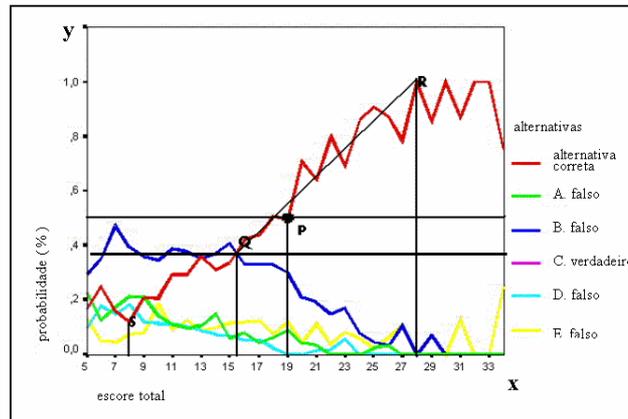
“um bom item apresenta porcentagens de resposta na alternativa correta que crescem com um aumento do escore total, e mostra porcentagens de resposta nas alternativas falsas (distratores) que decrescem com um aumento do escore total”

Isso implica que aqueles que têm todos os itens corretos (o escore máximo) têm uma probabilidade 1 (um) de terem marcado o item correto; aqueles que têm todos os itens errados, têm uma probabilidade 0 de terem marcado o item correto. Entretanto, num caderno com 39 itens de múltipla escolha com cinco alternativas, como é o caso do SAEB, poucos

alunos terão o escore 0 (zero) por causa da possibilidade de acerto ao acaso. Um aluno que somente “chuta” as questões terá uma chance de acertar aproximadamente oito questões ($39 \times 0,20$). Assim, pode ser esperado que a proporção de acerto ao item aumente de 0,20 para 1 com o escore total. Espera-se, também, que as alternativas falsas decresçam por causa da possibilidade de acerto ao acaso de 0,20 para 0 (quando todas as alternativas são igualmente atrativas) com um aumento do escore total. Até determinado escore, pode-se esperar que as alternativas certas e as alternativas falsas fiquem nos valores da chance de acerto ao acaso, nesse caso, 0,20. Depois deste escore total específico, a proporção de alternativas corretas aumenta e a proporção de alternativas falsas decresce. Se este escore total é um escore baixo, o item será provavelmente fácil. Se o escore total é alto, o item será provavelmente difícil. A dificuldade do item também depende da inclinação (*slope*). Na TRI, considerando-se o modelo de dois e três parâmetros, a dificuldade de um item é definida no ponto onde a linha de proporção 0,5 corta a “linha do item”. No modelo de três parâmetros, ao valor 0,5 é adicionado o valor do parâmetro c dividido por dois ($0,5+c/2$). Na abordagem visual isso é definido da mesma forma. A linha do item discrimina entre pessoas no intervalo acima das alternativas no máximo de 1 (um). Isso é chamado de intervalo de informação. Se as porcentagens de resposta à alternativa correta aumentam rapidamente com o escore total, o item terá um alto poder discriminativo, quando lentamente, terá um baixo poder discriminativo.

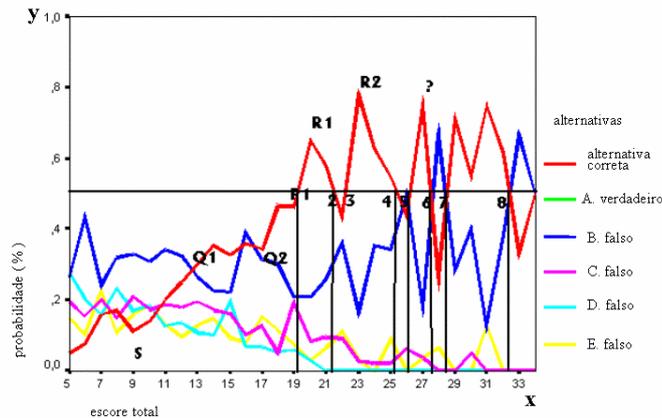
No método gráfico usado, as proporções das alternativas dos itens estão sendo plotadas em contraposição ao escore total. Inicialmente, usou-se o SPSS para fazer isso. Atualmente, J. A. Laros desenvolveu um software que, além de apresentar o gráfico do item, ele apresenta os índices de dificuldade e de discriminação do item e das alternativas. Os gráficos dão todas as informações necessárias para decidir selecionar ou não o item em um teste de realização e ainda oferece a chance de, por meio da análise de conteúdo, melhorar o item. Na figura 1, um exemplo de um gráfico do item e dos índices visualizados.

Figura 1 – Análise gráfica de um bom item



Na figura 1 (ilustração de Theo A. van Batenburg de um item de Física do SAEB 1999) observa-se no eixo Y as porcentagens de resposta nas alternativas t e no eixo X os escores totais (os escores de 0 a 5 e de 34 a 39 foram colocados juntos). As linhas das alternativas têm cores diferentes. A linha vermelha é a alternativa correta. As outras alternativas, exceto a rosa, são as alternativas falsas. A alternativa rosa não está no gráfico porque ela é a alternativa verdadeira e está sobreposta à linha vermelha. A linha de proporção 0,5 corta a linha vermelha no ponto P. O ponto P corresponde a um valor do escore total de 19, que pode ser interpretado como a dificuldade do item. Na linha da abscissa correspondente ao ponto Q, a linha vermelha da alternativa correta está acima das outras linhas. Aqui começa o intervalo de discriminação. O ponto Q corresponde a um valor do escore total de 15. A linha vermelha tem um valor de 1 no ponto R, o que demonstra o fim do intervalo de discriminação, no escore 28. A linha vermelha começa a crescer no ponto S, no escore 8. A linha vermelha aumenta até o máximo, no ponto R, no escore 28. A inclinação, que indica o valor discriminativo, pode ser calculada pela divisão $8/28 = 0,35$. Essas inclinações podem aumentar entre o mínimo de $5/35 = 0,14$ (intervalo máximo) e $34/35 = 0,97$. Esses números podem ser úteis na análise visual. Observa-se, ainda, neste exemplo, que alunos com escores baixos escolhem a alternativa azul mais frequentemente. A análise pedagógica do item pode explicar esse comportamento.

Figura 2 - Exemplo de um item ruim.



Também na Figura 2 a linha vermelha é a alternativa verdadeira. A linha vermelha corta a linha de proporção 0,5 oito vezes, correspondendo aos escores 19, 21, 25, 26, 27, 28 e 32. Por isso a dificuldade do item não pode ser determinada e este item deve ser, a princípio, excluído do teste. O item também nunca alcança o máximo de 1, assim o intervalo de discriminação e a inclinação (*slope*) não podem ser determinados. A alternativa azul escuro pode ser a causa de todos os problemas. A análise do conteúdo desse item também pode explicar o problema. Essa informação pode ser usada para melhorar o item e salvá-lo. Uma versão adaptada pode ser útil em outro teste.

2.2.2.7 Considerações sobre as análises

Ao relatar e discutir resultados de testes de desempenho de avaliações educacionais em larga escala, as análises apresentadas se tornam imprescindíveis, facilitando a compreensão de quem acompanha o processo e trazendo confiabilidade. Essas análises quando realizadas de forma integral asseguram a revelação de dados seguros a respeito do que os alunos sabem, de fato (Rodrigues 2002; 2006). A utilização dessas análises impede, sobremaneira, distorções na análise final dos resultados. Acrescente-se a isso, a necessidade imperativa de que os resultados sejam disponibilizados, não só para subsidiar a formulação ou reformulação de políticas públicas na área da educação, mas, principalmente, alcançar a pluralidade dos agentes intervenientes do processo educativo, no sentido também de formular ou reformular as práticas pedagógicas escolares.

A título de exemplificar a utilização das análises expostas, a seguir é apresentado um item do SAEB¹, do teste da 4ª série, Língua Portuguesa, edição 2001, selecionado para essa pesquisa, por apresentar boa qualidade.

Leia o poema abaixo.

O Porco e os Espinhos

Tem sempre uma pedra
no caminho
do amigo porco-espinho.
Ele corre no mato,
até se diverte,
dá susto nos outros
e tem namorada.
Mas, coitadinho,
não pode dar
abraço apertadinho.

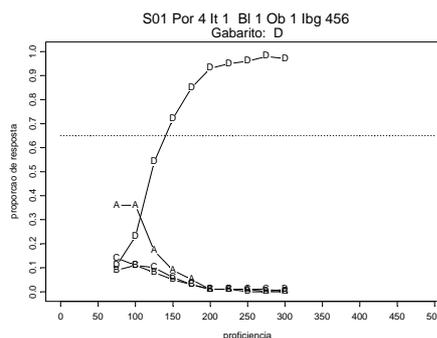
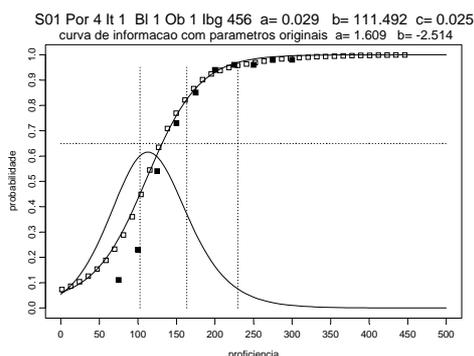
PIMENTEL, Luís. *Novas Idéias*. São Paulo. Editora do Brasil.

O porquinho não pode dar um abraço apertado porque ele

- (A) corre no mato.
- (B) dá susto nos outros.
- (C) tem namorada.
- (D) tem espinho.

ÍNDICES									PROPORÇÕES DE RESPOSTA					COEFICIENTES BISSERIAIS						
ITEM	BL	OB	GAB	DIFI	DISCR	ABAI	ACIM	BISE	A	B	C	D	""	""	A	B	C	D	""	""
1	1	1	D	.71	.60	.35	.95	.67	.11	.05	.06	.71	.08	.01	-.48	-.31	-.30	.67	-.46	-.63

¹ As análises psicométricas referentes a esse item foram realizadas pela Fundação Cesgranrio e pertencem ao SAEB.



Item: 1 Nível: 150

Tema: (4) Coerência e Coesão no Processamento do Texto

Descritor: (8) Estabelecer relação causa/conseqüência entre partes e elementos do texto.

A boa qualidade desse item é então confirmada pelas análises psicométrica e pedagógica. Psicometricamente, pode-se atestá-la pelos índices de discriminação, correlação bisserial do gabarito e das alternativas, pelos parâmetros a e c , pela curva característica do item e pela análise gráfica do item.

Os índices de discriminação, gerados pela TCT, o índice “D” e a correlação bisserial, e o parâmetro a , gerado pela TRI, foram, respectivamente, 0,60, 0,67 e 1,61, sendo todos eles indicadores de boa discriminação. As alternativas falsas apresentaram correlações bisseriais iguais ou inferiores a -0,30, indicando que essas alternativas só foram escolhidas por alunos que não alcançaram bom desempenho. O parâmetro c , de probabilidade de acerto ao acaso, foi 0,02, o que é considerado muito bom. O gráfico da CCI mostra um bom ajuste entre as respostas esperadas e as observadas. A análise gráfica do item, elaborada pela Fundação Cesgranrio, confirma o conceito emitido por van Batenburg e Laros, (2001), de que “um bom item apresenta proporção de resposta à alternativa correta que cresce à medida que aumenta o escore total e mostra proporções de respostas às alternativas falsas que vão decrescendo com o aumento do escore total”.

Para realizar a análise pedagógica dos itens, essa pesquisadora desenvolveu o quadro ilustrado no Apêndice A1, em que constam todas as informações consideradas importantes e citadas neste capítulo.

2.3. DESCRIÇÃO DOS RESULTADOS

A descrição dos resultados constitui-se uma das partes mais importantes do processo de avaliação educacional de sistemas. Beaton & Johnson (1992), p. 163, afirmam que *uma avaliação educacional é fundamentalmente um sistema designado a relatar a educadores, elaboradores de políticas e ao público geral sobre o status e as mudanças ocorridas no desempenho dos estudantes*. Os resultados descritos constituem o coração dos relatórios e é por meio deles que a avaliação alcançará seus propósitos. Eles afirmam ainda que, enquanto sistema de informações, deve-se ter a preocupação de que as informações produzidas e relatadas sejam úteis às diferentes audiências, não deixando de ser acuradas. Mas, além disso, um sistema de informações bem sucedido precisa relatar resultados de modo conciso e compreensível, de maneira a satisfazer às várias necessidades de informações para as quais a avaliação foi empreendida.

Os resultados são geralmente apresentados em escalas embasadas nas análises da TRI. *Essas escalas são indeterminadas; qualquer transformação linear reproduzirá bem as respostas aos itens* (Beaton & Johnson, 1992, p. 167). A maioria dos programas de computador que trabalha com a TRI produz escalas com média 0 e desvio padrão 1; portanto, os resultados se apresentam na escala, dispostos em intervalos negativos e positivos. Ao leitor dos resultados, a interpretação se torna mais difícil de compreender, pois um escore negativo tem outra conotação. Dessa forma, o NAEP, que foi a primeira instituição que trabalhou com a escala TRI, para dar um resultado mais significativo e intuitivo relata seus resultados em uma escala hipotética de extensão de 0 a 500, ficando os escores distribuídos, mais especificamente entre 100 e 400. Essa métrica foi escolhida para minimizar confusões com os escores de QI, do SAT (teste padronizado para admissão em faculdades ou universidades americanas) ou outras métricas equivalentes.

O processo avaliativo não abrange todas as proficiências dos estudantes, assim, uma amostra de tarefas das áreas a serem contempladas é selecionada. No caso do sistema norte-americano, esse processo é realizado por um grande comitê de especialistas das áreas avaliadas, educadores e cidadãos interessados. Esses comitês por área de aprendizagem especificam os objetivos para a avaliação em termos do que os estudantes deveriam realizar durante os determinados períodos de escolarização. Os objetivos são definidos e organizados em matrizes de dupla entrada, abrangendo conteúdos por processo, nas quais a porcentagem de itens é especificada para cada habilidade contemplada pelo comitê da área a ser avaliada.

Para satisfazer os objetivos da avaliação e assegurar que as tarefas selecionadas cubram uma extensão de níveis de dificuldade, o NAEP seleciona um grande número de tarefas dentro de cada área avaliada.

Beaton & Johnson (1992) alertam que nenhum sumário é capaz de capturar todas as informações que são disponibilizadas pelos dados. Algumas informações são perdidas. Entretanto, organizar os dados em escalas é um meio prático de reduzir dados complexos de avaliações.

Constata-se, por meio de revisão bibliográfica, que a maioria dos sistemas de avaliação fornece resultados acerca do desempenho alcançado pelos alunos relacionado às diversas variáveis de contexto pesquisadas. Normalmente, os resultados são apresentados por populações de estudantes, como por exemplo, alunos da 4^a, 8^a e 3^a séries, e subgrupos dessas populações, como gênero, raça, entre outros.

O desempenho dos alunos nas disciplinas pode ser relatado de duas formas: por meio de escores observados distribuídos em uma escala, geralmente, de 0 a 500, e por níveis que são categorizados conforme o desempenho alcançado e esperado para cada ciclo avaliado. As escalas com os escores retratam em termos quantitativos o desempenho dos alunos. Os níveis de desempenho são usados para relatar os resultados por meio de um conjunto de padrões que expressa o que os estudantes deveriam saber e serem capazes de fazer em cada nível previamente definido.

A distribuição dos escores observados dos alunos na escala de proficiência pode ser decorrente da opção pela metodologia de análises que utiliza a TRI. Por meio dessa análise, os parâmetros de proficiência dos alunos são estimados e distribuídos em uma única escala, representando o desempenho de todos os alunos de um período de ensino como, por exemplo, a educação básica, ou são distribuídos em escalas da mesma extensão, representando o desempenho dos alunos por série. Ao fazer essa distribuição são arbitrados os intervalos da escala, ou os pontos âncora. Segundo Koretz e Deibert (1993), o *National Center for Education Statistic* “descreveu os pontos âncora parcialmente em termos dos graus nos quais os conteúdos refletidos nos itens âncora são introduzidos ou pensados”. Assim, os pontos âncora são pontos na escala escolhidos na base da distribuição dos escores observados. Geralmente são apresentados intervalos de 25 ou 50 pontos (125, 150, 175, 200, ...). A caracterização do desempenho em cada ponto âncora é dada pelos itens selecionados baseados no desempenho dos estudantes em cada um dos intervalos arbitrados, com base nos seguintes critérios:

- a proporção de respostas corretas (*p-value*) a cada item dentro do grupo do intervalo tem que ser no mínimo 0,65;
- no grupo do intervalo anterior, o percentual de acertos deve ser inferior a 0,50;
- a diferença entre a proporção de acertos nos dois grupos deve ser no mínimo, 0,30;
- essas proporções de acertos têm que ser baseadas em amostras de no mínimo 100 alunos.

Desde 2001, o SAEB, sem comprometer o método de ancoragem dos itens, flexibilizou esses critérios, com o intuito de selecionar mais itens para descrever as escalas. Os itens selecionados são repassados aos painéis de especialistas. Os participantes recrutados para esses painéis devem ser especialistas nas áreas avaliadas, elaboradores de políticas públicas, professores de sala de aula e alguns membros do público geral. Eles têm a função de caracterizar os aspectos de desempenho desses itens. Suas descrições tornam a base para a descrição verbal dos pontos âncora usados nas publicações.

Quanto aos níveis de desempenho, como já foi dito anteriormente, as categorias são definidas com base no desempenho esperado para cada ciclo avaliado. Há diversos qualificativos que tentam exprimir com fidelidade as expectativas reunidas em torno de cada uma dessas categorias. Os níveis de desempenho geralmente refletem julgamentos acerca de como os estudantes deveriam desempenhar fundamentados em diversos painéis de juizes que utilizam diferentes métodos para fixarem os níveis. Assim, no Brasil, o SAEB categorizou esses níveis como: *muito crítico, crítico, intermediário, adequado e avançado* (Qualidade da educação: uma nova leitura do desempenho dos estudantes da 4ª série do Ensino Fundamental, 2003); nos Estados Unidos, os níveis foram qualificados como *básico, proficiente e avançado* (*National Education Goals Panel, 1995*); na Inglaterra, País de Gales e Irlanda do Norte, foram adotados níveis representados por unidades de 1 a 8, incluindo um primeiro nível chamado de entrada. Neste caso, eles descrevem resultados da aprendizagem, e indicam uma demanda relativa à complexidade e/ou à profundidade do desempenho e informam a autonomia do aluno demonstrada pelo seu desempenho (*Guidance for using unit level descriptors within the Qualifications and Credit Framework tests and trials, Northern Ireland Council for the Curriculum, Examinations and Assessment, 2006*).

Geralmente, o método estabelecido para a definição das categorias deve percorrer os seguintes passos (Koretz e Deibert, 1993):

- painéis de juízes dão definições simples do que os estudantes deveriam ser capazes de fazer para serem considerados como tendo alcançado cada nível;
- juízes são então convidados para estimar qual a proporção de estudantes limitada a cada nível poderia responder corretamente a itens específicos;
- essas proporções estimadas são então mapeadas na escala de escores;
- com base neste processo e em outras informações, juízes refinam e elaboram suas descrições de desempenho em cada nível;
- itens são então escolhidos baseados em uma variedade de considerações, incluindo a adequação do conteúdo e os padrões reais de desempenho, para exemplificar cada um dos níveis.

Koretz e Deibert (1993), no relato de pesquisa sobre interpretações dos pontos âncora e níveis de desempenho do NAEP pela mídia impressa em 1991 [*Interpretations of National Assessment Educational Progress (NAEP) Anchor and Achievement Levels by the Print Media in 1991*] chamam a atenção para o extremo cuidado que se deve ter com as descrições das métricas e dos níveis de desempenho baseados nos resultados das avaliações. Alguns dos pontos mais discutidos no referido artigo referem-se à redução das descrições dos pontos âncora, à descontinuidade do desempenho e à inconsistência da descrição dos níveis de desempenho.

A má interpretação dos resultados nas escalas de escores e nos níveis de desempenho pode conduzir a conclusões errôneas em relação ao que os alunos podem e deveriam fazer em relação aos ciclos de escolarização avaliados. No Brasil, por exemplo, têm sido observadas, na imprensa falada e escrita, afirmações como: “cerca de 59% dos alunos de 4ª série não dominam habilidades de leitura”, ou, “cerca de 52% dos alunos da 4ª série não identificam, ou mal identificam, um tipo de operação de soma ou subtração envolvida em um problema”. Essas afirmações falaciosas são decorrentes de má interpretação das descrições apresentadas nas categorias definidas para o SAEB, em 2001. A responsabilidade por essas descrições deveria ser compartilhada por diversos setores relacionados à área educacional, evitando-se assim, que a comunidade educacional fosse exposta a desconhecimento de gestores.

No capítulo 4 será apresentada pesquisa do processo de avaliação brasileiro sob a perspectiva do modelo norte-americano. As análises discutidas neste capítulo serão retomadas em uma proposta de aplicação concreta.

2ª PARTE

METODOLOGIA, ANÁLISE E DISCUSSÃO DOS RESULTADOS

3. MÉTODO

O tipo de pesquisa empreendida neste estudo foi descritiva, com base em investigação bibliográfica e documental. Os procedimentos metodológicos, em sua maioria, estiveram relacionados à análise de dados coletados em instituições que realizam processos de avaliação educacional em larga escala.

Pretendeu-se, por meio deste estudo, elucidar alguns problemas que permeiam o processo de avaliação educacional, em especial, os que se relacionam ao instrumento que avalia o desempenho dos alunos em avaliação em larga escala. As questões que nortearam o trabalho foram as seguintes:

- (1) Como deve ser constituído um teste que pretende avaliar o processo de construção do conhecimento dos alunos brasileiros e que informações ele pode oferecer ao sistema?
- (2) Como deve ser constituída e o que deve ser descrito em uma escala de proficiência?
- (3) Como devem ser apresentados os resultados para que sejam compreendidos por todos os agentes capazes de promover melhorias no sistema?
- (4) Que políticas têm sido implementadas com base nos resultados do SAEB?
- (5) Quais contribuições o SAEB pode fornecer para a adoção de políticas públicas que busquem melhorar a qualidade da educação básica brasileira tomando como base essas informações?

3.1. MATERIAL

A pesquisa foi realizada com base em fontes disponibilizadas pela diretoria do Sistema Nacional de Avaliação da Educação Básica e por instrumento de consulta desenvolvida por essa pesquisadora. Esses materiais são os seguintes:

1. Itens de Língua Portuguesa e Matemática de 4ª série, com análises psicométricas referentes aos anos 1995 a 2003;
2. Matrizes de Referência relacionadas aos anos 1997, 1999 e 2001;
3. Relatórios de resultados referentes aos anos 1995 a 2003;
4. Resultados das consultas feitas ao Ministério da Educação e às Secretarias de Estado de Educação sobre as políticas públicas implementadas com base nos dados do SAEB.

3.2. PROCEDIMENTOS METODOLÓGICOS PARA A ANÁLISE DOS DADOS

A seguir, são apresentados os tópicos principais que compuseram a base das análises da pesquisa:

- 3.2.1. Pesquisa dos referenciais adotados pelos sistemas norte-americano e brasileiro para proceder à avaliação, considerando-se as tendências para os testes de desempenho;
 - a. Foram apontados pontos comuns e divergentes dos propósitos e das características dos sistemas educacionais e dos marcos teóricos que fundamentam a avaliação da educação básica dos dois países;
 - b. Foram apontados pontos comuns e divergentes da metodologia, das análises e dos relatos dos resultados que particularizam o processo de avaliação da educação básica dos dois países.
- 3.2.2. Análise crítica das avaliações do SAEB (1995 a 2003) em termos das habilidades aferidas nos testes de 4ª série, de Língua Portuguesa e Matemática e publicadas na forma de escalas e níveis de desempenho;
 - a. Procedeu-se à análise da qualidade dos itens pelos índices psicométricos registrados. A qualidade foi determinada por critérios estatísticos e pedagógicos.

- b. Consideraram-se os critérios estatísticos para determinação da qualidade psicométrica dos itens fundamentados com base na literatura e na experiência dessa pesquisadora com análise de itens (Apêndice A1). Os critérios iniciais para separar os itens considerados como de boa qualidade foram os seguintes:
- correlação bisserial da alternativa correta - $r_{bis} \geq 0,30$;
 - bisserial das alternativas falsas $\leq -0,01$;
 - parâmetro a , de discriminação, da TRI - $a \geq 0,60$;
 - parâmetro $c \leq 0,35$, para itens com quatro alternativas.
- c. Em seguida, os itens selecionados como bons psicometricamente quanto os rejeitados, foram analisados do ponto de vista pedagógico, (Apêndice A2) considerando-se os seguintes critérios:
- adequação do item ao desenvolvimento cognitivo no qual o aluno de 4ª série se encontra;
 - formato do item em relação às regras de construção relacionadas ao texto, ilustração, enunciado e alternativas;
 - adequação do conteúdo avaliado pelo item;
 - no caso de item com boa qualidade psicométrica, mas qualidade pedagógica duvidosa, optou-se por não incluir o item;
- d. Após análise da qualidade psicométrica e pedagógica dos itens e seleção daqueles considerados adequados para a descrição da escala, foi realizada análise crítica de cada teste, apontando os principais problemas a eles relacionados;
- e. Em seguida a essa avaliação, verificaram-se as habilidades avaliadas por todos os itens selecionados;
- f. Foi elaborada tabela com descrição dessas habilidades e com dados sobre a qualidade psicométrica e pedagógica de todos os itens de cada disciplina avaliada de todas as edições do SAEB.
- g. O passo seguinte foi a organização dessas descrições por intervalo dos níveis de desempenho, por edição do SAEB (Apêndice A3);
- h. A escala foi descrita com base na organização das descrições por nível, agrupando-se as habilidades comuns e adequando a linguagem para a escala;
- i. Procedeu-se à transcrição das escalas de proficiência publicadas por edição ao corpo do trabalho;

- j. Compararam-se as descrições de habilidades pesquisadas com as referidas edições publicadas em que foi apontado o que os alunos brasileiros demonstraram ser capazes de fazer, com base nos dados avaliativos.

3.2.3. Análise crítica das formas de disseminação dos resultados do SAEB ao longo de todas as edições;

- a. Procedeu-se à leitura e análise crítica das publicações e disseminação dos relatórios com resultados das avaliações do SAEB, com base no referencial de outros países e nas próprias políticas implementadas. A análise crítica foi fundamentada nas seguintes categorias de análise:
 - público-alvo dos relatórios;
 - nível de compreensão dos dados relatados;
 - adequação da linguagem para os públicos-alvo dos relatórios;
 - formas de disseminação dos resultados

3.2.4. Levantamento e análise das políticas públicas implementadas na área da educação, em nível estadual e federal, com base nos resultados do SAEB;

- a. Elaborou-se instrumento para coletar informações sobre as políticas adotadas a partir dos resultados do SAEB, no Ministério da Educação e Secretarias de Educação dos Estados. (Texto enviado para a realização da consulta no Apêndice A4)
- b. O instrumento foi enviado para todas as Secretarias de Estado de Educação, para os gestores de avaliação, por meio de e-mail;
- c. Passados 30 dias, a mensagem foi reenviada aos gestores, com um novo apelo;
- d. Foi efetuada ligação telefônica para todos os responsáveis pela avaliação nas secretarias de educação de estado que não haviam retornado a resposta;
- e. Dos dados coletados foi feita uma tabulação, com o objetivo de categorizar as respostas obtidas;
- f. No Ministério da Educação a consulta foi feita pessoalmente na Secretaria de Educação Básica, no Departamento de Políticas de Educação Infantil e Ensino Fundamental e Ensino Médio com técnica responsável, que prontamente, respondeu às informações solicitadas.

4. PESQUISA DO PROCESSO DE AVALIAÇÃO BRASILEIRO SOB A PERSPECTIVA DO MODELO NORTE-AMERICANO

4.1. PROPÓSITOS PARA O ENSINO E CARACTERÍSTICAS DO SISTEMA EDUCACIONAL

4.1.1 PROPÓSITOS DO ENSINO

Como já discutido em capítulo anterior, há alguns pressupostos que devem ser observados na efetivação de um processo de avaliação. O primeiro deles diz respeito ao conhecimento dos propósitos do ensino e das características do sistema a ser avaliado (Linn, 2001).

O ensino no Brasil é regulamentado pela Lei de Diretrizes e Bases (LDB) 9.394 de 20 de dezembro de 1996 e em seu Título I, é determinado que a educação escolar deverá vincular-se ao mundo do trabalho e à prática social; e no Título II, que rege sobre os princípios e fins da educação, em seu Art. 2º, é determinado o propósito da educação. Nele, está prescrito que: “A educação é dever da família e do Estado, e que deve estar inspirada nos princípios de liberdade e nos ideais de solidariedade humana, e que deve ter por finalidade o pleno desenvolvimento do educando, seu preparo para a cidadania e sua qualificação para o trabalho”. Na LDB foi prevista a implementação legal do Plano Nacional de Educação que foi instituído na forma da Lei 10.172 em 9 de janeiro de 2001.

Neste Plano foram delimitadas as diretrizes e metas para os dez anos posteriores, em sintonia com a Declaração Mundial sobre Educação para Todos, documento elaborado no âmbito da Conferência Mundial sobre Educação para Todos, realizada em Jomtien, Tailândia,

de 5 a 9 de março de 1990, onde reuniram 155 países do mundo (Castro, 2000). As metas gerais previstas para serem atingidas até o ano 2011, em consonância com os princípios que orientam o alcance da finalidade da educação brasileira, são as seguintes:

- a elevação global do nível de escolaridade da população;
- a melhoria da qualidade do ensino em todos os níveis;
- a redução das desigualdades sociais e regionais no tocante ao acesso e à permanência, com sucesso, na educação pública; e,
- a democratização da gestão do ensino público nos estabelecimentos oficiais, obedecendo aos princípios da participação dos profissionais da educação na elaboração do projeto pedagógico da escola e a participação das comunidades escolar e local em conselhos escolares ou equivalentes.

Nos Parâmetros Curriculares Nacionais – PCN (1998), documento originado da LDB 9.394/96, foram apresentadas as linhas norteadoras para o ensino básico e que constituíram uma proposta de reorientação curricular oferecida às secretarias de educação, escolas, instituições de pesquisa e formadoras de professores e a todos os interessados na educação. Nos PCN é discutida e feita análise sobre a questão central que permeia nosso sistema educativo: educação *x* cidadania.

As discussões sobre essa questão recaem sobre as análises da conjuntura mundial que em paralelo com a conjuntura nacional, definem como papel do estado democrático facilitar o acesso à educação e torná-la instrumento fundamental na luta contra as exclusões à medida que se torna capaz de promover e integrar todos os brasileiros na prática efetiva de construção da cidadania.

Com base nas tendências mundiais, o referido documento apresenta a educação como eixo central na formação do indivíduo a qual deve tê-lo preparado para, ao longo de sua vida, enfrentar os desafios de um mundo em constante transformação.

Entre as diversas recomendações para efetivação de uma educação que seja permanente e que alcance os seus propósitos, são citados os quatro pilares que devem fundamentá-la (PCN, p. 17, 1998):

- 1) aprender a conhecer, que significa ter capacidade de aprender a aprender ao longo de toda a vida;
- 2) aprender a fazer, que pressupõe desenvolver a competência de resolver problemas, se relacionar em grupo e adquirir uma qualificação profissional;

- 3) aprender a viver com os outros, que consiste em desenvolver a compreensão do outro e a percepção das interdependências, preparando-se para gerir conflitos, respeitando a própria identidade e a dos outros, o pluralismo de idéias, a compreensão mútua e a busca pela paz;
- 4) aprender a ser, desenvolvendo melhor a personalidade, a autonomia, a expressão de opiniões próprias e assumindo responsabilidades pessoais.

Procurando seguir essas recomendações, os PCN sugerem que a educação escolar tente acolher e socializar os alunos, por meio da valorização dos conhecimentos e da forma de expressão de cada um. Para o alcance dessa proposta, a construção do conhecimento deve ocorrer respeitando-se as dúvidas e inquietações que permeiam a experiência de cada aluno e a promoção de situações de aprendizagem que tenham significado.

Nos Estados Unidos, no âmbito da educação básica, a educação é regulamentada pela “*Lei de Educação Elementar e Secundária*” (*Elementary and Secondary Education Act – ESEA*), decretada em 1964, que é re-autorizada a cada cinco anos ou menos, sob algum título específico. Além disso, existem estratégias nacionais de reforma da educação as quais são rubricadas dentro da lei. Isso inclui, por exemplo, a re-autorização denominada “Metas 2000” (*Goals 2000* - www.ed.gov/legislation/GOALS2000/TheAct) que partiu do comprometimento de diversos setores da sociedade, tanto no âmbito político como no civil, em executar a reforma educacional em torno de um conjunto de aspirações que promovessem “*escolas seguras e disciplinadas, que usassem as melhores práticas de ensino e tecnologia apropriada para assegurar que todos os estudantes fossem adequadamente preparados para enfrentar os desafios do século XXI*”. Desse esforço, foi estabelecido, em 1992, o Plano Nacional para a Educação Americana (“*Educate America Act*”), que foi constituído por seis metas, com a adição de outras duas incluídas mais tarde.

Esse Plano teve forte impacto no sistema nacional de avaliação e estabelecia as seguintes oito metas:

- (1) toda criança da América iniciaria a escola pronta para aprender;
- (2) a taxa de graduação da escola secundária (*high school*) aumentaria para no mínimo 90%;

- 3) todos os estudantes ao findarem os graus² 4, 8 e 12 (idades 10, 14 e 18 respectivamente) teriam que demonstrar competência ao enfrentar desafios relacionados às disciplinas estudadas; e cada escola na América asseguraria que todos os estudantes aprendessem a usar suas mentes bem (raciocínio), pois assim, estariam preparados para uma cidadania responsável, favorecendo o conhecimento e o trabalho produtivo na moderna economia da nação;
- (4) o ensino da nação determinaria o acesso a programas para a melhoria continuada de habilidades profissionais e a oportunidade de adquirir o conhecimento e as habilidades necessárias para instruir e preparar todos os estudantes americanos para o próximo século;
- (5) os estudantes dos Estados Unidos seriam os primeiros do mundo no desempenho em matemática e ciências;
- (6) todo adulto americano seria alfabetizado e desenvolveria conhecimento e habilidades necessárias para competir em uma economia global e exerceria os direitos e responsabilidades de cidadania;
- (7) cada escola nos Estados Unidos estaria livre de drogas, violência e a presença não autorizada de armas de fogo e álcool, e ofereceria um ambiente disciplinado e favorável à aprendizagem;
- (8) cada escola estimularia a sociedade a aumentar o envolvimento e a participação dos pais na promoção do crescimento social, emocional e acadêmico das crianças.

O estabelecimento dessas metas foi visto como um passo crítico no desenvolvimento de um sistema com responsabilidade integrada nacionalmente com uma forte ênfase sobre resultados. Gestores de políticas acreditam que as escolas têm se tornado cada vez mais dominadas por regras e procedimentos demandados por distritos e estados. O cenário das “metas nacionais”, especialmente aquelas relacionadas à participação, retenção e resultados, em termos de desempenho dos estudantes, foi programado a aumentar a autonomia profissional e gerencial das escolas, faculdades e universidades colocando o foco no produto-fim: nos resultados obtidos.

Em 1999, houve nova re-autorização do *ESEA*, que teve como temática *Educational Excellence for All Children Act* (Excelência Educacional para Todas as Crianças). Com essa re-autorização foi enfatizada a necessidade de propor a tentativa de reforma do estado para a

² Os graus 4, 8 e 12 do sistema educacional norte-americano correspondem às séries 4ª e 8ª EF e 3ª série EM em nosso sistema educacional.

escola, expandindo esforços para melhorar a qualidade do professor e assegurar responsabilidade para resultados (U. S. Department of Education: *Educational Excellence for All Children Act of 1999*).

Em janeiro de 2001, com a eleição do Presidente George W. Bush, iniciou-se, novamente, mais um processo de regulamentação de uma reforma estratégica para a educação que redundou em outra re-autorização do *ESEA*, conhecida como “*No Child Left Behind Act*” (Nenhuma Criança Deixada à Margem), apelidada Bill, foi formalmente aprovada em dezembro de 2001 e colocada em vigor a partir de janeiro de 2002. Entre outras provisões, essa Lei prevê:

- (1) aumento da testagem nacional regulamentada;
- (2) melhoria das escolas fracassadas;
- (3) garantia de que até o final do ano escolar 2005-06, todos os professores em escolas públicas seriam altamente qualificados;
- (4) aumento da despesa em programas de alfabetização;

Além disso, por meio dessa Lei, fixou-se uma meta de 12 anos para melhorar a proficiência acadêmica daqueles estudantes que são pobres, ou que tenham deficiências físicas, ou na linguagem inglesa (<http://www.whitehouse.gov/infocus/earlychildhood>: Education: The Promise of America).

Esse histórico mais recente que define os propósitos atuais da educação americana mostra o interesse e o empenho tanto do governo como da sociedade civil em superar os problemas que impactam o desempenho dos alunos e melhorá-lo.

Comparando-se os propósitos do ensino do Brasil e dos Estados Unidos, destacam-se quatro pontos nos quais as diferenças ficam evidentes. O primeiro diz respeito ao foco da educação americana: os resultados dos alunos; o segundo, diz respeito à estreita relação entre os propósitos do ensino e a avaliação institucional e externa da educação. O terceiro ponto relaciona-se à preocupação americana em atualizar, de forma prática e regulamentada por lei, os objetivos instrucionais, focando-os sobre as necessidades que o período histórico exige; o quarto, diz respeito ao orgulho e à cidadania do povo americano, que parece apoiar, participar, discutir e se comprometer com políticas implementadas com a finalidade de elevar e superar o nível de qualidade da educação.

Nos próprios Estados Unidos têm sido feitas críticas freqüentes quanto à associação da melhoria da educação aos resultados. Sabe-se que os resultados devem estar relacionados à

conseqüência dos propósitos e não devem se constituir no propósito em si. Inúmeros fatores propiciam ou não melhores resultados. E melhores resultados do desempenho estão relacionados à superação desses fatores intervenientes no processo educacional.

No Brasil, verifica-se que embora tenham leis e um plano de superar as dificuldades encontradas na área da educação, há pouca participação da sociedade na busca de superação dos problemas. Percebe-se que quase não há diálogo entre as instituições que poderiam se responsabilizar e empenhar pela melhoria da qualidade da educação. Por exemplo, não se observa comunicação entre a instituição que cuida da avaliação da educação básica no país, a secretaria responsável por esse nível de ensino (Secretaria de Educação Básica) e as instituições formadoras de docentes desse nível de ensino (universidades). Esse diálogo poderia redundar em esforços e resultados mais eficazes.

Há leis no Brasil que regulamentam questões fundamentais do âmbito da implementação de uma educação de qualidade, mas constata-se que na prática há uma dificuldade de comunicar e disseminar essas ações. Foi executado por essa pesquisadora, no período de 2004 a 2006, um trabalho de coordenação e execução de oficinas pedagógicas para apresentação de resultados de avaliações e elaboração de itens de testes de desempenho em oito capitais do Brasil e onze cidades de diversas regiões do Estado do Rio de Janeiro. Houve intenso contato com professores de sala de aula que lecionam para a 4ª, 8ª e 3ª séries, e constatou-se que somente cerca de 10% dos professores conhecem os PCN que dão novas orientações curriculares e pedagógicas para o ensino brasileiro. Considerando-se a dimensão do país e a transformação que a educação pode promover, verificam-se muito poucas iniciativas com intuito de reciclar e atualizar a prática pedagógica dos professores. Evidencia-se, que a maioria fica no âmbito das relações humanas.

Quando se relaciona o sistema avaliativo e os propósitos do ensino, observa-se que ao planejarem-se as avaliações, não são considerados os propósitos do ensino, pois os instrumentos de avaliação adotam, ainda, uma perspectiva conteudista, distanciando-se da finalidade maior da educação, que é o pleno desenvolvimento do educando, seu preparo para a cidadania e sua qualificação para o trabalho. Se os propósitos fossem considerados, a constituição dos instrumentos refletiria esse teor. Os itens dos testes estariam abordando tarefas que correspondessem à aplicação prática dos conteúdos aprendidos na escola.

Em relação aos propósitos do ensino, e considerando-se o momento histórico que estamos vivendo, conclui-se que o Brasil encontra-se muito distante de atingir o “pleno desenvolvimento do educando, seu preparo para a cidadania e sua qualificação para o

trabalho”. Essa conclusão é embasada não só na comparação com a sociedade norte-americana, mas comparada a países vizinhos e com dificuldades econômicas semelhantes às do nosso país. Países como Argentina, Uruguai e Chile têm avançado no alcance de uma educação melhor qualificada e que prepare, de fato, o aluno para o exercício pleno da cidadania. Esses países têm se atualizado nos objetivos instrucionais e programas de formação de professores que os preparam para a atual fase que estamos vivendo, de tecnologia avançada e de rapidez das informações (ver Quadro da situação da Sub-Região MERCOSUL - Políticas e Estratégias para a Prevenção do Fracasso Escolar - Organização dos Estados Americanos (OEA)/Agência Interamericana para a Cooperação e o Desenvolvimento (AICD) / Ministério da Educação, Ciência e Tecnologia/Instituto Internacional de Planejamento Educacional IPE/Unesco/Buenos Aires - <http://tq.educ.ar/fracasoescolar/documentos/mercosurPORT.doc>)

4.1.2 CARACTERÍSTICAS DO SISTEMA EDUCACIONAL

Os dois países têm algumas semelhanças e divergências no que diz respeito ao que constitui características do sistema de ensino: a estrutura do sistema, as orientações curriculares e a formação dos professores. A estrutura da educação básica no Brasil *é formada pela educação infantil, ensino fundamental e ensino médio* e tem por finalidades *desenvolver o educando, assegurando-lhe a formação comum, indispensável para o exercício da cidadania e fornecer-lhe meio para progredir no trabalho e em estudos posteriores* (Art. 21 e Art. 22 da LDB 9.394/96).

A partir desse eixo, foi dada autonomia aos estados para delinearem a estrutura que melhor lhes conviesse. Assim, há estados que adotam regime seriado (1º a 9º anos do ensino fundamental e 1ª a 3ª séries do ensino médio e outros que adotam o regime de ciclos (quatro ciclos para o ensino fundamental e um ciclo para o ensino médio). A responsabilidade pela educação é exercida, em parte, pela legislatura de cada um dos estados, e dentro de cada estado, o funcionamento das escolas públicas é monitorado pelo estado e pelos municípios, que da mesma forma que os estados têm autonomia para definir a estrutura que melhor convier. O Art. 24 estabelece como carga horária mínima para a educação básica, nos níveis fundamental e médio, 800 horas distribuídas em um mínimo de 200 dias de efetivo trabalho escolar. O ano letivo no Brasil inicia, normalmente em fevereiro e se encerra em dezembro, com um período de férias entre os dois semestres.

Quanto à alocação de fundos para financiamento da educação, foi definido no Art. 69 da LDB, que a União aplicaria anualmente no mínimo 18% da receita e os Estados, o Distrito Federal e os Municípios, no mínimo, 25% de suas receitas.

Em relação ao currículo aplicado nas escolas, a orientação é dada pelas Diretrizes Curriculares Nacionais, que são definidas como “*o conjunto de definições doutrinárias sobre princípios, fundamentos e procedimentos na Educação Básica, (...) que orientarão as escolas brasileiras dos sistemas de ensino, na organização, na articulação, no desenvolvimento e na avaliação de suas propostas pedagógicas*” (Diretrizes Curriculares Nacionais). Citando Bonamino e Martinez (2002), p. 371, *o federalismo constitui uma dimensão significativa e está presente no “reconhecimento pelo Conselho Nacional de Educação (CNE) da competência da Câmara de Educação Básica para exercer função deliberativa sobre as Diretrizes Curriculares Nacionais, mas reservando para os entes federativos e para as escolas as tarefas de plenificação e implementação curriculares.*

Nessa perspectiva, pela LDB 9.394/96, Art.26, é assegurado que todos os níveis de ensino devem ter uma base nacional comum, a ser complementada em cada sistema de ensino e estabelecimento escolar, e uma parte diversificada, exigida pelas características regionais e locais da sociedade, da cultura, da economia e da clientela. Embora não haja nenhuma lei federal que prescreva um padrão curricular, os programas de educação, geralmente incluem as mesmas disciplinas e conteúdos curriculares, variando conforme os níveis de ensino.

O documento “Parâmetros Curriculares Nacionais” é oficialmente o orientador da prática pedagógica nas escolas. Esse documento adota como concepção de ensino e aprendizagem a perspectiva construtivista que toma por base a “*teoria genética* de Jean Piaget, a “*teoria da atividade*, nas formulações de Vygotsky, Luria e Leontiev e cols, a “*psicologia cultural*”, fundamentada nos trabalhos de M. Cole e K. Wakai e as “*aprendizagens significativas*”, de Ausubel. Esse documento apresenta idéias modernas e orientações curriculares que atenderiam as demandas da educação na atualidade.

Quanto à formação dos profissionais da educação, há sete artigos da Lei 9.394/96 que a regulamenta (do Art. 61 ao Art. 67). Esses artigos determinam como formação mínima dos docentes que atuam na educação infantil e nas quatro primeiras séries iniciais da educação básica, a formação oferecida em nível médio, na modalidade Normal, e para o exercício docente nas séries posteriores, a formação em nível superior, em curso de licenciatura ou de graduação plena, em universidades e institutos superiores de educação.

Nos Estados Unidos, a estrutura da educação é determinada tanto pelo estado quanto pelo local (distrito) (*National Center for Education Statistics, 1999: The Condition of Education*). Os estados formulam políticas sobre a alocação de fundos, a certificação de professores, os livros didáticos e os serviços de bibliotecas, e a provisão de registros e estatística. Alguns estados também especificam currículos. Comissões de diretores de escolas distritais se responsabilizam pela coleta de taxas, construções de edifícios, e têm tradicionalmente, adquirido equipamentos, determinado políticas de formação dos professores e empregado professores e outros profissionais relacionados à área de educação (*Education Week, 2002; Cambron-McCabe, 2002*).

Quanto ao ano letivo, embora algumas decisões em relação ao calendário escolar sejam de atribuição das comissões de direção locais, cada estado tem um período letivo mínimo a ser cumprido, geralmente expresso em dias. O padrão comum é o cumprimento de 180 dias letivos por ano, distribuídos em uma carga horária média de 1.010 horas, dadas em 5 dias por semana (segunda a sexta-feira).

Em relação ao financiamento da educação americana, a contribuição dos estados e distritos é de cerca de 91%, cabendo à instância federal apenas 7%, e 2% a empresas privadas. Cerca de 91% dos alunos na faixa de idade entre 5 e 17 anos são alunos de escolas públicas.

Nos Estados Unidos, no que diz respeito ao currículo praticado nas escolas, não há uma orientação nacional para nenhum dos níveis considerados da educação básica americana. São os estados que têm o direito e o dever de estabelecer padrões específicos ou orientações curriculares características do próprio estado para cada uma das fases. Como no Brasil, embora não haja nenhuma lei federal que prescreva um padrão curricular, os programas de educação dos 50 estados, geralmente incluem as mesmas disciplinas e conteúdos curriculares. As intenções dos currículos são determinadas no nível dos distritos escolares de acordo com as orientações dos estados. Todos os 50 estados têm algum tipo de política de testagem local relacionado de alguma forma a documentos curriculares oficiais e padrões específicos de aprendizagem centralizados no inglês, matemática e ciências. Tem havido no país um movimento para cumprir o que determina a Lei “*No Child Left Behind*”, no sentido de desenvolver, em nível nacional, padrões de conteúdos centrais, os quais proporcionem consistência maior em termos de conteúdo curricular. Quase todos os estados têm os padrões locais, embora eles variem em termos de especificidade, áreas cobertas, formatos, etc. (Doppen & Yeager, 1998; Cooper, 1992).

A formação dos professores americanos não é regulamentada por lei. Mais de 1.000 instituições de ensino preparam futuros professores e não existe nenhuma entidade oficial que supervisione a execução desses programas. Esse sistema descentralizado, ao mesmo tempo em que permite flexibilidade e correspondência às necessidades locais, tem se tornado um foco de preocupação para gestores de políticas que objetivam fortalecer a formação dos professores (*National Council on Teacher Quality and US Department Of Education, International Affairs Office, 2004*). Até o grau 6, correspondente à idade em torno de 11 – 12 anos, os professores apresentam uma formação generalista, embora alguns, em especial nas áreas de educação física, artes e ciências, tenham uma formação especializada. Atualmente, entretanto, embora sem regulamentação legal, a maioria dos estados tem requerido que as instituições de formação de professores apresentem certificado de acreditação fornecido pelo *National Council for Accreditation of Teacher Education – NCATE* (Conselho Nacional para Acreditação da Educação de Professores), ou que tenham adotado os padrões do *NCATE* nos programas de formação de professores.

Como pode ser observado pela exposição de algumas características dos dois sistemas de ensino podem-se constatar mais similitudes do que diferenças entre os dois países. Semelhanças e diferenças essas que não cabe aqui discuti-las ou criticar.

4.2 SISTEMA DE AVALIAÇÃO

Como já foi exposto anteriormente, a avaliação brasileira foi regulamentada pela Lei 9.394/96, que em seu Art. 9º, inciso VI, é assegurado *o processo nacional de avaliação do rendimento escolar no ensino fundamental, médio e superior, em colaboração com os sistemas de ensino, objetivando a definição de prioridades e a melhoria da qualidade do ensino*. O Ministério da Educação e Cultura que foi o responsável pela avaliação da educação básica no Brasil a partir de 1990, em 1997 atribuiu, definitivamente ao Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira o gerenciamento do Sistema Nacional de Avaliação da Educação Básica, pela Diretoria de Avaliação da Educação Básica (INEP/DAEB/SAEB). De 1990 até 2003, o sistema avaliativo realizou pesquisa amostral, em escolas de todas as instâncias representativas de todo o país, com uma regularidade bienal, com exceção da primeira edição que ocorreu com diferença de três anos.

Embora essa pesquisa esteja abordando o formato amostral, pois até o início desse estudo essa era a estrutura da avaliação no Brasil, o sistema atual alterou sua estrutura em

2005, por força de três Portarias. A Portaria Nº 931, de 21 de março de 2005, instituiu o Sistema de Avaliação da Educação Básica - SAEB e sua composição em dois processos de avaliação: a Avaliação Nacional da Educação Básica - ANEB, e a Avaliação Nacional do Rendimento Escolar – ANRESC; a Portaria Nº 69, de 4 de maio de 2005 - estabeleceu a realização da ANRESC em 2005 e a Portaria Nº 89, de 25 de maio de 2005 - estabeleceu a realização da ANEB em 2005.

A ANRESC, executada em 2005, teve por objetivo avaliar alunos em Língua Portuguesa, foco Leitura, e Matemática, foco Resolução de Problemas, de todas as escolas públicas urbanas que tivessem 30 ou mais alunos matriculados nas 4ª e/ou 8ª séries do ensino fundamental, das Unidades da Federação que concordassem em participar. Com essa avaliação, foram avaliados 3.304.350 alunos matriculados nas 4ª e 8ª séries, do EF. Esse quantitativo representa³ cerca de 9,8% dos alunos matriculados no ensino fundamental da rede pública e 46,6% dos alunos matriculados nessas respectivas séries (Censo Escolar 2005). Os alunos avaliados estavam distribuídos em 122.463 turmas, de 40.920 escolas⁴, representando cerca de 28% das escolas públicas que oferecem ensino fundamental (Censo Escolar 2005) do país, de 5.398 municípios.

A segunda avaliação, com população amostral e com características semelhantes às avaliações anteriores, a ANEB, continuou tendo por meta avaliar a educação básica brasileira, tendo abrangido, de forma amostral, 194.822 alunos de escolas públicas e particulares, matriculados na 4ª e 8ª séries do ensino fundamental e na 3ª série do ensino médio de 5.940 escolas de todas as unidades da federação. Essa avaliação continuou garantindo a série histórica da avaliação da educação básica brasileira.

A avaliação ANRESC, dado seu caráter universal, tem sido chamada pelo INEP “Prova Brasil”. Ela tem como foco cada unidade escolar. Assim, a Prova Brasil foi idealizada para produzir informações sobre o ensino oferecido por município e escola, individualmente, com o objetivo de auxiliar os governantes nas decisões e no direcionamento de recursos técnicos e financeiros, assim como a comunidade escolar no estabelecimento de metas e implantação de ações pedagógicas e administrativas, visando à melhoria da qualidade do ensino (<http://www.inep.gov.br/basica/saeb/anresc.htm>). A ANEB continua com o objetivo de

³ Segundo o Censo Escolar 2005 houve 33 534 561 alunos matriculados no ensino fundamental em 2005. Desses alunos, 4 040 704 encontravam-se matriculados na 4ª série e 3 053 168, na 8ª série.

⁴ Ainda segundo o Censo Escolar 2005, o número de estabelecimentos que oferecem o ensino fundamental é 162 727. Desses, 19 096 são estabelecimentos privados.

oferecer subsídios para a formulação, reformulação e monitoramento de políticas públicas, contribuindo, dessa maneira, para a ampliação da qualidade do ensino brasileiro.

As duas avaliações abrangeram questões elaboradas com base no que está previsto nas Matrizes Curriculares do SAEB, para as séries avaliadas nos currículos de todas as unidades da federação seguindo as recomendações dos Parâmetros Curriculares Nacionais. Além dos testes, os alunos que fizeram ANEB, responderam a um questionário que coletou informações sobre seu contexto social, econômico e cultural. Compôs ainda a avaliação a coleta de dados por meio de questionários destinados ao professor, diretor e escola.

Desenvolvido pelo INEP, autarquia do Ministério da Educação (MEC), o SAEB foi a primeira iniciativa brasileira, em âmbito nacional, no sentido de conhecer mais profundamente o sistema educacional nacional. Além de coletar dados sobre a qualidade da 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 respondidos 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.

As informações obtidas a partir dos levantamentos do SAEB também permitem acompanhar a evolução da qualidade da educação ao longo dos anos. Segundo publicações do INEP, as informações são utilizadas principalmente pelo MEC e pelas Secretarias Estaduais e Municipais de Educação na definição de ações voltadas para a solução dos problemas identificados, assim como no direcionamento dos seus recursos técnicos e financeiros às áreas prioritárias, com vistas ao desenvolvimento do sistema educacional brasileiro e à redução das desigualdades nele existentes.

O SAEB enquanto pesquisa amostral, procede à constituição de sua amostra por meio de informações das escolas cadastradas no Censo Escolar, as quais são separadas por estratos, a fim de que sejam produzidos resultados de desempenho generalizáveis não só para o todo do sistema educacional brasileiro, mas também para grupos específicos de alunos, chamados estratos de interesse.

Desta forma, os principais critérios para estratificação têm sido:

- série em que o aluno está matriculado e que são as de interesse do SAEB: 4ª e 8ª do Ensino Fundamental e 3ª do Ensino Médio, na modalidade regular;
- as Unidades da Federação (Estados), pois o SAEB produz resultados para cada um dos 26 Estados e para o DF

- dentro de cada UF as dependências administrativas a que as escolas pertencem: pública (estadual ou municipal) e particular;
- a localização da escola, isto é, se fica na capital (ou regiões metropolitanas) ou no interior, se é urbana ou rural;
- o tamanho da escola, calculado pelo número de turmas que a escola mantém nas séries avaliadas;

Dentro desses estratos, são sorteadas escolas e nelas turmas de alunos que farão os testes. Em cada escola é sorteada no mínimo uma e no máximo duas turmas de uma mesma série a ser avaliada.

Dessa forma, pela pesquisa amostral, não são obtidos resultados que reflitam a eficiência de cada escola isoladamente, porque as turmas sorteadas, na maioria das vezes, não são quantitativa nem qualitativamente representativas da realidade daquela escola, mas apenas em conjunto com outras turmas de outras escolas, o estrato que representam será refletido.

Quanto a políticas de testagem, o SAEB tem empreendido grande esforço para promover a disseminação da cultura da avaliação no país. Em outubro de 2005, em pesquisa realizada por essa pesquisadora, solicitada pela coordenação do Banco Nacional de Itens (BNI), havia 19 estados do Brasil que já tinha adotado algum sistema de avaliação. A maioria deles seguindo o modelo do SAEB, inclusive utilizando itens para proceder à comparação com a avaliação nacional.

A avaliação nos Estados Unidos⁵ é primeiramente responsabilidade do estado e dos distritos locais. A maioria dos estados fixa padrões designando o que e o quão bem os estudantes devem saber. Todos os estados têm desenvolvida alguma política de testagem que deve fornecer informações acerca do processo de aprendizagem dos alunos. Em todas as avaliações é especificado, na forma de padrões o que os alunos devem saber nas áreas essenciais de matemática, leitura, escrita, ciências e cidadania. Atualmente, a política nacional de avaliação, sob a tutela da Lei *No Child Left Behind*, tem focado sobre a escola. E para que a escola seja alcançada, o governo tem dado ainda mais autonomia aos estados e distritos locais. Os resultados das avaliações têm estado disponíveis para toda a comunidade. Os resultados dos alunos, quando considerados unidades de análise, somente são fornecidos à escola e aos pais. Neste caso, as avaliações estaduais e distritais devem produzir relatórios para cada aluno.

⁵ Todas as informações sobre as características do sistema de avaliação do NAEP foram extraídas do National Center for Education Statistics, 2005: *The Nation's Report Card* ; e 2001: *The NAEP 1998 – Technical Report*

A Lei “*No Child Left Behind*” determinou que no período do ano escolar 2005-06, cada estado devesse medir o progresso de cada aluno, em leitura e matemática, em cada um dos graus do ensino (elementar e baixo secundário, correspondentes ao ensino fundamental brasileiro), ou seja, do grau 3 ao 9, e no mínimo uma vez durante os graus 10 ao 12. Este sistema de avaliação deve atingir todos os alunos, mesmo aqueles com dificuldade de aprendizagem por causa de problemas com a língua (estrangeiros), ou deficiências físicas ou mentais. Perante a lei, todos têm algum potencial para progredir e esse progresso deve ser mostrado.

Foram também propostas medidas que incluem fundos adicionais para melhoria de escolas fracassadas, e assistência federal para tutorias, programas de verão, ou a provisão de outros serviços educacionais para crianças de escolas fracassadas. Além disso, desde setembro de 2002, tem sido requerido das escolas identificadas como escolas de baixo desempenho, pagar os custos de transporte para os estudantes que escolherem estar presentes em escolas com desempenho mais alto dentro do distrito.

Foram estimulados programas tais como, “Primeira Leitura” (*Reading First*) – um empenho focalizado em toda a nação para capacitar todos os estudantes a se tornarem, em breve, leitores bem sucedidos. Por meio dos programas, fundos federais seriam proporcionados para assistir a estados e distritos escolares no estabelecimento de instruções de leitura abrangente de alta qualidade, do jardim de infância até o grau 3 (de 5 anos até 8/9 anos de idade).

Os padrões acadêmicos de ensino determinados pelos estados para cada ciclo de escolarização são elaborados de tal forma, que eles devem ser referência para o ensino e a avaliação. Desse modo, fica amenizada a questão de os professores ensinarem para que os alunos se saiam bem nos testes, pois na verdade, o que o professor está ensinando é o que poderá ser avaliado. O foco central das avaliações estaduais e locais é avaliar se os alunos estão aprendendo o que foi referendado. Pois todas as políticas serão localizadas para promover o papel principal da escola, ou seja, ela deve exercer o papel de ensinar e promover a aprendizagem dos alunos.

Esse modelo descentralizador de avaliação proposto atualmente nos Estados Unidos tem por meta subsidiar professores, diretores e pais para promoverem melhoria na qualidade do ensino oferecido a todas as crianças. Assim, os professores recebem grandes quantidades de informação a respeito da testagem anual. Por exemplo, quando ocorrem resultados muito baixos no desempenho, esses podem ser atribuídos aos seguintes indicadores: (1) o currículo

precisa ser revisto e alinhado com os conteúdos sob os quais estão baseados os padrões de ensino do estado; (2) os professores precisam modificar seus métodos de ensino; (3) o professor não tem domínio em relação a determinados conteúdos relacionados à disciplina a qual leciona; (4) problemas relacionados a crianças individualmente são detectados e a escola deve resolvê-los ou encaminhá-las. De posse do conhecimento desses indicadores, a chance de serem sanadas diretamente as dificuldades e de realmente melhorar a qualidade do ensino são aumentadas. Com base nesses indicadores localizados, políticas podem ser implementadas com o objetivo de superar essas dificuldades.

Os diretores das escolas, por sua vez, usam esses indicadores para orientar ações acerca dos métodos de ensino, da organização curricular, da formação dos professores, dos recursos materiais a serem solicitados de acordo com as necessidades específicas da escola, enfim, eles terão subsídios para implementarem uma gestão mais eficiente.

Os pais, da mesma forma, também se sentem munidos de informações seguras e confiáveis acerca do desempenho de seus filhos, pois têm um referencial público (padrões) do que os seus filhos devem saber em cada grau estudado. O conhecimento desses dados é coadjuvante no processo de ensino, acompanhando o desempenho da gestão do diretor, do professor e do filho, dando abertura a mudanças no método de ensino, no programa curricular, nos sistemas de estudo, entre outros relacionados à própria criança.

Além das avaliações localizadas no nível dos estados ou dos distritos locais, a avaliação na qual esse estudo toma como referência é a avaliação que ocorre em nível nacional e é chamada *National Assessment Educational Progress* – NAEP – (Avaliação Nacional do Progresso Educacional). Esta é uma avaliação nacionalmente representativa e contínua que informa o que os estudantes americanos sabem e podem fazer em várias disciplinas. A primeira avaliação ocorreu em 1969, embora o NAEP tenha sido criado e regulamentado por lei em 1964. Desde 1969, as avaliações têm sido conduzidas periodicamente.

O NAEP é regulamentado por lei e é executado pelo Centro Nacional de Estatística em Educação (*NCES – National Center for Education Statistic*), o qual atribui o seu gerenciamento à Diretoria de Educação Estatística. O Comitê Governamental de Avaliação Nacional (*NAGB – National Assessment Governing Board*), nomeado pela Secretaria de Educação, mas independente do Departamento, ajusta políticas para o NAEP e é responsável por desenvolver a estrutura e as especificações dos testes que servem como suporte para as avaliações. O NAGB foi criado em 1988 e é um grupo bipartidário formado por membros que

incluem governadores, legisladores, representantes de escolas locais e estaduais, educadores, empresários e membros do público em geral.

O NAEP, como o SAEB, não fornece escores para estudantes individuais ou escolas; ao invés disso, ele oferece resultados relacionados: (a) ao desempenho dos alunos nas disciplinas; (b) à experiências instrucionais; e, (c) ao ambiente escolar para populações de estudantes (por exemplo, alunos do 4ª grau) e grupos dentro dessas populações, como estudantes femininos, hispânicos, entre outros. Os resultados do NAEP estão sempre associados a amostras de estudantes de populações de interesse.

O NAEP realiza um grande esforço para proporcionar medidas de desempenho acadêmico justas e acuradas. Nessa missão, o NAEP:

- informa ao público acerca das tendências relacionadas aos conteúdos aprendidos, decorrentes do processo de ensino;
- informa ao público acerca da proficiência dos estudantes em níveis nacionais e estaduais;
- fornece dados para comparação dos níveis de desempenho dos estudantes através dos estados;
- informa elaboradores de políticas acerca dos relacionamentos entre o desempenho dos estudantes e as variáveis de contexto;
- oferece dados comparativos sobre o desempenho de estados, regiões e grupos.

O NAEP proporciona, além do relatório nacional, informações descritivas acerca do desempenho para o público em geral, para os elaboradores de políticas e para os professores. A avaliação mede os resultados do ensino refletidos no comportamento dos alunos na resolução dos itens dos testes.

O NAEP mede o desempenho de estudantes capturando informações por meio da amostragem de estudantes. A amostra de estudantes é determinada por meio de uma estratificação randômica de escolas de toda a nação e em estados participantes.

Observam-se muitas diferenças relativas aos sistemas brasileiro e americano, principalmente, no momento atual de implementação no Brasil de uma política de avaliação considerada “quase universal”. Verifica-se que nos Estados Unidos há uma exigência legal para que todos os estados adotem a sua política de avaliação e há um incentivo grande para que ocorra também nos distritos locais. No Brasil, como nos Estados Unidos, é dada autonomia aos estados para deliberarem sobre sua estrutura de ensino e orientações

curriculares, dessa forma, não há muito sentido de a avaliação de escolas públicas brasileiras, no âmbito dos estados e municípios, ser administrada pela esfera federal. Cada estado tem as suas peculiaridades. Os testes desenvolvidos pelos próprios estados refletiriam muito mais a sua realidade e teriam muito mais poder de implementação de políticas que favorecessem a superação das dificuldades. Além disso, considerando-se o caráter de “quase universalidade”, que o governo brasileiro gostaria de alcançar, os próprios estados teriam mais possibilidade de atingir e avaliar todas as instâncias, incluindo a zona rural e pequenas escolas que não têm 30 alunos na 4ª ou 8ª séries, e que representam grande parte do quantitativo de escolas de regiões em que o desempenho se apresenta mais baixo (Região Norte e Nordeste). Dessa forma, o modelo norte-americano descentralizado, que toma como unidade de análise a escola, cuja responsabilidade pela avaliação recai sobre o estado ou o distrito local, parece ser muito mais eficiente que o brasileiro. E o sistema nacional de avaliação, salvaguardadas algumas diferenças com o sistema americano, que serão discutidas a seguir, deveria seguir o seu curso.

4.2.1 O QUE REFERENCIAM AS AVALIAÇÕES NOS DOIS PAÍSES?

No Brasil a construção dos testes do SAEB tem por base as Matrizes Curriculares de Referência (Anexo 1), que é um documento no qual estão descritas a fundamentação teórica e as orientações para a elaboração dos itens dos testes do SAEB. É importante pontuar que esses testes têm alcance nacional, sendo, portanto, fundamental que eles sejam orientados pelo estabelecimento dos conteúdos desejáveis e necessários às demandas e exigências implícitas no sistema educacional brasileiro, considerando-se todas as diferenças regionais.

Todas as matrizes curriculares foram desenvolvidas a partir de uma ampla consulta nacional e consensual sobre os conteúdos praticados nas escolas brasileiras de Ensino Fundamental e Médio e da reflexão de professores, pesquisadores e especialistas sobre a produção científica em cada área que se torna objeto de conhecimento escolar. Na última matriz publicada (2002) constam os conteúdos estabelecidos, hierarquizados e distribuídos em três ciclos, com terminalidade na 4ª e 8ª séries do Ensino Fundamental e na 3ª série do Ensino Médio. A partir da avaliação de 2001, têm sido contemplados os conteúdos relacionados à Leitura e à Resolução de Problemas, em Matemática.

A Matriz Curricular de Referência para cada área avaliada, ficou estruturada da seguinte forma: na dimensão conteúdos foram expostos os grandes temas e tópicos e/ou assuntos relacionados a cada disciplina e série e na outra dimensão foram expostos os

descritores que representam competências que devem ser alcançadas pelos alunos ao final de cada ciclo de escolarização.

O modelo teórico adotado (SAEB 2001 – Novas Perspectivas, 2002) apresenta as matrizes como uma associação dos conteúdos às competências cognitivas utilizadas no processo de construção do conhecimento. O conceito de competência foi adaptado de Phillippe Perrenoud (1993) que a considera como “a capacidade de agir eficazmente em um determinado tipo de situação, apoiando-se em conhecimentos, mas sem se limitar a eles”. Para enfrentar uma situação, geralmente, colocam-se em ação vários recursos cognitivos complementares, entre os quais os conhecimentos. Perrenoud (1993) afirma, ainda, que “quase toda ação mobiliza alguns conhecimentos, algumas vezes elementares e esparsos, outras vezes complexos e organizados em rede” (p.11).

Assim, no documento Novas Perspectivas, p. 11, competências cognitivas foram definidas como “as diferentes modalidades estruturais da inteligência que compreendem determinadas operações que o sujeito utiliza para estabelecer relações com e entre os objetos físicos, conceitos, situações, fenômenos e pessoas. As habilidades instrumentais referem-se especificamente ao plano do saber fazer e decorrem, diretamente, do nível estrutural das competências já adquiridas e que se transformam em habilidades”.

Segundo o mesmo documento, o processo de construção do conhecimento passa, necessariamente, pelo saber fazer antes de ser possível compreender e explicar, até que este processo contínuo termine numa fase posterior do desenvolvimento com a tomada de consciência dos instrumentos utilizados e das relações estabelecidas. Pode-se dizer que o processo de conhecer comporta um ciclo, pois a compreensão e a tomada de consciência dos instrumentos e das relações estabelecidas em um nível influenciam o fazer no nível seguinte.

A formulação e a utilização de descritores na matriz, como base para a construção dos itens de teste, e que são utilizados para avaliar o desempenho dos alunos, têm como pressuposto epistemológico o fato de que os conteúdos científicos, matemáticos, lingüísticos, históricos, etc., se constituem de princípios, conceitos e informações relacionadas por operações intelectuais (classificação, seriação, correspondência, causa e efeito, correlação, implicação etc.).

No documento consta, ainda, outro pressuposto central, do ponto de vista teórico, implícito nessa fundamentação, é o que concebe a construção do conhecimento por meio de

mudanças qualitativas que abrem novas possibilidades de interagir com objetos do conhecimento cada vez mais complexos, abrangentes e abstratos.

Foi levado em conta, ao elaborar as matrizes, que os alunos avaliados apresentam diferentes faixas etárias e escolaridade (p. 12). Assim, os conteúdos, competências e habilidades foram, também, diferenciados para que se pudesse detectar o que o aluno sabe (resolvendo os itens do teste) em função das etapas próprias do processo de seu desenvolvimento.

No documento, conforme a opção teórica, de natureza cognitivista, adotada para a construção dos descritores, é priorizada a avaliação de conteúdos na perspectiva das competências e habilidades, as quais trazem implícito o ato de raciocínio, de coordenação das informações em consonância com as questões propostas e, eventualmente, de produção de novas informações significativas e realização de inferências quando necessário. Os pressupostos básicos do cognitivismo – interação sujeito-objeto do conhecimento e as construções intelectuais necessárias às respostas – se mostram presentes tanto em Língua Portuguesa quanto em Matemática.

Dessa forma, na elaboração de itens, torna-se possível detectar em que etapa de construção dessas noções os alunos se encontram, possibilitando um diagnóstico mais preciso.

Nos Estados Unidos, os testes também têm como referência as matrizes (Anexo 2) que refletem pontos de vista profissionais, evidências empíricas de pesquisa sobre medida de alta qualidade e um bom delineamento técnico em uma avaliação que mede o desempenho de milhares de estudantes (*Reading Framework for the National Assessment of Educational Progress*, 2004). Elas abarcam um amplo rol de conceitos que são apoiados por descobertas de pesquisas contemporâneas que são relevantes para os padrões estaduais e locais e para os currículos. As matrizes estão fundamentadas em um balanceamento de abordagens e programas os quais são considerados necessários para tornar os estudantes competentes. As matrizes representam uma síntese de várias perspectivas, ênfases e opiniões entre profissionais nas universidades, estados, distritos locais e salas de aula.

Cada avaliação do NAEP é construída em torno de uma matriz organizada que serve de plano para orientar o desenvolvimento do instrumento de avaliação e determinar os conteúdos a serem avaliados. O Comitê Governamental de Avaliação Nacional (*NAGB – National Assessment Governing Board*) é quem as desenvolve.

As matrizes capturam uma extensão de conteúdos de disciplinas específicas e habilidades de pensamento necessárias pelos estudantes com o objetivo de lidar com questões complexas que eles encontram dentro e fora das salas de aula. As matrizes são determinadas por meio de um processo de desenvolvimento que assegura que elas sejam apropriadas para as necessidades educacionais atuais. Pelo fato de as avaliações precisarem ser flexíveis para espelharem mudanças nos objetivos educacionais e currículos, as matrizes são avançadas e abrangentes, balanceando práticas de aprendizagem com descobertas de pesquisa.

As matrizes nos Estados Unidos servem como orientação para planejar avaliações ou revisar currículos. Podem também proporcionar informações sobre habilidades apropriadas para os três níveis avaliados (graus 4, 8 e 12) e podem ser modelos para medir essas habilidades de modo inovador. O desenvolvimento das matrizes e o processo de revisão periódico envolvem representantes de vários setores da sociedade e seguem os seguintes passos:

- a) participação ampla e revisões por educadores e especialistas nas áreas avaliadas;
- b) revisões por comitês cujos membros representem elaboradores de políticas, profissionais da educação e membros do público em geral;
- c) envolvimento de supervisores especialistas de agências de educação;
- d) ouvidores públicos;
- e) revisões por acadêmicos, pelo *staff* do Centro Nacional de Educação Estatística (NCES) e por um painel político consultivo.

O desenvolvimento dos itens toma por base uma tabela de especificações que segue os moldes das matrizes de referência. Nessa tabela há um referencial para elaboração dos itens encaixado em cada tópico ou subtópico em cada área de conteúdo avaliado. Os elaboradores de itens devem seguir rigorosamente o que é determinado nessas tabelas. Existem também regras de construção. Na Tabela 1, um exemplo de uma tabela de especificação, fundamentada na matriz de referência, para a construção de itens para a área de leitura.

Tabela 1 – Orientação para elaboração de itens abordando elementos de Textos Literários e Aspectos da Leitura – NAEP

ELEMENTOS DE UM TEXTO LITERÁRIO	ASPECTOS DA LEITURA			
	FORMAR UM ENTENDIMENTO GERAL	INTERPRETAR	ESTABELECEER CONEXÕES ENTRE O LEITOR E O TEXTO	EXAMINAR O CONTEÚDO E A ESTRUTURA
TEMA	Qual é a moral da história? Use uma evidência da história na sua resposta.	Como o ambiente ajuda a ilustrar o tema da história?	Você pensa que a lição dessa história é verdadeira nos dias de hoje? Por que?	Explica o que torna essa história uma fábula.
PERSONAGENS PRINCIPAIS	Qual era a opinião do principal personagem de _____?	Qual a causa o personagem principal ter feito _____? Use evidências da história em sua resposta.	Que ações diferentes do personagem você pensaria para os dias de hoje? Busque apoio para sua resposta em evidências do texto.	Como a descrição do autor de _____ ajudaria a explicar as ações do personagem?
EVENTOS PRINCIPAIS	Escreva um resumo dos principais eventos da história.	O que se espera depois de _____?	Como você pensa que a história terminaria se _____ não acontecesse?	Como o primeiro evento ajuda você a prever o final?
PROBLEMA	O que pode tornar o problema pior?	Como pode ajudar a resolver o problema?	O problema da história pode ser comparado com o de outra que você já tenha lido? Inclua evidência do texto e a outra história.	Por que o autor explica o problema na primeira parte da história? Explique com evidências da história.
VOCABULÁRIO	Quais palavras descrevem o que principalmente a história diz respeito?	Quais palavras deixam você saber que tempo se passa a história?	Explique o duplo significado de _____. Descubra qual significado melhor explica as idéias principais do texto.	Por que o autor usa as palavras _____ para descrever como _____ sente?

Todas as questões do NAEP enfatizam o pensamento crítico e o raciocínio em vez da memorização de fatos. As questões de múltipla escolha requerem que os estudantes selecionem uma única resposta clara. As de respostas construídas, ao contrário, requerem que os estudantes exercitem sua capacidade de comunicação do raciocínio.

Após o processo de elaboração dos itens, um extenso processo de revisão assegura que as avaliações sejam consistentes e que reúnam todos os critérios descritos em cada tabela de especificações. Comitês externos e especialistas em testagem e nas áreas de conteúdo avaliadas em nível estadual devem checar os conteúdos das avaliações. Os itens devem ser revisados quanto a preconceitos e sensibilidade, e especialistas em testes devem rever a qualidade técnica. Um extenso editorial da revisão proporciona garantia de qualidade. Finalmente, como o NAEP é regulamentado por lei, o Conselho Governamental revê os itens tanto antes da testagem em campo como após a testagem para verificar a adequação e possíveis vieses.

4.2.1.1 O que se avalia em leitura nos dois países?

Os principais documentos que referenciam as informações deste tópico são: do Brasil, *SAEB 2001 – Novas Perspectivas*, nos Estados Unidos, *Reading Framework for the National Assessment of Educational Progress*, 2004.

No Brasil, embora no documento supracitado, seja mencionada matriz de língua portuguesa, ela cobre habilidades relacionadas somente à leitura. É considerado que um bom leitor recorre a seus conhecimentos lingüísticos para ser capaz de perceber os sentidos, as intenções, implícitas e explícitas do texto e os recursos utilizados pelo autor para significar e atuar verbalmente.

Segundo o mesmo documento, no Brasil, o referencial teórico que fundamentou as matrizes, está alicerçado na finalidade do ensino de Língua Portuguesa definida nas Diretrizes Curriculares Nacionais e em como vem sendo adotada nas diversas propostas curriculares. Segundo esse documento, a finalidade do ensino é promover o desenvolvimento do aluno para o domínio ativo do discurso, sobretudo nas instâncias públicas de uso da linguagem, de modo a possibilitar sua inserção efetiva no mundo da escrita, ampliando suas possibilidades de participação social no exercício da cidadania. Nas Diretrizes como em outras propostas, a linguagem é concebida, ao mesmo tempo, como atividade cognitiva e como ação entre indivíduos. Assim, considera-se que o indivíduo que realmente domina uma língua, em suas variedades, sabe fazer uso de seus recursos expressivos em diversas situações e consegue cumprir melhor seus objetivos comunicativos, suas intenções interlocutórias.

É considerado que o aluno competente será capaz de considerar que todo texto oral ou escrito é um ato de linguagem e, nesse sentido, ao lê-lo ou ouvi-lo, descobrirá as finalidades e as intenções, os ditos e os não-ditos. Ao mesmo tempo, esse aluno saberá, por meio do texto, manifestar seus próprios desejos e convicções; e saberá também informar, persuadir, divertir, emocionar, argumentar, posicionar-se, criticar.

Essas são as justificativas apresentadas no documento para adotar o texto como eixo central das atividades de ensino-aprendizagem e avaliação da língua. Seria a partir do texto que as atividades e os conteúdos deveriam ser selecionados e organizados. No documento, é definido que o ensino da Língua Portuguesa deve centrar-se em três práticas:

1. A prática de compreensão de textos;
2. A prática de produção de textos; e
3. A prática de análise lingüística.

Dessa forma, a organização da matriz é resultante da compreensão de que, ao realizar essas práticas, o aluno modifica, reorganiza e reconstrói sua visão da língua em funcionamento, tornando-se capaz de estabelecer um número gradativamente maior de relações substantivas entre ele, como sujeito de conhecimento, e a língua e o objeto a serem conhecidos.

É chamada a atenção para a redução da matriz em relação à multiplicidade de competências e habilidades que são desenvolvidas pelo aluno durante sua vida escolar, não devendo, portanto, ser tomada como referência na implementação de propostas curriculares. Isso é justificado pelo fato de a matriz estabelecer apenas um conjunto de saberes e de habilidades cujo domínio é esperado de alunos de 4^a e 8^a séries do ensino fundamental e 3^a série do ensino médio.

O grau de complexidade dos textos, segundo a matriz, é resultante, entre outros, da temática desenvolvida, das estratégias textuais usadas em sua composição, da escolha de um vocabulário mais ou menos incomum, dos recursos sintático-semânticos utilizados, bem como das determinações específicas do gênero e da época em que foi produzido. E o grau de complexidade das tarefas contempladas nos testes resulta de aspectos relativos aos diversos saberes que o sujeito terá que mobilizar para resolver o problema proposto. Ao se considerarem as articulações entre a complexidade da tarefa e a do objeto, outros graus de dificuldade são introduzidos, de modo que não existirão padrões fixos de complexidade. O mesmo descritor, dependendo do texto escolhido e do tipo de tarefa solicitada, pode trazer graus de dificuldade diferentes. Por exemplo, quando se solicita que o leitor localize uma informação no texto, a tarefa poderá ser mais ou menos difícil, dependendo da complexidade desse texto e do tipo de questão proposta.

No referencial adotado pelo SAEB para a avaliação da Língua Portuguesa é reafirmada a disposição de ressaltar a relevância da exploração do teor discursivo do ensino da língua. Pois, dessa forma, considerou-se que seria rompida a tradição “conteudista” de abordagens descontextualizadas e, estaria assim, favorecido o desenvolvimento das múltiplas capacidades comunicativas de que o indivíduo deve dispor para responder às exigências de sua condição de ser social e participativo, chamado a intervir nos destinos da História, da qual é, irremediavelmente, sujeito.

A Matriz de Língua Portuguesa do SAEB, nessa perspectiva, aborda seis tópicos: (1) Procedimentos de Leitura; (2) Implicações do Suporte, do Gênero ou do Enunciador na Compreensão do Texto; (3) Relação entre Textos; (4) Coerência e Coesão no Processamento

do Texto; (5) Relações entre Recursos Expressivos e Efeitos de Sentido; e, (6) Variação Lingüística. A distribuição percentual de itens por tópico, planejada para o ano 2005, é apresentada na Tabela 2, considerando-se as prioridades para cada ciclo de escolarização.

Tabela 2 – Distribuição percentual dos itens por tópico em Língua Portuguesa - SAEB

TÓPICOS	4ª SÉRIE		8ª SÉRIE		3ª SÉRIE	
	PRIORIDADE	%	PRIORIDADE	%	PRIORIDADE	
1 - Procedimentos de leitura	P1	41	P3	19	P3	19
2 - Implicações do suporte, do gênero e/ou enunciador na compreensão do texto	P3	13	P5	6	P5	6
3 - Relação entre textos	P5	4	P4	8	P4	7
4 - Coerência e coesão no processamento do texto	P2	27	P1	42	P1	44
5 - Relação entre recursos expressivos e efeitos de sentido	P4	11	P2	21	P2	21
6 - Variação lingüística	P5	4	P5	4	P5	3

A seguir, um resumo do que é avaliado em cada tópico⁶.

- Procedimentos de Leitura – Neste tópico são requeridas dos alunos as habilidades em empregar estratégias para localizar informações explícitas e inferir informações implícitas em um texto. A segunda habilidade é mais complexa que a primeira, pois exige que o leitor ultrapasse os limites do que está escrito e reconheça também o que não está textualmente registrado, mas, subentendido.
- Implicações do Suporte, do Gênero ou do Enunciador na Compreensão do Texto – Neste campo, é requerido ao aluno estabelecer relações entre informações advindas de fontes diversas (suporte e gênero), bem como identificar a finalidade de um texto e de textos diferentes, relações entre o gênero do texto e sua função comunicativa.
- Relação entre Textos – Para demonstrar ter construído a competência relacionada a este tópico, o aluno deve ser capaz de comparar textos, analisando a relação entre o modo de tratamento do tema e as condições de produção, recepção e circulação deles, assim como de comparar textos que expressem opiniões diferentes sobre um mesmo fato ou tema.
- Coerência e Coesão no Processamento do Texto – Este tópico exige do leitor habilidades que o levem a identificar a linha de coerência do texto, o conflito gerador do enredo e os elementos que constroem a narrativa. O aluno deve compreender o

⁶ Informações retiradas do *site* do INEP/SAEB: <http://www.inep.gov.br/basica/SAEB/objetivos.htm>, em 12/11/2006

texto, não como um simples agrupamento de frases justapostas, mas como um conjunto contínuo e harmonioso, em que há interligações e relações entre suas partes e entre seus elementos, estabelecidas por repetições ou substituições lexicais.

- Relações entre Recursos Expressivos e Efeitos de Sentido – Pelos descritores componentes deste tópico, é exigido do aluno o conhecimento de diferentes gêneros textuais para que ele possa construir e antecipar a construção de significados a partir de recursos expressivos, como a ortografia, a ironia, o humor, a pontuação e outras notações promotoras da leitura para além dos elementos presentes na superfície do texto.
- Variação Lingüística – Este tópico refere-se às manifestações da fala no discurso direto e indireto e é medido pelo reconhecimento das marcas lingüísticas identificadoras do locutor e do interlocutor no texto.

Nos Estados Unidos, desde 1992, a Matriz de Leitura do NAEP (*Reading Framework for the National Assessment of Educational Progress*, 2004) proporciona orientações e a base teórica para desenvolver as avaliações da leitura administradas para amostras randômicas de estudantes. Ela reflete as idéias de muitos indivíduos e organizações envolvidas no ensino da leitura, incluindo pesquisadores, elaboradores de políticas, professores, empresários e outros membros do público. Além disso, a Matriz reflete pesquisa que vê a compreensão da leitura como um processo dinâmico e interativo.

Nesse contexto, a definição do NAEP de letramento em leitura tem sido usada internacionalmente para ajudar o público a entender o que os estudantes devem saber e serem capazes de desempenhar na área de leitura. Esse entendimento ajuda a unificar as importantes dimensões da leitura para o desempenho dos estudantes. A leitura inclui a habilidade de entender e usar textos escritos tanto para o próprio prazer quanto para a aprendizagem, para participar da sociedade e para desempenhar metas.

Segundo o referencial teórico constante na matriz do NAEP, p.3, o letramento em leitura é um direito fundamental de cada pessoa. Ele dá acesso à informação e a habilidade para exercer as funções da vida. A leitura promove desde a melhoria do poder da linguagem até a apreciação da beleza da poesia. O ato de ler prolonga a experiência humana por meio da exploração de eventos em trabalhos literários. A leitura é a chave para o conhecimento e a informação.

Em um mundo conduzido pela tecnologia da informação, ocorre uma crescente complexidade do domínio da leitura conforme vai se tornando diverso o formato de textos. Variedades de textos como CD-ROMs, páginas da Web, jornais e revistas de lugares diferentes demandarão do leitor novas estratégias de leitura.

Em um Relatório do Painel Nacional de Leitura é descrito, de forma sumária, o papel do leitor nas atividades de pesquisas como as empreendidas pelo NAEP: “*Na pesquisa cognitiva, leitura é intencional e ativa. De acordo com essa visão, um leitor lê um texto para entender o que está lendo, para construir representações na memória do que é entendido, e para colocar esse entendimento em prática*” (National Institute for Child Health and Human Development (NICHD), 2000, p. 4. Esses processos são a base para estabelecer metas amplas para a avaliação de leitura pelo NAEP.

Um dos meios de o NAEP refletir as definições correntes de letramento é pela diferenciação entre três contextos e quatro aspectos da leitura. Esses contextos e aspectos da leitura são os fundamentos da avaliação da leitura do NAEP.

No estabelecimento do que os alunos devem saber e serem capazes de fazer, a Matriz do NAEP especifica três contextos para a leitura:

- I. Leitura para experiência literária;
- II. Leitura para informação;
- III. Leitura para o desempenho de uma tarefa.

A leitura para experiência literária implica o exame de temas, eventos, personagens, ambientes, problemas e a linguagem característica de obras literárias pelo leitor. Dessa forma, o leitor considera sua experiência e conhecimento para o texto em atividades tais como eventos esperados, ambientes imaginados, conseqüências previstas, ações analisadas e considerando a linguagem de obras literárias. O leitor que lê para adquirir experiência literária deve pensar na perspectiva de autores e personagens e deve considerar a estrutura e a linguagem da história. Vários tipos de textos são associados com o contexto de leitura para experiência literária, incluindo histórias curtas, romances, poemas, peças de teatro, biografias, lendas, mitos e contos folclóricos.

A leitura para informação implica o engajamento do leitor com aspectos do mundo real. Ela é mais freqüentemente associada com livros didáticos, livros de fontes primárias e secundárias, artigos de jornais e revistas, ensaios e discursos. Algumas características que distinguem o texto informativo do literário são a organização e a forma em que a informação

é apresentada. O texto informativo é organizado por tópicos e detalhes secundários, enquanto o literário é organizado pela estrutura da história, poema, ou drama. Textos informativos podem ter títulos grandes, gráficos, ilustrações e letras grandes que assinalam importância no texto. Entretanto, algumas comunalidades existem entre a leitura de textos informativos, literários e as habilidades e estratégias requeridas para o entendimento de cada um. Ambos requerem que os leitores analisem criticamente o texto, reflita sobre ele e tire suas conclusões.

Quando a leitura é para informação, os leitores precisam conhecer padrões específicos ou formas de organização desse tipo de texto (por exemplo, causa e consequência, ordem seqüencial, comparações/contrastes, opiniões e argumentos secundários), para poder interpretá-lo. Geralmente, as pessoas têm diferentes estratégias para a leitura de textos dessa natureza. Por exemplo, pessoas podem precisar encontrar detalhes de uma informação para responder a uma questão, ou obter alguma informação quando passa os olhos em algum artigo de revista. A principal característica da leitura de textos informativos é que os leitores ao lê-los, têm como foco a aquisição de informações.

A leitura para o desempenho de uma tarefa implica a leitura para realizar ou fazer alguma coisa. Assim, quando pessoas lêem para realizar tarefas, elas usam as expectativas de seus propósitos e a estrutura própria desse tipo de texto para se orientarem na busca da informação. São considerados textos dessa natureza: gráficos, horários ou planos de viagem de ônibus ou de trens, orientações ou montagens de jogos, procedimentos de sala de aula e de bibliotecas, formulários de taxas ou de seguros, receitas, bulas, documentos de identidade, títulos de eleitor e cédulas de votação, mapas, garantias de consumidor, memorandos, comunicados internos, entre outros. A orientação para o leitor implica olhar para a informação específica para fazer alguma coisa. Nesse tipo de leitura, é exigido que o leitor aplique a informação que está sob a forma de texto escrito, não simplesmente a entenda. O leitor, nesse caso, não observa o estilo ou pensa no texto como faria na leitura para experiência literária. No quarto grau, considera-se que o aluno ainda não tenha maturidade para o contexto de leitura para o desempenho de uma tarefa.

Há uma distribuição proporcional de itens relacionados a cada contexto para leitura que muda de um grau para outro para refletir as mudanças de demandas conforme os alunos amadurecem. Essa proporção é especificada na Tabela 3.

Tabela 3 – Proporção de itens por contexto e grau

Grau	Contexto para leitura		
	Experiência literária (%)	Informação (%)	Desempenho de uma tarefa (%)
4	55	45	Nenhum
8	40	40	20
12	35	45	20

Os aspectos da leitura que fundamentam a avaliação da leitura do NAEP na composição da Matriz caracterizam a forma de os leitores buscarem informações para alcançarem a total compreensão do texto e têm sua origem no:

- I) entendimento geral do texto;
- II) desenvolvimento da interpretação;
- III) estabelecimento de conexões entre o texto e o leitor;
- IV) exame do conteúdo e da estrutura do texto.

A compreensão integral do texto pelo leitor se desenvolve em diferentes etapas. O leitor foca sua atenção em tópicos ou temas gerais, interpretam e integralizam idéias dentro e através do texto, faz conexões com o conhecimento e com as experiências adquiridas anteriormente e examinam o conteúdo e a estrutura do texto. As questões propostas para a avaliação deverão estar fundamentadas nesses quatro aspectos da leitura que requerem a seleção e a integração de várias estratégias de leitura em vez de aplicar uma única estratégia ou habilidade. As estratégias que os leitores usam para construir seu entendimento e ajustar sua abordagem são inerentes a esses aspectos.

Para alcançar a meta de aproximar experiências reais de leitura, os textos utilizados pelo NAEP são tipicamente aqueles que o estudante lê todos os dias. Os trechos abordados são de textos autênticos encontrados nos ambientes dos estudantes nos graus 4, 8 e 12. Textos selecionados de fontes originais, não simplificações, paráfrases, resumos, ou modernizações. Somente mudanças mínimas são permitidas pelo NAEP (por exemplo, substituir uma palavra desconhecida, revisar uma frase confusa, ou recolocar um nome não familiar por um popular). Tais mudanças não afetam o nível de dificuldade, o propósito, ou o significado do trecho do texto. Histórias completas, artigos, ou capítulos de livros são usados para estimularem o interesse dos estudantes. Esses textos devem tornar possível contemplar questões que produzam várias estratégias de leitura.

Os textos variam no tamanho. No 4º grau, estudantes lêem textos de 250 a 800 palavras; no 8º grau, 400 a 1.000 palavras, e no 12º grau, 500 a 1.500 palavras. Nos graus 8 e 12, podem ser dados até dois textos em um bloco de 50 minutos.

Os textos são selecionados para cada grau conforme o nível de desenvolvimento e o tema apropriado. Além disso, os textos devem expressar uma idéia completa e reunir critérios para gênero, linguagem, interesse e organização. Esses critérios incluem considerações curriculares, adequação ao tema, estilo, perspectiva e clareza. Os documentos são originais e se relacionam às tarefas que são apropriadas a cada grau e às experiências dos estudantes que estão sendo avaliados.

A dificuldade do item é uma função da dificuldade e da complexidade do texto, e da quantidade de conhecimento do contexto requerido para responder corretamente ao item. A dificuldade de um texto é uma função da complexidade de seus argumentos, da abstração de seus conceitos, da inclusão de pontos de vista não usuais e da projeção de tempo para a sua compreensão – fatores esses não conduzidos por medidas tradicionais de leitura. Assim, quando a dificuldade do texto for grande, a dificuldade das questões a ele relacionadas também o será, pois as questões focam sobre pontos importantes do texto.

4.2.1.2. O que se avalia em Matemática nos dois países?

Os documentos que referenciam as informações relacionadas às matrizes de Matemática do SAEB e do NAEP, foram respectivamente, “*SAEB 2001 – Novas Perspectivas, 2002*” e “*Mathematics Framework for the National Assessment Educational Progress, 2004*”.

Segundo documentos brasileiros, as matrizes que referenciam a elaboração dos testes de Matemática do SAEB basearam-se em três premissas básicas: (1) os conceitos matemáticos não se constituem verdades absolutas e se formam de maneira inter-relacionada, contemplando diferentes procedimentos de solução; (2) a aquisição do conhecimento de Matemática dá-se por meio de aprendizagens significativas relacionadas com o mundo real do sujeito, interpretado e construído em diferentes linguagens; (3) a avaliação deve aproximar-se o máximo possível da situação de aprendizagem do aluno (SAEB 2001 - Relatório Nacional, p. 13). Essas três premissas, junto às limitações impostas ao tipo de avaliação a ser realizada, indicaram a proposição de uma matriz compreendida basicamente de situações-problema

buscando a possibilidade de avaliar satisfatoriamente as competências evidenciadas pela aprendizagem dos conteúdos matemáticos.

No Doc. *SAEB 2001 – Novas Perspectivas, 2002*, é ressaltada a diferença entre o que se prescreve e o que de fato se realiza no processo de ensino e aprendizagem da Matemática, na educação básica. É destacada a necessidade de mudanças, as quais exigem novas orientações metodológicas e uma revisão dos objetivos. Segundo o documento, tem havido consenso entre os professores de que o ensino da Matemática não pode limitar-se a um processo que tenha como finalidade a simples memorização de regras e técnicas.

Dessa forma, é esperado que os professores que ensinam Matemática considerem a resolução de problemas como eixo norteador da atividade matemática, pois, a resolução de problemas possibilita o desenvolvimento de capacidades como: observação, estabelecimento de relações, comunicação (diferentes linguagens), argumentação e validação de processos, além de estimular formas de raciocínio como intuição, indução, dedução e estimativa. Essa opção traz implícita a convicção de que o conhecimento matemático ganha significado quando os alunos têm situações desafiadoras para resolver e trabalham para desenvolver estratégias de resolução.

Quanto à avaliação em Matemática, o documento sugere repensar certas idéias ainda dominantes entre os professores, notadamente as que concebem como prioritário avaliar a memorização de fórmulas, regras e esquemas. Ressalta-se que a avaliação em Matemática tem uma dimensão social, quando fornece informações ao aluno sobre seu desenvolvimento a respeito de capacidades e competências matemáticas exigidas socialmente. Considera-se, ainda, que a avaliação possui, evidentemente, uma dimensão pedagógica: ao fornecer dados sobre a aprendizagem de seus alunos, o professor poderá refletir sobre sua prática e propor novas abordagens para os conceitos e procedimentos ainda não consolidados.

É destacado no documento que o professor deve ter clareza sobre os critérios da avaliação e das limitações e possibilidades do instrumento a ser utilizado. A matriz de matemática do SAEB é um documento que visa identificar as competências em conteúdos matemáticos desenvolvidas no ensino fundamental e médio e que são passíveis de serem verificadas por meio de avaliações escritas. No entanto, ressalta-se que no documento é alertado que a matriz não pode ser interpretada como um conjunto de indicações que norteie as estratégias de ensino nas escolas, papel reservado a parâmetros, currículos e diretrizes curriculares.

Segundo consta, a matriz não traz orientações ou sugestões de como trabalhar em sala de aula, tampouco sugere progressão e hierarquia de conteúdos. Além disso, não menciona certas habilidades e competências que embora sejam importantes, não podem ser medidas por meio de uma prova escrita. Em outras palavras, a matriz sofre as limitações do tipo de instrumento (teste) utilizado na medição do desempenho.

No mesmo documento é afirmado não ser válido explicitar competências relacionadas a conhecimentos e procedimentos que não possam ser objetivamente verificados. Um exemplo: o conteúdo “utilizar procedimentos de cálculo mental”, que consta nos PCN, apesar de indicar uma importante capacidade que deve ser desenvolvida ao longo de todo o ensino fundamental, não tem, nesta Matriz, um descritor correspondente.

É levado em conta no processo de elaboração dos itens do SAEB, que para confirmar que um aluno tenha aprendido uma noção é necessário que ele seja capaz de utilizá-la para resolver um problema diferente daquele que deu origem à construção da noção, ao questionar as resoluções efetuadas e as respostas encontradas. Por isso, o teste deverá constituir-se, prioritariamente, por situações em que a resolução de problemas seja significativa para o aluno.

Por problemas significativos para o aluno, no documento é explicitado, que são as situações que permitem “recontextualizar” os conhecimentos que foram, por esse aluno, “descontextualizados” por ocasião de seu processo de aprendizagem. Ao contrário da situação-problema, cuja utilização está estreitamente ligada às situações de aprendizagem, os problemas significativos têm por objetivo verificar em que medida essa aprendizagem foi efetivamente realizada (p. 24).

Segundo consta, esta opção pela resolução de problemas significativos não exclui totalmente a possibilidade da proposição de alguns itens que avaliem se o aluno tem domínio de determinadas técnicas (p. 24). Por fim, é lembrado que os conhecimentos e competências matemáticas indicadas nos descritores da Matriz de Referência de Matemática estão presentes, de forma consensual, nos currículos das unidades da federação e nas Diretrizes Curriculares Nacionais. Esses descritores são apresentados em três níveis: 4ª e 8ª séries do ensino fundamental e a 3ª série do ensino médio, sendo assinalado que os descritores da 4ª série geralmente estão contemplados de forma mais abrangente nos descritores da 8ª, da mesma forma que estes estão incluídos nos descritores da 3ª série do ensino médio, e supõe-se que apresentam, evidentemente, graus de complexidade diferenciados.

A Tabela 4 apresenta os percentuais planejados para o SAEB 2005, relativos à distribuição dos itens nas provas, em função das prioridades e do número de descritores associados a cada tema e para cada série avaliada.

Tabela 4 – Distribuição percentual dos itens por tema em Matemática - SAEB

TEMA	4ª SÉRIE			8ª SÉRIE			3ª SÉRIE		
	Prior	Nº Desc	%	Prior	Nº Desc	%	Prior	Nº Desc	%
1 – Espaço e Forma	P3	05	17	P2	11	29	P3	10	28
2 – Grandezas e Medidas	P2	07	31	P3	04	11	P4	03	10
3 – Números e operações/Álgebra e funções	P1	14	46	P1	20	55	P1	20	57
4 – Tratamento da Informação	P4	02	6	P2	02	5	P2	02	5

Nos Estados Unidos, a avaliação de Matemática do NAEP é similar à de Leitura, pois são os dois focos principais das avaliações nacionais nos anos recentes. Os testes de desempenho em Matemática têm a função de avaliar quão bem os estudantes americanos desempenham nos conteúdos que exigem habilidade matemática.

Segundo o documento “*Mathematics Framework for the National Assessment Educational Progress, 2004*”, desde 1992, por meio de um trabalho realizado pelo *College Board*, a Matriz de Matemática do NAEP proporciona orientações e a base teórica para desenvolver as avaliações administradas para amostras randômicas de estudantes. Ela reflete as idéias de muitos indivíduos e organizações envolvidas no ensino da Matemática, incluindo pesquisadores, elaboradores de políticas, professores, empresários e outros membros do público. Além disso, a Matriz reflete pesquisas atuais que consideram a matemática como essencial no desenvolvimento de habilidades que requerem o desenvolvimento da capacidade de raciocinar.

A Matriz de Matemática do NAEP tem passado por várias mudanças na tentativa de alcançar o rápido avanço tecnológico e as novas tendências e exigências que os momentos atuais exigem. Como resultado da análise e revisão da Matriz, os conselhos e comitês de direção e planejamento sancionaram recomendações para a avaliação de Matemática de 2005 nos seguintes tópicos: áreas de conteúdo; complexidade matemática dos itens; distribuição de itens; e, formato dos itens.

Desde as primeiras avaliações no final dos anos 70 e início dos 80, o NAEP tem regularmente reunido dados sobre o entendimento dos estudantes a respeito dos conteúdos

matemáticos. Embora os nomes das áreas dos conteúdos que compõem as matrizes, bem como alguns de seus tópicos nessas áreas, tenham mudado um pouco de uma avaliação para outra, desse período até hoje, um consistente foco tem sido mantido frente à coleta de informações sobre o desempenho de estudantes nas cinco áreas principais: (1) Propriedades do Número e Operações; (2) Medidas; (3) Geometria; (4) Análise de dados e probabilidade; e (5) Álgebra.

Com essas divisões não se tenciona separar matemática dentro de elementos discretos. A idéia é proporcionar um esquema de classificação útil que descreva um espectro completo de conteúdos matemáticos avaliados pelo NAEP. No grau 12, as cinco áreas são colapsadas em quatro, com geometria e medida, combinadas em uma. Isso reflete o fato de que a maioria dos tópicos de medida se ajusta à natureza geométrica. É importante observar que certos aspectos da matemática ocorrem em todas as áreas de conteúdo. O melhor exemplo disso é o tópico cálculo. Calcular é a habilidade de desempenhar operações sobre os números. Não deve ser confundido com a área de Propriedades do Número e Operações, a qual abarca uma ampla extensão de conceitos acerca de nosso sistema de numeração. Certamente a área de Propriedades do Número e Operações inclui uma variedade de habilidades de cálculo, que vai desde operações com números inteiros até trabalho com decimais e frações. Mas trabalha-se com cálculo também em Medidas e Geometria, como por exemplo, no cálculo do perímetro de um retângulo, na estimativa da altura de uma construção, ou no cálculo da hipotenusa de um triângulo retângulo. Análise de dados freqüentemente envolve cálculo, como o cálculo de uma média, ou a abrangência de um conjunto de dados. Probabilidade freqüentemente exige trabalho com números racionais. Resolver equações algébricas envolve cálculo numérico. O cálculo, portanto, se configura em um exemplo de uma habilidade fundamental em todas as áreas de conteúdo e desfaz a idéia da abordagem de conteúdos isolados na avaliação.

A área “Propriedade dos números e operações” contempla na avaliação cinco subtópicos: sentido dos números, estimação, operações com números, razão e proporcionalidade e propriedades do número e operações. A área de Medidas apresenta dois subtópicos: atributos físicos e sistemas de medida. A área de Geometria apresenta cinco subtópicos: dimensão e forma, transformação de formas e preservação das propriedades, relacionamentos entre figuras geométricas, posição e direção e raciocínio matemático. A área de Análise de dados e Probabilidade apresenta quatro subtópicos: representação dos dados, características dos conjuntos de dados, experimentos e amostras e probabilidade. A área de

Álgebra apresenta quatro subtópicos: padrões, relações e funções, representações algébricas, variáveis, expressões e operações e equações e inequações.

A matriz de matemática do NAEP relaciona os tópicos e subtópicos em toda a extensão da educação básica. A matriz é apresentada de forma que se observa o desenvolvimento das habilidades ao longo dos ciclos de escolarização. A título de exemplo, na Tabela 5 é apresentado um trecho da matriz relacionado ao tópico “Propriedade dos números e operações”.

Tabela 5 – Subtópico Operações com números, relacionado a Propriedade dos números e Operações (NAEP)

OPERAÇÕES COM NÚMEROS		
GRAU 4	GRAU 8	GRAU 12
a) Adicionar e subtrair: <ul style="list-style-type: none"> números inteiros, ou frações com denominadores comuns, ou décimos até centésimos. 	a) Executar cálculos com números racionais.	a) Executar cálculos com números reais, incluindo números irracionais ou números de valores absolutos.
b) Multiplicar números inteiros: <ul style="list-style-type: none"> não maiores que dois dígitos por dois dígitos, com cálculo com lápis e papel, ou, números maiores com uso de calculadora. 		
c) Dividir números inteiros: <ul style="list-style-type: none"> até três dígitos por um dígito, com cálculo com lápis e papel, ou até cinco dígitos por dois dígitos, com uso de calculadora. 		
d) Descrever o efeito de operações sobre tamanho. (números inteiros)	d) Descrever o efeito de multiplicar e dividir por números, incluindo o efeito de multiplicar ou dividir um número racional por: <ul style="list-style-type: none"> zero, ou um número menor que zero, ou um número entre zero e um, ou, um, ou um número maior que um. 	d) Descrever o efeito de multiplicar e dividir por números, incluindo o efeito de multiplicar ou dividir um número real por: <ul style="list-style-type: none"> zero, ou um número menor que zero, ou um número entre zero e um, ou, um, ou um número maior que um.
	e) Proporcionar um argumento matemático para explicar operações com duas ou mais frações.	
f) Interpretar operações com números inteiros e as relações entre eles.	f) Interpretar operações com números racionais e as relações entre eles.	
g) Resolver problemas de aplicação envolvendo números e operações.	g) Resolver problemas de aplicação envolvendo números racionais e operações usando respostas exatas ou estimativas apropriadas.	g) Resolver problemas de aplicação envolvendo números racionais e irracionais comuns, operações usando respostas exatas ou estimativas apropriadas.

Com a intenção de assegurar que a Matriz 2005 mantenha o mesmo propósito que a anterior, avaliando um conteúdo apropriado e equilibrado por meio da abordagem de uma variedade de meios e ações matemáticas, foi proposta nessa matriz a complexidade matemática relacionada ao item, uma outra dimensão baseada nas propriedades de um item,

em vez de fundamentar-se nas habilidades do estudante. A complexidade matemática de um item responde à questão: “O que o item pergunta aos estudantes?”.

Cada nível de complexidade inclui aspectos do saber e do fazer matemáticos, tais como o raciocínio, os procedimentos de execução, os entendimentos dos conceitos, ou a resolução dos problemas. Os níveis são ordenados de forma que itens em níveis mais baixos demandariam que estudantes executassem procedimentos mais simples como o entendimento de conceitos elementares, ou solução de problemas simples. Os de complexidade moderada, envolvem mais flexibilidade de pensamento e escolha entre alternativas. Dos itens de alta complexidade, é esperado que os estudantes raciocinem e comuniquem conceitos sofisticados, executem procedimentos complexos, ou solucionem problemas não rotineiros. A ordenação dos níveis não se relaciona com uma seqüência desenvolvimental ou pela seqüência em que ocorrem o ensino e a aprendizagem. Em vez disso, há uma descrição de diferentes demandas feitas aos estudantes por meio de itens de teste.

A distribuição de itens entre as várias áreas de conteúdo matemático é uma característica crítica do delineamento da avaliação, que reflete a importância relativa e o valor dado a cada uma das áreas de conteúdo curricular dentro da matemática. Como tem sido nas avaliações passadas do NAEP, em matemática, as categorias têm recebido uma ênfase diferencial em cada grau avaliado e assim continuou na Matriz 2005. A Tabela 6 apresenta a distribuição percentual recomendada de itens por conteúdo para a avaliação de 2005.

Tabela 6 – Distribuição percentual de itens por grau e áreas de conteúdo (NAEP)

ÁREA DE CONTEÚDO (2005)	GRAU 4 (%)	GRAU 8 (%)	GRAU 12 (%)
Propriedades do Número e Operações	40	20	10
Medida	20	15	30
Geometria	15	20	
Análise de dados e Probabilidade	10	15	25
Álgebra	15	30	35

Ao longo da descrição do que referencia as avaliações de leitura e matemática para os dois países, muitas diferenças puderam ser constatadas. Verifica-se que a avaliação americana já possui um referencial consolidado e bem constituído.

Uma observação importante a ser discutida a respeito dos referenciais é a consistência do referencial teórico das matrizes do SAEB. Constata-se que há uma dissonância entre o discurso e a prática, além de uma falta de lógica entre o que é abordado como descritor e o referencial teórico. Neste sentido constata-se a necessidade de rediscutir e repensar as matrizes enquanto referenciais para a avaliação. Ainda em relação a isso, verifica-

se que nos Estados Unidos as matrizes são pensadas de forma que elas se tornem padrões para o processo ensino-aprendizagem. No Brasil, há a recomendação expressa de que as matrizes não sejam tomadas como referência, o que de fato, na forma em que elas se encontram, não seria aconselhável. Mas, propõe-se que seja feita uma reflexão sobre isso, considerando-se a falta de padrões que o sistema educacional enfrenta.

Em relação à tabela de especificações para elaboração de itens, que é uma prática do sistema norte-americano, poderia também ser adotada pelo sistema brasileiro. Pois, constata-se a dificuldade de elaboradores em terem um norte além da matriz, porque os descritores abordam habilidades muito amplas. Dessa forma, o construtor de itens teria um suporte objetivo, podendo exercitar só a criatividade, melhorando, quem sabe a qualidade dos itens.

Quanto aos conteúdos avaliados em leitura e matemática, que são abordados nas matrizes brasileiras, constata-se que se faz necessário revê-los e adequá-los às diretrizes de ensino e ao momento atual que exige um novo desenho do conhecimento voltado para a construção do processo de letramento de todas as áreas do currículo escolar. A proposta americana é muito interessante e bem consistente com esses momentos atuais. Enfim, por meio dessa exposição, têm-se subsídios para refletir sobre o que sustenta os testes atuais em nosso país.

4.2.2 Como se procede à análise dos dados nos dois países?

O processo de análise dos instrumentos dos dois países são muito semelhantes. São utilizadas, basicamente as mesmas análises que têm a finalidade de garantir a qualidade psicométrica dos instrumentos e assegurar uma calibração confiável dos itens que darão um perfil do sistema em relação à qualidade do aprendizado. As informações aqui veiculadas foram coletadas dos Relatórios técnicos do SAEB 99, 2001 e 2003, escritos pelo Professor Ruben Klein e pelo Technical Report: The NAEP 1998, organizado por Nancy L. Allen, John R. Donoghue e Terry L. Schoeps.

Tanto no Brasil quanto nos Estados Unidos são utilizados diferentes instrumentos de coleta de dados, sendo os testes que avaliam o desempenho um deles, o qual está sendo objeto de análise nesse estudo.

Nos relatórios pesquisados dos dois países consta como primeiro passo a revisão dos testes quanto ao conteúdo e à forma, que compreende uma certificação da qualidade dos

aspectos teóricos, pedagógicos e lingüísticos dos itens que contribuirá com a validação empírica, que verificará, por meio de técnicas estatísticas e psicométricas, as características de comportamento dos itens após sua aplicação em campo.

As análises estatísticas certificam a qualidade psicométrica dos itens. Por meio delas é estimado o poder de discriminação do item, ou seja, sua capacidade de diferenciar os alunos que conhecem o conteúdo e já desenvolveram as competências requeridas dos demais que ainda não atingiram a proficiência necessária para respondê-lo; o índice de dificuldade de cada questão, que permite equilibrar as provas com questões de diferentes graus de dificuldade; e a probabilidade de acerto ao acaso, que indica a chance de acerto do item sem o conhecimento e a construção da habilidade requerida (acerto pelo “chute”).

As diferenças das análises adotadas pelos dois países relacionam-se principalmente aos variados formatos dos itens adotados pelo sistema norte americano (múltipla escolha, resposta construída curta e a estendida), à seleção da amostra e às análises por subdomínios de conteúdos por meio de valores plausíveis.

Os testes do SAEB e do NAEP utilizam um delineamento amostral de itens denominado “Blocos Incompletos Balanceados” (*Balanced Incomplete Block*) – BIB – que permite que um quantitativo maior de itens seja dividido em subconjuntos menores chamados blocos. Esses blocos são organizados em grupos de diferentes combinações de forma espiralada o que garante que cada bloco se emparelhará a outro, e em todas as posições possíveis. Nos Estados Unidos é também utilizada uma variante desse delineamento chamado “Blocos Incompletos Parcialmente Balanceados” (*Partially Balanced Incomplete Block*) – BIPB. Cada combinação resulta em um caderno de prova, e todas as combinações em cadernos diferentes. Nas Tabelas 7 e 8, a seguir, são apresentados exemplos de cada variante citada do delineamento amostral de itens. Os números de blocos por teste variam nos dois países. No Brasil, desde 1995, adotou-se um delineamento com treze blocos e número variável de itens.

Tabela 7 – Exemplo de BIB

Caderno	Bloco I	Bloco II	Bloco III
1	A	B	D
2	B	C	E
3	C	D	F
4	D	E	G
5	E	F	A
6	F	G	B
7	G	A	C

Tabela 8 – Exemplo de BIPB

Caderno	Bloco I	Bloco II
1	A	C
2	B	A
3	C	D
4	D	B
5	A	D
6	B	C
7	H	E
8	E	F
9	F	G
10	G	H
11	G	E
12	H	F
13	C	G
14	D	H
15	E	B
16	F	A

Essa distribuição assegura que sejam superados fatores intervenientes na resposta aos itens, como o cansaço do aluno ou a falta de tempo para responder às questões que aparecem no final do teste. Essa técnica propicia a cobertura de um amplo espectro curricular em cada levantamento, permitindo inferências sobre o sistema educacional e não sobre os conhecimentos individuais de cada aluno (Beaton, Johnson e Ferris, 1987).

Para garantir a comparabilidade entre os anos, mantêm-se alguns blocos de itens já aplicados em anos anteriores. Já a comparabilidade do desempenho dos alunos entre as três séries avaliadas, é assegurada pela aplicação de blocos da 4ª série à 8ª série do ensino fundamental, bem como da 8ª série à 3ª série do ensino médio.

Após a aplicação dos testes, eles são analisados por meio de duas teorias: a Teoria Clássica do Teste e a Teoria de Resposta ao Item. Os modelos que fundamentam as duas análises se apóiam em estatísticas que buscam analisar o desempenho dos alunos resultante da aplicação dos testes, considerando-se os níveis de dificuldade, de discriminação e de acerto ao acaso dos itens.

As primeiras análises são efetuadas com base na TCT. Os parâmetros gerados referem-se ao cálculo da proporção de acerto do item e de escolha a cada alternativa, ao índice de discriminação, ao índice “D”, definido pela diferença de 27% dos alunos com maiores e menores medidas de desempenho, respectivamente, e ao coeficiente bisserial do item e por alternativa. Essas análises são realizadas para o Brasil, para cada uma das cinco regiões geográficas e para cada unidade da federação (UF).

No Brasil, ao fazer as análises, o peso amostral dos alunos e o delineamento BIB são levados em conta. Considerando-se a análise com o BIB, o escore dos alunos apresenta a inconveniência dos cadernos não terem dificuldades iguais e, portanto, há problemas de comparabilidade do número de acertos. Por exemplo, espera-se que um aluno de 8ª série obtenha um escore maior em um item de um bloco comum com a 4ª série. Para efetuar esta análise, a Fundação Cesgranrio, que foi quem realizou as análises nesse período de 1995 a 2003, utiliza um programa desenvolvido por Klein e Klein (1998). Esse Programa, além de levar em conta o BBI e os pesos dos alunos, calcula a proporção de respostas em cada alternativa e a correlação bisserial entre cada opção de resposta de múltipla escolha e uma medida de desempenho, no caso, os escores. As correlações bisseriais são calculadas usando o *normit*, uma variável obtida por uma transformação não linear a partir dos escores dos examinandos.

Antes de realizar as análises utilizando o procedimento da *normit*, torna-se necessário para cada série/disciplina, uma análise da dificuldade de cada caderno e de cada bloco, levando em conta o bloco e sua posição no caderno.

Nos Estados Unidos, é feita uma intensa análise exploratória dos itens que inclui análises convencionais, com e sem os pesos amostrais dos examinandos incorporados. As análises são conduzidas no nível do bloco, sendo produzidos escores conforme o número de respostas corretas a cada item, número de respostas a cada alternativa, número de casos omissos, e do não alcance. Dentro dos blocos, há itens de múltipla escolha e de respostas construídas curtas e extensas. Para cada um desses tipos, são realizadas análises que embora sejam análogas, requerem procedimentos diferenciados.

O processo de correção das respostas dos testes, etapa anterior às análises estatísticas, exige rigor e tempo para a sua execução. As perguntas de respostas construídas requerem que os estudantes formulem suas próprias respostas. Os avaliadores qualificados e treinados atribuem um escore para as respostas construídas. Atribuir esses escores a um grande número de questões de respostas construídas com um elevado grau de qualidade – em nível de confiabilidade – e dentro de uma estrutura de tempo limitado, é essencial ao sucesso da avaliação conduzida pelo NAEP (no ano 2005, mais de três milhões de respostas construídas foram corrigidas). Para garantir confiabilidade e correção rápida, o NAEP adota os seguintes passos nessa etapa:

- desenvolve guias com escores explicitados, focalizados a cada questão, que emparelham os critérios delineados segundo as matrizes da avaliação;

- recruta avaliadores qualificados e experimentados, treinando-os, e verificando a habilidade para atribuir escores para questões específicas por meio dos testes qualificados;
- emprega um processamento de imagem e um sistema de correção que encaminha as imagens das respostas dos estudantes diretamente aos avaliadores que podem assim realizar a avaliação mais rápida;
- monitora a consistência dos avaliadores por meio de verificações da fidedignidade durante o andamento da avaliação;
- avalia a qualidade da tomada de decisão do avaliador por meio do monitoramento freqüente por especialistas em avaliação do NAEP; e
- documenta todo o treinamento, a atribuição de escores, e os procedimentos de controle de qualidade nos relatórios técnicos.

Enquanto as questões de respostas construídas são avaliadas usando o sistema de processamento da imagem, as respostas às perguntas de múltipla escolha são lidas por meio do escaneamento dos cadernos de teste.

No Brasil como nos Estados Unidos, para assegurar medidas de desempenho justas e confiáveis, após o processo de atribuição de escores às respostas, os dados e os desempenhos são checados por vários meios em relação à fidedignidade da correção, à análise dos itens e ao funcionamento diferencial do item (*Differential Item functioning - DIF*). Quanto à análise do funcionamento diferencial do item (DIF). Ela se refere a procedimentos que avaliam se os itens são diferencialmente difíceis para grupos diferentes de examinandos. Os procedimentos utilizados incluem o controle das diferenças entre grupos observando-se um critério, geralmente, os escores do teste. O desempenho de cada item entre os grupos é então comparado dentro do conjunto de examinandos que têm o mesmo escore total no teste.

Os itens dicotômicos corrigidos são analisados usando procedimentos padrão que resultam em um registro para cada item. Esse registro é resultante das seguintes análises do item: número de estudantes que omitiram ou que não tenham atingido o item; o número de examinandos que responderam; a porcentagem; a média e o desvio padrão de respostas corretas para o bloco no qual o item aparece; a porcentagem de examinandos que dá uma resposta sem atenção; p^+ , que é a proporção de examinandos que dá uma resposta correta ao item (razão entre o número de respostas corretas pelo número de respostas corretas mais as erradas e as omitidas); Δ , que é a inversa normalmente transformada do p^+ escalado para média 13 e desvio padrão 4; o coeficiente de correlação bisserial entre o item e o número de

respostas corretas para o bloco no qual o item aparece, o coeficiente de correlação ponto-bisserial.

Os itens politômicos corrigidos, prevalentes nos testes americanos, são analisados por métodos paralelos aos itens dicotômicos, os quais produzem também registros individuais para cada item. Categorias de respostas são definidas em duas formas: uma baseada nos códigos originais para as respostas, como especificado pelos avaliadores, e outra usada na definição do modelo da TRI, a qual é definida pelo tratamento de cada categoria de respostas dada a classificação pela análise da TRI. Assim, na classificação original pode-se considerar até sete categorias de respostas, levando em conta desde aqueles que não acertam nada até aqueles que acertam completamente, além daqueles que omitiram e não atingiram o item. Considerando-se a análise da TRI, algumas categorias podem ser combinadas, redundando em quatro ou cinco categorias.

As seguintes estatísticas, análogas àquelas dos itens dicotômicos corrigidos, são computadas: a proporção de examinandos que dá uma resposta sem atenção; no lugar do $p+$, a razão da média do escore no item pelo escore máximo possível que foi atribuído ao item; no lugar do Δ , a razão da inversa normalmente transformada do escore médio do item pelo escore máximo possível do item para a média 13 e desvio padrão 4; o coeficiente de correlação polisserial é usado no lugar da bisserial; o coeficiente de correlação de Pearson é usado no lugar da ponto bisserial.

Após essas análises, são geradas tabelas com resultados ao nível dos itens. Essas tabelas sumarizam as estatísticas dos itens para todos eles dentro de cada bloco, considerando-se o tamanho amostral, sem peso e com peso. Nos relatórios americanos não são apresentadas tabelas informando os índices gerados da TCT. No Brasil, são informados os índices da TCT por item e por alternativa (proporção de respostas e correlação bisserial), além da análise gráfica por item. Esses dados facilitam a compreensão do desempenho dos alunos, permitindo uma integração entre as análises estatísticas e pedagógicas e conseqüentemente uma melhor análise da qualidade dos itens.

A atribuição dos pesos para as amostras participantes da avaliação americana se diferencia da brasileira, principalmente em função dos resultados da avaliação por estado, conforme pode ser observado a seguir.

Os pesos amostrais se relacionam às escolas e aos estudantes participantes que são escolhidos para serem nacionalmente representativos e que são selecionados em cada estado,

no Distrito de Columbia e nas escolas do Departamento de Defesa. Desde 2002, uma amostra combinada de escolas públicas tem sido selecionada para o estado e para a nação. Assim, da amostra selecionada para a avaliação de um estado, uma subamostra é designada para ser representativa dos estudantes da nação. No nível do estado, os resultados são relatados somente para escolas públicas e para alguns agrupamentos demográficos que permitem análises. Nas avaliações em que os estados participam, os resultados são também relatados em nível estadual. A amostra para o NAEP nacional é composta de todas as subamostras estaduais de estudantes de escolas públicas, bem como de amostras de estudantes de escolas privadas. Em estados não participantes, um determinado número de escolas e estudantes são selecionados para completar a amostra em nível nacional.

Nos estados, uma amostra de aproximadamente 2.500 estudantes de 100 escolas é selecionada, por grau e disciplina. A seleção das escolas é randômica dentro das categorias de escolas com características similares. Mas algumas escolas ou grupos de escolas (distritos) podem ser escolhidos em todos os ciclos de avaliação se eles forem os únicos representantes de determinadas características. Normalmente, 30 estudantes por disciplina e por série são selecionados aleatoriamente em cada escola. Como a meta é avaliar todos os estudantes selecionados na amostra, inclusive aqueles que apresentam alguma deficiência, o NAEP tem proposto algumas adaptações nos testes ou nos padrões de aplicação para permitir que todos participem. A avaliação nacional relata resultados para escolas públicas e não-públicas, para todas as regiões do país, bem como, por sexo, raça, grau de urbanização, posição da escola, instrução dos pais e participação no Programa Nacional de Almoço na Escola (*National School Lunch Program - NSLP*).

Os resultados dos estudantes avaliados são combinados por meio dos valores plausíveis atribuídos para proporcionar estimativas acuradas do desempenho geral da nação e do desempenho dos estados individuais e de algumas jurisdições. Os pesos atribuídos aos resultados, leva em consideração o fato de que estados e escolas dentro dos estados, representam diferentes proporções da população nacional geral e estadual.

Tanto no Brasil, quanto nos Estados Unidos, após essas análises iniciais, os dados das amostras da avaliação são calibrados usando modelos da TRI. Para itens dicotômicos corrigidos, formas logísticas dos modelos de dois e três parâmetros são usadas, enquanto para os itens politômicos, é utilizada a forma do modelo de crédito parcial generalizado. Esses dois tipos de itens e modelos são combinados gerando as escalas do SAEB e do NAEP. Estimativas dos parâmetros do item são obtidas em uma escala provisória, usando os

programas Bilog e Parscale. A adequação do modelo de TRI para os dados observados é analisada em cada escala comparando as funções de resposta ao item empíricas com as curvas teóricas. Gráficos dessas funções de respostas ao item empíricas e curvas teóricas são, então, comparados. As análises de DIF, previamente descritas, também proporcionam informações relacionadas à adequação do modelo através das subpopulações.

Outro diferencial, tomando por base as análises americanas, diz respeito às estimações dos parâmetros que são geradas também separadamente para cada uma das três escalas de Leitura, definidas aos contextos contemplados na matriz (literária, informativa e para desempenho de tarefas) e, em Matemática, aos cinco conteúdos avaliados (propriedades dos números e operações, medidas, geometria, análise de dados e probabilidade e álgebra). Para leitura e matemática,

Este processo é feito após a calibração conjunta dos itens. Valores plausíveis são extraídos da distribuição preditiva dos valores dos escores calibrados para cada estudante (esse processo é chamado condicional). As variáveis usadas para calcular os valores plausíveis para uma escala de avaliação nacional ou para os grupos de escalas incluem um amplo espectro de variáveis de contexto, de atitude e experimentais, e as composições de tais variáveis. Para melhorar a estabilidade numérica das escalas, as variáveis de contexto originais são padronizadas e transformadas em um conjunto de variáveis linearmente independentes pela extração dos componentes principais da matriz de correlação das variáveis originais. Assim, em vez de se utilizar as variáveis originais, são usados os componentes principais como variáveis independentes para calcular os valores plausíveis daquelas escalas.

Tanto o SAEB quanto o NAEP realizam também a análise da unidimensionalidade dos itens. Várias estatísticas e simulações são utilizadas verificando a inflação do número de dimensões e examinando, quando evidenciada mais de uma dimensão, a força de cada uma dessas dimensões para apoiar a utilização ou não de uma escala composta para descrever os dados. Depois dos itens calibrados, é feita a transformação para a métrica das escalas por procedimentos de *linkagem* via população comum. Dessa forma, os dados da última avaliação são calibrados juntos com os da atual, pelos programas BILOG-MG e PARSCALE, especificando amostras de cada avaliação como vindas de diferentes populações. Após esse procedimento são realizadas as transformações sendo estabelecidas novas escalas.

Para extrair inferências dos resultados das análises estatísticas é importante observar que todo o processo é produto de estimativas, estando, portanto, sujeito a medidas de incerteza, ou erros de medida. E sobre as fontes de incerteza pode-se afirmar que existem

duas: a primeira, que o SAEB e o NAEP usam amostras de estudantes em vez de testar todos os estudantes; e a segunda, que as avaliações não questionam todo o espectro relacionado aos conteúdos contemplados, em vez disso, é selecionada uma amostra desses conteúdos. A magnitude dessas incertezas é, então, refletida no erro padrão de cada estimativa. Portanto, as comparações apresentadas e discutidas nos relatórios são baseadas sobre testes estatísticos que consideram os erros padrão dessas estatísticas e a magnitude da diferença entre as médias ou as porcentagens.

4.2.3 Como os resultados são relatados nos dois países?

Em relação à forma de relatos dos resultados dos dois países, grandes diferenças são observadas. Os últimos resultados publicados pelo SAEB foram os relativos ao SAEB 2001 e os resultados da Prova Brasil 2005. Os resultados de 2001 foram apresentados em relatórios contendo resultados nacionais, pedagógicos, regionais e estaduais, por série e disciplina. O relatório nacional apresentou dados relativos ao contexto completo da avaliação, desde aspectos técnicos das análises, até o contexto educacional da década e reflexões sobre a eficácia e a equidade do sistema educacional brasileiro. Foram apresentados resultados do desempenho dos alunos em escala única e por série, com interpretações e exemplos de itens para cada uma delas. Nos relatórios pedagógicos foram feitas análises relacionando o que foi avaliado, o que os resultados indicam e em que a prática pedagógica do professor pode ser melhorada. Outra publicação feita também com base nos resultados de 2001, embora publicada em 2003 (Qualidade da Educação: uma nova leitura do desempenho dos estudantes, 2003), em outra gestão da DAEB, foi em relação leitura do desempenho dos estudantes.

Os resultados da Prova Brasil foram publicados na Internet e enviados para cada escola na forma de um boletim. Os dados disponibilizados na Internet são quantitativos descritivos da escola relacionados a todos os participantes da avaliação e os indicadores de desempenho da escola em relação às médias de outras categorias avaliadas. Todas essas informações encontram-se na forma de tabelas e gráficos. Em relação à escala de proficiência, ela foi publicada na forma de um boletim que foi enviado, em separado às escolas.

Os resultados das avaliações de leitura e matemática do NAEP são disponibilizados para diversas categorias de interesse em duas formas: impressos e via Internet. Dos que são impressos, aquele que tem maior divulgação é o apresentado na forma de um sumário

executivo em que são relatados escores médios para grupos de estudantes na escala 0–500 e as porcentagens de estudantes que alcançam cada um dos três níveis de desempenho: Básico, Proficiente e Avançado. Os níveis de desempenho oferecem uma média de porcentagens de estudantes identificados que demonstram certas proficiências na leitura e na matemática. Autorizado pela legislação do NAEP e adotado pelo Conselho Governamental, os níveis de desempenho são padrões baseados em julgamentos coletivos acerca do que os estudantes deveriam saber e serem capazes de fazer, fundamentados no que é previsto nas Matrizes de Leitura e Matemática do NAEP. Na Tabela 9, são apresentadas as definições políticas dos níveis de desempenho do NAEP.

Tabela 9 – Definição política dos níveis de desempenho do NAEP

AVANÇADO	Desempenho superior
PROFICIENTE	Desempenho acadêmico sólido para cada grau avaliado. Estudantes que alcançam esse nível demonstram competência em responder aos desafios da disciplina avaliada, incluindo o conhecimento da matéria, aplicação de tais conhecimentos a situações do mundo real e habilidades analíticas apropriadas à matéria.
BÁSICO	Domínio parcial de conhecimento e habilidades que são fundamentais para um trabalho proficiente em cada grau.

Fonte: U. S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Reading and Mathematics Assessment

Como parte do processo para fixar níveis de desempenho em cada disciplina avaliada pelo NAEP, painéis de professores, especialistas em currículo, elaboradores de políticas e membros do público em geral usam definições políticas genéricas para desenvolver descrições do que os estudantes deveriam saber e serem capazes de fazer em cada um dos níveis, nos três graus avaliados. Na Tabela 10 são apresentadas como exemplo, as descrições que ilustram os níveis de desempenho em Leitura do quarto grau.

Tabela 10 – Níveis de Desempenho em Leitura para o 4º grau – NAEP 2005

NÍVEIS DE DESEMPENHO EM LEITURA	DESCRIÇÃO
Avançado	<p>Estudantes do quarto grau, que alcançam um nível avançado, devem ser capazes de generalizar acerca de tópicos na seleção da leitura e demonstrarem uma consciência de como os autores compõem e usam artifícios literários. Quando lêem um texto adequado ao quarto grau, eles devem ser capazes de julgar criticamente os textos e, em geral, dar respostas completas que indicam pensamento cuidadoso.</p> <p>Por exemplo, quando lêem um texto literário, estudantes do nível avançado devem ser capazes de fazer generalizações acerca de um ponto da história e estender seu significado integrando experiências pessoais e outras leituras com idéias sugeridas pelo texto. Eles devem ser capazes de identificar artifícios literários tais como uma linguagem figurativa.</p> <p>Quando o texto é informativo, esses estudantes devem ser capazes de explicar a intenção do autor usando material de apoio do texto, de fazer julgamentos críticos da forma e conteúdo do texto e explicar seu julgamento claramente.</p>
Proficiente	<p>Estudantes do quarto grau que alcançam um nível proficiente devem ser capazes de demonstrar um entendimento geral do texto, indicando inferências bem como informação literal. Quando a leitura do texto é adequada a alunos do quarto grau, eles devem ser capazes de ampliar as idéias no texto, fazendo inferências, esboçando conclusões e fazendo conexões com suas próprias experiências. As conexões entre o texto e o que o estudante infere devem ser claras.</p> <p>Por exemplo, quando lêem um texto literário, estudantes do nível proficiente devem ser capazes de resumir a história, esboçando conclusões acerca dos personagens ou do enredo, e reconhecer relações de causa e efeito.</p> <p>Quando lêem um texto informativo, estudantes desse nível, devem ser capazes de resumir a informação e identificar a intenção ou o propósito do autor. Devem ser capazes de esboçar razoavelmente conclusões do texto, reconhecer relações de causa e efeito ou similaridades e diferenças e identificar o significado de conceitos principais na seleção.</p>
Básico	<p>Estudantes do quarto grau que alcançam um nível básico devem ser capazes de demonstrar um entendimento geral do que leram. Quando a leitura do texto é adequada a alunos do quarto grau, eles devem ser capazes de fazer conexões óbvias entre o texto e suas próprias experiências e estender as idéias no texto para fazer inferências simples.</p> <p>Por exemplo, quando lêem um texto literário, eles devem ser capazes de contar acerca do que a história diz respeito, proporcionando detalhes para apoiar o seu entendimento e conectar aspectos da história com suas próprias experiências.</p> <p>Quando lêem um texto informativo, estudantes desse nível, devem ser capazes de contar sobre o que o texto diz respeito ou identificar o propósito para lê-lo, proporcionando detalhes para apoiar o seu entendimento e conectar idéias do texto para seu conhecimento contextual e experiências.</p>

Fonte: U. S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Reading Assessment

Os desempenhos dos alunos em Leitura e Matemática são também apresentados na forma de mapas que posicionam as descrições dos itens ao longo da escala composta, onde são apresentados os itens respondidos com sucesso pelos estudantes. As descrições usadas no mapa focalizam sobre as habilidades necessárias para responder a cada uma das questões em cada disciplina relacionada a cada série. Para cada item do teste indicado, estudantes que escoraram acima do ponto da escala têm uma probabilidade mais alta de responder com

sucesso à questão e aqueles que escoraram abaixo do ponto considerado, têm uma probabilidade mais baixa de obter sucesso na questão. Os mapas de itens são indicadores úteis de probabilidades mais altas ou mais baixas de responder com sucesso a questões dependendo da habilidade geral dos estudantes como medida pela escala NAEP.

Em relação ao nível de desempenho dos estudantes que se encontram abaixo do nível Básico, o mapa é um formato interessante e útil para descrever o que eles são capazes, pois não se pode considerá-los como não leitores ou não capazes de desempenhar nenhuma habilidade matemática. Por exemplo, pela avaliação de Matemática, eles demonstram ser capazes de ler, mas não suficientemente bem para alcançar o escore mínimo para o nível Básico.

Pela Tabela 11, pode-se observar o que, pelas descrições apresentadas dos itens do quarto grau, os alunos que estão abaixo do nível Básico são capazes de fazer.

Tabela 11 – Mapa de descrição dos itens de Matemática para o grau 4 – NAEP 2005

	500	
	↑	
	300	
	290	294 Identificar equação para descrever padrões dados em uma tabela
	282	288 Identificar medidas dadas sobre uma régua
	280	284 <i>Subtrair frações com denominadores comuns</i>
AVANÇADO		278 <i>Aproximar fração de uma hora dada em minutos</i>
		273 Solucionar um problema descrito envolvendo números grandes (calculadora disponível)
	270	272 Determinar números ausentes em uma sentença numérica
		272 <i>Solucionar um problema descrito envolvendo multiplicação (calculadora disponível)</i>
	260	260 <i>Determinar a altura de um retângulo depois de ele ser dobrado</i>
		258 <i>Representar uma situação com uma expressão algébrica</i>
	249	254 <i>Identificar qual figura sobre uma rede quadriculada tem a maior área</i>
	250	253 Completar um gráfico de barras a partir da descrição de dados
PROFICIENTE		246 Determinar o valor de um ponto sobre uma reta numérica
	240	
		232 Determinar o próximo número em um dado padrão
	230	228 Classificar números como pares ou ímpares
		223 <i>Determinar quais atributos podem ser medidos com um metro</i>
	220	
		218 <i>Subtrair números de dois dígitos para resolver um problema descrito</i>
	214	
BÁSICO		211 <i>Identificar quais formas são cilíndricas</i>
	210	211 Subtrair números de dois dígitos de um de três dígitos
	200	203 <i>Identificar um número dado em uma notação expandida</i>
		197 <i>Determinar o resultado mais provável em um problema descrito</i>
	190	
	↓	
	0	

Fonte: U. S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Mathematics Assessment

Nota: Tipo regular de letra denota questões de resposta construída. Tipo itálico denota questões de múltipla escolha. As questões de Matemática foram mapeadas na escala de 0–500. A posição da questão na escala representa o escore alcançado por estudantes que tinham 65% de probabilidade de alcançarem sucesso nas questões de respostas construídas e para múltipla escolha, 74%.

Comparando-se os dois sistemas, respeitando-se as peculiaridades brasileiras e norte-americanas e considerando-se que o sistema de avaliação brasileiro adota como modelo o sistema norte-americano, constata-se que muito pode ser feito para melhorar o nosso sistema.

Em relação ao sistema de avaliação adotado nos dois países, percebe-se um diferencial significativo em relação às avaliações nacionais, estaduais e locais. Verifica-se, do lado americano, pouca ingerência da esfera federal nas esferas estaduais e locais. São os estados e distritos locais que administram os processos avaliativos, que são realizados com maior frequência e fornecem resultados mais pontuais, dando condições de serem implementadas políticas mais diretas e eficientes. Há exigências de cumprimento a leis federais no âmbito da educação, como é o caso do que é prescrito na re-autorização da ESEA 2001 (*No Child Left Behind*), havendo, conseqüentemente, maior comprometimento pelas instâncias diretamente relacionadas com a superação das dificuldades constatadas.

Quanto às matrizes, parece ser conveniente que o SAEB proceda à sua revisão. Afinal, os sistemas educacionais de todo o mundo têm buscado estratégias para se adaptar aos novos paradigmas de aprendizagem e o SAEB, como já discutido anteriormente, pode se constituir um eixo orientador dessa tendência.

Em relação às análises utilizadas para fornecer os resultados da avaliação, constata-se que o Brasil encontra-se muito avançado, tendo inclusive, desenvolvendo tecnologias próprias. Talvez, fosse interessante rever o delineamento amostral adotado a partir de 2002 nos Estados Unidos, que parece garantir resultados mais confiáveis aos estados. Se o esquema norte-americano para o delineamento amostral para os estados fosse adotado no Brasil, a amostra de alunos brasileiros aumentaria para cerca de 486.000 alunos⁷, considerando-se duas disciplinas avaliadas, assegurando resultados mais confiáveis no nível dos estados.

Além disso, também parece interessante, principalmente, tomando como unidade de análise a escola, a produção de escalas por temas/tópicos delimitados para as avaliações de Leitura e Matemática.

Os valores plausíveis atribuídos às proficiências dos alunos, prática comum no sistema de avaliação norte-americano, talvez pudessem minimizar o peso de alguns fatores intervenientes no desempenho dos alunos e apresentar resultados mais justos, considerando-se a realidade brasileira.

Quanto à apresentação dos resultados, sugere-se uma discussão mais profunda, com estratos de interesse da sociedade, a respeito dos qualificativos dos níveis de desempenho e dos pontos de corte relativos a cada um deles. O mapa de descrição dos itens também se constitui uma forma interessante de apresentação de resultados em algum tipo de relatório.

⁷ A amostra consideraria 30 alunos por escola, 100 escolas por estado, duas disciplinas e três séries avaliadas.

Após essa pesquisa parece razoável considerar que o sistema brasileiro de avaliação pode ser repensado. Vários comentários foram feitos no decorrer deste capítulo que apóiam esse posicionamento. No capítulo a seguir, é feita uma análise mais profunda do sistema brasileiro, no que diz respeito aos testes de desempenho e às descrições das escalas produzidas com base nos resultados, ao longo de cinco edições do SAEB (1995 a 2003).

5. ANÁLISE CRÍTICA DOS TESTES NA PERSPECTIVA DAS DESCRIÇÕES DAS ESCALAS NO PERÍODO 1995-2003

Neste capítulo, é destacada a importância do teste enquanto instrumento utilizado para avaliar o desempenho do aluno no processo de descrição dos resultados. Como já foi exposto anteriormente, a eficácia de uma avaliação educacional depende, sobretudo, do teste elaborado para conferir o desempenho dos alunos e de sua eficiência para medir, avaliar e descrever os resultados observáveis da aprendizagem. O teste se constitui, dessa forma, o referencial das habilidades aferidas. Assim, em um processo nacional, ele será o único e o principal instrumento confirmatório do que os alunos de toda a nação demonstram ser capazes de fazer ou não.

É necessário ser transparente e honesto ao devolver os resultados. Sabe-se que inúmeros fatores contribuem para a fidedignidade ou não dos resultados. Eles vão interferindo na sucessão de operações que compõem o processo, incluindo desde o planejamento, o delineamento amostral, a aplicação, a correção, as próprias análises, até chegar à descrição dos resultados. Há que se ter um rigor extremado por todo o processo para que a avaliação alcance os objetivos planejados.

Este tópico da pesquisa teve por finalidade chamar a atenção para o que se apresenta como resultado do que foi avaliado. Como já fora dito ao longo deste estudo, os resultados devem estar integralmente relacionados ao instrumento utilizado. Eles devem ser relatados de forma clara o suficiente para que possam ser entendidos e assimilados por todos os interessados na melhoria da qualidade do sistema educacional, sejam elaboradores de

políticas, secretários de educação, diretores de escolas, professores, jornalistas, pais, enfim quem quer que seja.

Para realizar a análise crítica da descrição dos resultados, enquanto habilidades aferidas e publicadas na forma de escalas e níveis de desempenho, foram necessários dois procedimentos distintos: (1) (re)descrever as escalas, com base nos posicionamentos adotados ao longo deste estudo e (2) analisar criticamente as descrições publicadas.

Em síntese, para realizar essa análise, examinou-se a qualidade psicométrica e pedagógica de cada item e de cada teste, selecionando os itens adequados para compor a descrição do que os alunos demonstraram saber fazer. Alguns itens foram considerados na descrição, mesmo não tendo obedecido aos critérios determinados. Estes apresentaram boa qualidade pedagógica. O número de itens que não atingiu esses critérios, mas que foram considerados na análise são os seguintes: correlação bisserial – sete itens; parâmetro a – cinco itens; e parâmetro c – 22 itens. Em seguida, procedeu-se à análise do teste, apontando os problemas encontrados em cada um deles. O terceiro passo foi descrever as habilidades relacionadas a cada item selecionado. As habilidades foram organizadas em tabela conforme os níveis a elas atribuídos, resultantes das análises realizadas pela Fundação Cesgranrio. E, por fim, procedeu-se às descrições na forma de uma escala que refletisse o que, de fato, os alunos demonstraram saber fazer tomando por base o desempenho no teste (ver modelos dos procedimentos nos Anexos A1, A2 e A3).

A (re)descrição das escalas foi feita com base no referencial teórico subjacente à TRI, ou seja, pressupõe-se que um aluno que tenha alcançado uma proficiência mais alta sabe mais que outro aluno que tenha alcançado uma proficiência mais baixa, em um processo cumulativo. Portanto, um item representativo de um intervalo 175-200 exige um nível de habilidade menor do que outro item do intervalo 200-225. Dessa forma, ao descrever o que os alunos são capazes, esse diferencial deve ficar explícito. Por exemplo, se se considera a competência “*localizar informação explícita em um texto*”, os itens representativos dos diferentes intervalos da escala, deverão apresentar habilidades que demonstrem a construção dessa competência, isto é, deve ficar explícito que o seu alcance ocorre por meio do processo de desenvolvimento de habilidades, das mais simples às mais complexas.

Realizar essa (re)descrição e criticar as descrições publicadas foi um grande desafio, posto que, se considerou o olhar de apenas uma pessoa, que embora esteja fundamentada nos estudos que ora apresenta e que tenha experiência como professora de séries iniciais, como

avaliadora e como pesquisadora, não desfrutou da interlocução com especialistas nas áreas tomadas como referência (língua portuguesa e matemática).

Na Tabela 12 constam alguns dados descritivos das avaliações consideradas neste estudo.

Tabela 12 - Informações relacionadas aos testes analisados (SAEB)

TESTES	Nº DE QUESTÕES ANALISADAS	ALUNOS AVALIADOS
Língua Portuguesa 1995	130	12.033
Matemática 1995	130	11.886
Língua Portuguesa 1997	143	23.404
Matemática 1997	143	23.535
Língua Portuguesa 1999	169	21.542
Matemática 1999	169	21.572
Língua Portuguesa 2001	169	57.254
Matemática 2001	169	57.258
Língua Portuguesa 2003	169	46.067
Matemática 2003	169	46.131
Total	1.560	320.682

As discussões serão feitas com base nos dados analisados por edição do SAEB, publicados a partir de 1995. Ao final, serão feitas algumas reflexões sobre os propósitos da escala e o que poderia ser feito para melhorar esse processo de descrição.

5.1. SAEB – EDIÇÃO 1995

5.1.1. AVALIAÇÃO DE LÍNGUA PORTUGUESA

O teste de Língua Portuguesa do SAEB 1995 teve por foco a leitura (Resumo Executivo do SAEB 1995, 1997). Os itens foram elaborados a partir de uma matriz referencial (não constante em nenhum documento na DAEB) que especificava os conteúdos e habilidades a serem avaliados. Em síntese, segundo consta nesse Resumo Executivo, as habilidades de leitura previstas para serem avaliadas em toda a educação básica, considerando-se as duas séries do final do ciclo do ensino fundamental (4ª e 8ª) e a do final do ensino médio (3ª),

foram classificadas em três categorias: (1) estabelecimento de significado; (2) extensão do significado; e (3) exame crítico do significado. Para cada uma dessas categorias foram relacionadas habilidades. Foram dadas as seguintes definições para cada uma das categorias (p. 3):

- 1) **Estabelecimento de significado:** é a consolidação, motivada por aprendizagem, da relação entre uma palavra e a idéia a que ela se refere, que corresponde a uma denotação.
- 2) **Extensão do significado:** é a aquisição de outros significados pela palavra original, que corresponde a uma conotação.
- 3) **Exame crítico do significado:** corresponde a um exercício de crítica, ou seja, um exercício em que a coerência do aluno é convocada para a construção de um juízo.

O teste foi constituído de 130 questões referentes a 48 textos. Dessas questões, nesse estudo, apenas 68, ou seja, 52% foram consideradas para proceder à descrição do que os alunos brasileiros sabem ou deveriam saber fazer. Em síntese, pôde-se verificar que, em relação ao teste, os 48 textos, de forma geral, foram muito pobres. Em sua maioria, foram constituídos de um a dois parágrafos (dos 24 textos informativos, 17 eram de frases curtas com no máximo três linhas; oito textos literários; uma propaganda, duas instruções, três frases para completar com termos gramaticais; duas receitas; duas tabelas; e seis bilhetes). O gênero literário foi contemplado em apenas 17% e o informativo foi o mais adotado (50%), sendo, estes, basicamente constituídos de pequenos trechos de informações veiculadas em jornais. Houve várias questões e textos com linguagem inadequada para a faixa etária avaliada. Faltou plausibilidade à maioria das alternativas dos itens. O foco não foi somente leitura, pois vários itens cobraram sinonímia, concordância verbal e nominal, entre outros. Observaram-se também diversas questões com conteúdos repetidos e com o mesmo nível de dificuldade, o que impediu a abrangência de conteúdo representativo do que os alunos brasileiros deveriam saber fazer. Houve, ainda, entre as questões, as que apresentaram cinco alternativas, o que, certamente, causou confusão aos alunos.

Retirando os 62 itens que não foram adequados ou apropriados (apresentaram problemas pedagógicos ou psicométricos) para avaliarem os alunos, a distribuição do conteúdo, comparada aos tópicos da atual matriz 2001, ficou da seguinte forma: Procedimentos de leitura: 72% (49 itens); Implicações do suporte, do gênero e/ou enunciador

na compreensão do texto: 10% (7 itens); Relação entre textos: 3% (2 itens); Coerência e coesão no processamento de texto: 12% (8 itens); Relações entre recursos expressivos e efeitos de sentido: 1% (1 item); Variação lingüística: 1% (1 item).

Na Tabela 13 são apresentadas as estatísticas geradas dos itens selecionados para a descrição da escala de proficiência dos alunos em Língua Portuguesa, 4ª série, 1995. Essa seleção foi feita considerando-se os critérios descritos na metodologia, tomando por base os dados cedidos pelo INEP, com as referidas análises psicométricas realizadas pela Fundação Cesgranrio⁸.

Tabela 13 - Estatísticas - Teste de Língua Portuguesa - 4ª Série - SAEB - Edição 1995

	Parâmetro a	Parâmetro b	Parâmetro c	Discriminação Índice D	Proporção de acerto
N	68	68	68	68	68
Média	1.39	-1.65	0.13	0.55	0.61
Desvio Padrão	0.49	0.95	0.09	0.10	0.17
Mínimo	0.58	-3.45	0.00	0.30	0.25
Máximo	2.83	0.11	0.40	0.71	0.89

Na Tabela 14 é apresentada a (re)descrição dos níveis de proficiência que representa o que os alunos brasileiros demonstraram saber fazer, com base no teste aplicado, e o percentual de alunos da 4ª série, que atingiu cada um desses níveis. Essa (re)descrição foi feita com base nas informações cedidas pelo INEP, as quais apontavam o nível referente à cada item. A seleção desses itens para a (re)descrição foi feita com base nos critérios estabelecidos na metodologia deste estudo.

Tabela 14 – Descrição e percentual de alunos nos níveis com base no teste de Língua Portuguesa - 4ª Série - 1995

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$\theta < 125$ 9,1%	Não houve itens para descrever este nível da escala.
$125 \leq \theta < 150$ 11,5%	Alunos deste nível de proficiência são capazes de localizar uma informação explícita nas primeiras duas palavras do texto; identificar o sinônimo de uma palavra comum do vocabulário (atleta) e identificar o gênero textual receita;
$150 \leq \theta < 175$ 18,9%	Alunos deste nível são capazes de localizar informações explícitas no início de textos curtos informativos e em tabelas; identificar um texto como do gênero poesia; comparar o que está dito em dois textos à vivência pessoal; identificar marcas lingüísticas que evidenciam o locutor e o interlocutor de um bilhete e inferir o sentido de uma palavra comum do vocabulário deles em um texto.
$175 \leq \theta < 200$ 21,1%	Neste nível, os alunos demonstram ser capazes de localizar informações explícitas no meio de textos simples e início de textos mais complexos; inferir informações implícitas e o sentido de uma expressão em textos com vocábulos simples; identificar o gênero textual do texto (bilhete).
$200 \leq \theta < 225$ 18,4%	Neste nível de proficiência, os alunos são capazes de localizar informações explícitas parafraseadas e que se encontram no final de textos simples; inferir o sentido de uma palavra, de uma expressão e uma informação implícita pelos contextos dos textos; reconhecer a consequência de uma instrução não seguida;

⁸ Todas as análises foram feitas somente com os itens selecionados e previamente analisados.

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	interpretar uma informação com base em uma ilustração gráfica; identificar a fala de um dos personagens de uma fábula; identificar um dos motivos de um bilhete.
225 ≤ θ < 250 11,6%	Os alunos que atingiram esse nível de proficiência são capazes de localizar informação explícita parafraseada em um texto mais complexo; inferir sobre uma informação implícita em quadrinhos e no meio de textos complexos; identificar o tema ou o assunto principal de um texto; interpretar texto com auxílio de material gráfico mais abstrato (quadrinhos); identificar o gênero de um texto (informativo – enciclopédico); estabelecer conclusão com base na informação de outro texto; estabelecer relação causa/conseqüência entre partes do texto; reconhecer o efeito de sentido decorrente do uso de um ponto de exclamação (espanto).
250 ≤ θ < 275 5,5%	Neste nível os alunos demonstram capacidade para localizar uma informação explícita no final de um texto fabular; inferir sobre uma informação implícita do texto relacionada ao tema e de um bilhete que contém diversas informações; identificar o assunto principal de um bilhete com muitas informações; identificar o gênero do texto (informativo – notícia de jornal); estabelecer relação causa/conseqüência entre partes e elementos do texto com base em sua compreensão global; identificar característica da personagem principal de um texto segundo opinião do narrador.
275 ≤ θ < 300 2,6%	Alunos deste nível são capazes de localizar informação explícita em um texto informativo de um parágrafo, cujo enunciado é colocado de modo positivo e cuja informação é dada de forma negativa; identificar uma substituição de referência (deles) por meio do estabelecimento de relações entre partes de um texto.
θ > 300 1,2%	Não houve itens para descrever este nível da escala.

Nota: Os dados cedidos pelo INEP permitiram a descrição da escala em intervalos menores e o fornecimento de informações sobre os percentuais de alunos por intervalo foi feito com base em cálculo feito pelos dados cedidos a essa pesquisadora.

A escala publicada referente à Língua Portuguesa, da edição de 1995, foi descrita no formato único, considerando-se os testes de Língua Portuguesa aplicados nos alunos nas três séries avaliadas. As descrições (p. 5) são apresentadas na Tabela 15.

Tabela 15 – Níveis de proficiência de Língua Portuguesa publicados e percentuais de alunos nos níveis - SAEB - 1995

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
θ < 150 20,6%	Não há itens que descrevam este nível.
150 ≤ θ < 225 58,4%	Os alunos mostram compreensão literal e leitura fragmentária; reconhecem informações e pormenores e conseguem fazer seleções de informações simples; fazem interpretação de caráter limitado: estabelecem relações diretas entre fatos, idéias e conceitos sem usar suporte inferencial abstrato.
225 ≤ θ < 300 19,7%	Abordam textos extensos, sem perder significado e lidam com textos práticos (telegrama, circular, manual de instrução) fazendo inferências a partir de informações contidas no texto; percebem relações entre fatos principais e secundários e identificam características específicas de personagens; trabalham com tipologia textual variada estabelecendo relação entre fatos e temas comuns.
300 ≤ θ < 375 1,2%	Trabalham com desenvoltura com variada tipologia de textos, sendo capazes de integrar entre si, partes de textos diversos; sintetizam a idéia principal a partir da análise do conjunto e não por detecção em fragmentos; relacionam informações com sua vivência pessoal, ampliando a compreensão do texto; fazem inferências tanto textuais como extratextuais, construindo relações lógicas e temporais, relacionando texto e contexto; são capazes de depreender o objetivo do autor.
θ ≥ 375 0,0%	São capazes de ler o figurativo complexo, localizando a seqüência lógica, detectando similaridades e diferenças e estabelecendo relações de causa e feito; mostram domínio de vocabulário e conceitos mais complexos, associando idéias no sentido denotativo e conotativo; diferenciam termos específicos de textos informativos e jornalísticos; percebem e distinguem a crítica, a ironia e o humor contidos na linguagem.

Fonte: MEC/INEP²

Nota: A descrição publicada tomou por base intervalos maiores para os níveis de proficiência e, embora a descrição seja referente a todos os alunos da educação básica, o percentual de alunos é referente exclusivamente aos níveis de proficiência dos alunos da 4ª série, calculados por essa pesquisadora.

Considerando-se a Tabela 15, constata-se um excesso de redução do que poderia ser descrito em relação aos níveis de proficiência que representam o que os alunos brasileiros demonstraram saber fazer, com base nos testes aplicados. Além da redução, observa-se também uma linguagem distante da praticada por professores de Língua Portuguesa, mesmo considerando-se especialistas. Com certeza, mesmo que os resultados dessa avaliação chegassem às escolas, estaria comprometida a identificação do professor à sua prática pedagógica.

5.1.2. AVALIAÇÃO DE MATEMÁTICA

O teste de Matemática do SAEB 1995 abrangeu cinco áreas de conteúdo: Números e operações; Medidas; Geometria; Análise de dados, Estatística e Probabilidade; e Álgebra e funções (Resumo Executivo do SAEB 1995, 1997). Os itens foram elaborados a partir da matriz referencial (não constante em nenhum documento na DAEB) que especificava esses conteúdos e habilidades a serem avaliados. As habilidades relacionadas a esses conteúdos foram classificadas em três categorias: (1) compreensão de conceitos; (2) conhecimento e utilização de procedimentos; (3) aplicação ou resolução de problemas. Foram dadas as seguintes definições para cada uma das categorias (p. 4):

- 1. Compreensão de conceitos:** nessa categoria, segunda consta no documento, as habilidades a ela associadas devem responder à pergunta: **O que é?** Devem-se utilizar conceitos termos, definições e princípios.
- 2. Conhecimento e utilização de procedimentos:** nessa categoria, os procedimentos relacionam-se à pergunta: **Como se faz?** Esses procedimentos envolvem cálculos, formulações, construções, estimações, execução de algoritmos, manipulações algébricas.
- 3. Aplicação ou resolução de problemas:** nessa categoria, a aplicação corresponde à pergunta: **Para que serve?** As aplicações referem-se à provas, inferências, solução de problemas expressos em linguagem natural e aplicações no mundo real.

O teste foi constituído de 130 questões das quais apenas 76 foram consideradas para proceder à descrição do que os alunos brasileiros sabem ou deveriam saber fazer. Em síntese, analisando-se o teste pôde-se verificar que não foi apresentada nenhuma questão contextualizada, além disso, cerca de 95% do teste contempla questões de reconhecimento, ou seja, para se chegar à resolução, só exigiu-se memorização. Neste teste, também foram

apresentadas várias questões muito semelhantes e com o mesmo nível de dificuldade, tornando mais difícil, dessa forma, abranger conteúdo representativo do que os alunos brasileiros deveriam saber fazer.

Quanto à distribuição do conteúdo, tomando-se como referência a matriz atual (2001), constatou-se que 86% dos itens do teste dizem respeito ao tema números e operações; 8% ao tema espaço e forma; 5% ao tema grandezas e medidas; e, 1% diz respeito ao tema tratamento da informação.

Em relação à qualidade pedagógica, pôde-se constatar que houve várias questões inadequadas para o nível de escolaridade avaliado (44 questões). Da mesma forma que em língua portuguesa, foram apresentadas questões com cinco alternativas e com conteúdos de 8ª série.

Na Tabela 16 são apresentadas as estatísticas geradas dos itens selecionados para a descrição da escala de proficiência dos alunos em Matemática, 4ª série, 1995. Essa seleção foi feita considerando-se os critérios descritos na metodologia, tomando por base os dados cedidos pelo INEP, com as referidas análises psicométricas realizadas pela Fundação Cesgranrio.

Tabela 16 - Estatísticas - Teste Matemática - 4ª Série - SAEB - Edição 1995

	Parâmetro a	Parâmetro b	Parâmetro c	Discriminação Índice D	Proporção de acerto
N	76	76	76	76	76
Média	1,67	-0,69	0,17	0,54	0,44
Desvio Padrão	0,59	0,94	0,10	0,11	0,16
Mínimo	0,62	-3,24	0,01	0,29	0,16
Máximo	3,75	1,09	0,50	0,75	0,83

A seguir, são apresentadas as (re)descrições dos níveis de proficiência que representam o que os alunos brasileiros demonstraram saber fazer, com base no teste aplicado, e o percentual de alunos da 4ª série, que atingiu cada um desses níveis.

Tabela 17 – Descrição e percentual de alunos nos níveis com base no teste de Matemática - 4ª Série - 1995

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$\theta < 150$ 15,8%	Sem itens para descrever este nível da escala.
$150 \leq \theta < 175$ 23,2%	Alunos desse nível de proficiência são capazes de identificar o menor número natural entre quatro da ordem das unidades de milhar; de somar frações de mesmo denominador; e de identificar uma unidade de medida de massa (tonelada).
$175 \leq \theta < 200$ 24,2%	Neste nível, os alunos são capazes de reconhecer uma fração em desenho representando relação entre parte-todo; comparar tamanhos de segmentos de reta e identificar o menor; reconhecer um número natural escrito por extenso, pela interpretação do valor posicional de cada uma das ordens (número da ordem das dezenas de milhares e com um zero intermediário); indicar o valor da incógnita de uma expressão, dadas quatro alternativas; armar e efetuar uma soma de duas parcelas, com reserva; identificar a unidade de

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	medida utilizada para pesar um objeto (grama).
200 ≤ θ < 225 17,7%	Neste nível de proficiência, os alunos são capazes de armar e efetuar uma subtração, com recurso; calcular uma multiplicação envolvendo número racional; reconhecer que um quadrilátero partido ao meio gera duas partes iguais; reconhecer horas e minutos em relógios de ponteiros; resolver problema envolvendo soma de frações de mesmo denominador; estabelecer trocas de moedas em função do valor.
225 ≤ θ < 250 10,8%	Os alunos que atingiram esse nível de proficiência são capazes de decompor um número natural em sua forma polinomial (dezena de milhar); identificar, por meio de figuras, relação parte-todo e corresponderem números inteiros e fração a número fracionário; calcular o resultado de multiplicações por 10 e por 100; calcular o resultado de uma divisão de números naturais; ordenar números fracionários e calcular o resultado de uma subtração de frações com o mesmo denominador; armar e efetuar uma subtração com recurso, tendo um zero intercalado no subtraendo; identificar um círculo como uma curva fechada simples; resolver um problema de conversão de horas em dia, dias em semana e horas por semana; decompor um número pelos valores posicionais; identificar um ponto no interior de uma figura fechada; resolver expressões simples com uma incógnita e substituindo figuras por números; ler um gráfico de colunas, interpretá-lo e calcular o resultado da soma de todas as colunas.
250 ≤ θ < 275 4,8%	Neste nível, os alunos demonstram capacidade para identificar a característica de paralelismo em uma figura plana; calcular o quociente e o resto de uma divisão de um algarismo da ordem das centenas por um da ordem das dezenas; armar e efetuar adição de quatro parcelas, envolvendo números racionais, de diferentes ordens decimais; realizar trocas entre cédulas e moedas; estabelecer regras para seqüência de números correlacionados em duas colunas; identificar, entre unidades de medida de comprimento, a mais adequada para medir a área de um território (km ²); calcular o resultado de uma divisão de um algarismo da ordem das unidades de milhar, com um zero intermediário, por um da ordem das dezenas; indicar os fatores que podem constituir um número da ordem das dezenas; resolver problema que envolve soma e divisão; calcular o resultado de uma subtração, com base na propriedade distributiva; reconhecer o significado de frações como quocientes; identificar equivalência de frações, envolvendo inteiros.
275 ≤ θ < 300 2,2%	Alunos desse nível são capazes de identificar um pentágono pelo número de lados; simplificar frações; armar e efetuar adição de números mistos (natural e racional - decimal, centesimal, milésimo) e multiplicar um número da ordem das centenas, sendo que um deles possui um zero intermediário; dividir um número da ordem das unidades de milhar por um da ordem das dezenas; corresponder a escrita de um número decimal à escrita de uma fração; reconhecer: (1) a ordem crescente e decrescente de uma seqüência numérica alternada de algarismos da ordem das centenas e das unidades de milhar; (2) uma regra para uma seqüência numérica; (3) a propriedade da multiplicação de que a ordem dos fatores não altera o produto; (4) números ímpares numa determinada seqüência de algarismos das unidades de milhar; e associar os dados de uma tabela a um gráfico;
θ > 300 1,2%	Sem itens para descrever este nível da escala.

Da avaliação de 1995 foram produzidas escala única para a disciplina Matemática. Assim, na Tabela 18 são apresentadas as descrições para a educação básica (p. 6), com os respectivos percentuais de alunos nos níveis.

Tabela 18 – Níveis de proficiência de Matemática publicados e percentuais de alunos nos níveis - SAEB - 1995

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
θ < 150 15,8%	Não há itens que descrevam este nível.
150 ≤ θ < 225 65,2%	Identificam grandezas; reconhecem uma representação gráfica de números fracionários; adicionam frações com o mesmo denominador; constroem uma fração, dados o seu numerador e denominador; efetuam adições de dois números naturais de até cinco algarismos; conhecem unidade de medida de massa e de comprimento; resolvem problemas concretos envolvendo unidades monetárias.
225 ≤ θ < 300 17,8%	Demonstram conhecimento consolidado das quatro operações com números naturais e resolvem problemas simples, aplicando-as; reconhecem regras que relacionam duas seqüências de números inteiros; identificam termos desconhecidos em sentenças matemáticas; ordenam números fracionários em ordem crescente e decrescente; operam com números fracionários; resolvem problemas simples envolvendo

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	frações; operam com números decimais; sabem efetuar conversões com medidas de comprimento; sabem selecionar unidades apropriadas para medidas de superfície; relacionam dados de tabelas simples com gráficos de setores; interpretam gráficos de barras; conhecem unidades de tempo e efetuam suas conversões.
$300 \leq \theta < 375$ 1,2%	Dominam os algoritmos e os conceitos envolvendo as quatro operações e sabem aplicá-las na resolução de problemas de mais de um passo; reconhecem o valor posicional relativo de um algarismo e sabem aplicar este conceito para obter o maior número que se pode formar com algarismos dados; reconhecem frações equivalentes, lidam com frações em forma decimal e sabem empregar aproximações para inteiros para facilitar o cálculo com decimais; operam com números inteiros relativos, utilizam regra de três e calculam porcentagens; identificam leis de formação de seqüências formadas por inteiros ou por figuras geométricas; resolvem equações simples do 1º grau e aplicam propriedades simples de números em progressão aritmética; sabem operar com monômios; discriminam ângulos agudos, retos e obtusos; identificam ângulos opostos pelo vértice como congruentes; reconhecem retas paralelas, aplicam proporcionalidade em triângulos semelhantes ou em retas cortada por um feixe de paralelas; interpretam gráficos cartesianos; relacionam dados envolvendo porcentagem com gráficos de setores; estimam volumes por meio de cálculos e efetuam conversões de unidades de medida de massa e as aplicam em problemas simples.
$\theta \geq 375$ 0,0%	Sabem posicionar um número fracionário positivo entre dois inteiros; efetuam expressões com frações envolvendo as quatro operações e sabem efetuar divisões de números decimais; resolvem problemas simples envolvendo frações e porcentagens; sabem calcular aproximação inteira de raiz quadrada não exata de números inteiros; aplicam propriedades simples de progressões geométricas; constroem matrizes (ou seqüências bivariadas) através de lei recursiva de formação; resolvem sistemas de 1º grau; lidam com polinômios; sabem fatorar expressões algébricas e aplicá-las na simplificação de frações algébricas; simplificam expressões envolvendo radicais; dominam os conceitos básicos relativos a equações do 2º grau e trabalham com funções simples; aplicam a definição de logaritmo; resolvem problemas simples envolvendo probabilidade; resolvem problemas envolvendo ângulos, áreas e volumes de figuras simples; operam com conceitos simples relativos a círculo, tangente e ângulo de um setor.

Fonte: MEC/INEP

Em relação à escala de matemática publicada, embora ela tenha tido uma descrição mais detalhada do que os alunos demonstram saber fazer com base nos testes aplicados, o fato de ter uma descrição única, como em Língua Portuguesa, dificulta o professor, principalmente de séries iniciais, se identificar com sua prática pedagógica.

É importante considerar que os dados que tiveram por base essa avaliação continuam sendo apresentados e comparados aos atuais. Após análise desses testes (língua portuguesa e matemática) e das escalas, constata-se que o instrumento mostrou-se muito diferente das atuais avaliações. As orientações para a construção dos itens e as matrizes de referência eram outras. É difícil compatibilizar os itens aos atuais descritores. Pode-se apenas indicar os temas aos quais se referem os itens de 1995 aos de 2001.

A análise dos testes apresentou problemas com textos (curtos, sem sentido, sem referências), com elaboração dos itens, com a abrangência dos conteúdos e com o próprio formato do teste.

Percebe-se que os testes referentes a essa edição são compostos por itens que, em sua maioria, cobram memorização. Eles reproduzem a prática pedagógica tradicional de transmitir conteúdos sem contextualização e se assemelham àqueles contemplados na sala de aula por professores partidários dessa prática.

A configuração de um teste para avaliação em larga escala deve ser outra. Por meio dos testes, deve-se querer investigar, sobretudo, se o aluno é capaz de aplicar o que aprendeu, após um período de escolarização, na resolução de problemas cotidianos. Trata-se de responder à questão: A escola está sendo eficaz na formação do cidadão?

Pode-se assegurar que os testes de Língua Portuguesa e Matemática de 4ª série, referentes a essa edição, avaliaram conteúdos sem significado prático, no exercício da cidadania. Dessa forma, considera-se desaconselhável continuar comparando os atuais resultados da avaliação com os dessa edição, pois os testes e os referenciais são bastante diversos.

5.2. SAEB – EDIÇÃO 1997

No documento Matrizes Curriculares de Referência para o SAEB, 1997, p. 7, consta que a elaboração da matriz foi orientada pela *busca do estabelecimento dos conteúdos desejáveis e necessários às demandas e exigências implícitas no sistema educacional, considerando as diferenças regionais*. A esses conteúdos, foram associadas competências cognitivas e habilidades instrumentais. As competências foram categorizadas em três níveis de ações e de operações mentais que se estabelecem conforme as relações entre o sujeito e o objeto do conhecimento. Assim, foram determinados três níveis: básico ou presentativo, operacional ou procedural e global ou operatório.

No nível básico estão as ações que tornam o objeto do conhecimento presente para o sujeito. Essas ações podem ser descritas pelas seguintes atividades: apontar, constatar, descrever, discriminar, identificar, indicar, localizar, ler, nomear, observar, perceber, posicionar, representar.

No nível operacional estão as ações que estabelecem relações com e entre os objetos do conhecimento e o sujeito. As ações podem ser representadas pelas seguintes atividades: associar, comparar, conservar, compor, compreender, correlacionar, decompor, diferenciar, estabelecer, estimar, incluir, interpretar, medir, ordenar, organizar, relacionar, transformar.

No nível global estão as ações que envolvem a aplicação de conhecimentos e a resolução de problemas pelo sujeito. As seguintes ações representam esse nível: analisar, antecipar, avaliar, aplicar, abstrair, construir, criticar, concluir, supor, deduzir, explicar, generalizar, inferir, julgar, prognosticar, resolver e solucionar.

5.2.1. AVALIAÇÃO DE LÍNGUA PORTUGUESA

A matriz curricular de Língua Portuguesa teve por base o estabelecimento de um conjunto de saberes significativos para os alunos dos três ciclos de escolarização avaliados, os quais deveriam privilegiar o conhecimento lingüístico operacional, ou seja, ações que se fazem com e sobre a linguagem, e as implicações culturais decorrentes do uso social da língua em uma sociedade complexa. Essa escolha decorre do fato de que os conhecimentos constituídos por um indivíduo, aqueles advindos da escola e da vivência de mundo, articulam-se entre o seu grau de letramento e o grau de letramento do ambiente do qual faz parte (p. 11). Ao se privilegiar o conhecimento lingüístico operacional, enfatiza-se que o texto é elemento essencial na constituição do teste. Os itens são sempre referenciados a um texto, e é por meio dele que se determina o grau de complexidade dos itens.

O teste de Língua Portuguesa do SAEB 1997 foi dividido em duas instâncias, consideradas complementares: a leitura e a produção de textos. A orientação foi que a produção de textos fosse necessariamente dissertativa, abordando conhecimentos lingüísticos e aspectos que se esperava que fossem desenvolvidos pelos alunos na escola. Dessa forma, os testes elaborados com base nas Matrizes Curriculares de Referência para o SAEB 1997 abrangeram três áreas de conteúdos, conforme pode ser observado na Tabela 19.

Tabela 19 - Conteúdos da Matriz de Referência de Língua Portuguesa 1997

Áreas	Subáreas	Descritores
Práticas de leitura de textos	Estratégias de leitura determinadas pelos diferentes objetivos da própria leitura	15
	Articulação, texto e contexto	3
	Utilização de mecanismos básicos de coesão no processamento do texto	6
	Relações na progressão temática do texto	4
	Relações entre recursos expressivos e efeitos de sentido	8
Análise e reflexão lingüística	Texto e gênero	7
	Variação lingüística	4
	Operações lingüísticas de estabelecimento de relações entre forma e sentido	6
	Paradigmas lingüísticos (regras, classes)	6
Práticas de produção de textos	Tipos de textos	0
	Habilidades observadas	0

A Matriz, embora represente uma proposta inovadora na elaboração dos itens, parece não ter sido submetida a professores de sala de aula. Observa-se que a linguagem adotada

para explicitar o que se espera avaliar nos alunos de 4ª série é muito rebuscada, distanciando o professor que trabalha com as séries iniciais, dos quais não é exigida formação na área específica de Letras. Para exemplificar essa complexidade, pode-se citar o descritor 24: “*identificar o antecedente próximo de uma elipse do sujeito*”; ou o 32: “*analisar o efeito de sentido conseqüente do uso de um recurso prosódico*”.

O teste foi constituído de 143 questões das quais apenas 53 (37%) foram consideradas para proceder à descrição do que os alunos brasileiros sabem ou deveriam saber fazer. Em síntese, analisando-se a qualidade do instrumento, pôde-se verificar que a maioria dos textos foi de temas inadequados a alunos desse ciclo de escolarização. Além disso, o teste apresentou poucas questões que diziam respeito à realidade contextual dos alunos brasileiros e abrangeu apenas conteúdos representativos do que os alunos aprendem na sala de aula, ou seja, faltaram questões que abordassem aplicações práticas da finalidade do ensino da língua depois de um período de escolarização. Novamente, foram apresentadas várias questões com conteúdos repetidos e com o mesmo nível de dificuldade, impedindo que outros níveis de proficiência fossem avaliados.

Em relação às questões houve várias formuladas de modo inadequado para o nível de escolaridade avaliado. Houve também mudança dos intervalos considerados na métrica da escala e dos critérios para determinação dos níveis representativos dos itens. Com isso, a posição de vários itens comuns passou para a posição seguinte, mostrando que aparentemente, seria necessária uma proficiência maior para respondê-los, quando na verdade, a invariância dos parâmetros é um pressuposto da TRI que deve ser mantida nesse tipo de avaliação.

De modo geral, retirando os itens que não foram adequados ou apropriados para avaliarem os alunos (apresentaram problemas pedagógicos ou psicométricos), pode-se dizer que, entre os itens de múltipla escolha, o teste abrangeu um alto percentual de itens relacionados ao tópico “Práticas de leitura de textos” (98%), e apenas 2% sobre “Análise e reflexão lingüística”.

Em relação aos itens de resposta construída, apenas oito tiveram esse caráter. A intenção inicial era que a produção de textos fosse necessariamente dissertativa, abordando conhecimentos lingüísticos e aspectos que se esperava que fossem desenvolvidos por meio da escolarização. Não foram informadas a essa pesquisadora quaisquer análises relacionadas a esses itens. Mas, observando-os, constatou-se que apenas uma questão “*criar um texto para um cartaz proibindo o estacionamento de carros para facilitar o trânsito de pedestres*”,

abordou uma produção própria de texto. Todas as outras consistiram basicamente de reescrever frases, dar o significado de palavra e expressão, entre outras.

Na Tabela 20 são apresentadas as estatísticas geradas dos itens selecionados para a descrição da escala de proficiência dos alunos em Língua Portuguesa, 4ª série, 1997. Essa seleção foi feita considerando-se os critérios descritos na metodologia, tomando por base os dados cedidos pelo INEP, com as referidas análises psicométricas realizadas pela Fundação Cesgranrio.

Tabela 20 - Estatísticas - Teste Língua Portuguesa - 4ª Série - SAEB - Edição 1997

	Parâmetro a	Parâmetro b	Parâmetro c	Discriminação Índice D	Correlação bisserial	Proporção de acerto
N	53	53	53	53	53	53
Média	1,46	-1,25	0,15	0,51	0,55	0,53
Desvio Padrão	0,48	0,84	0,14	0,09	0,08	0,14
Mínimo	0,58	-3,24	0,00	0,31	0,38	0,25
Máximo	2,59	-0,05	0,40	0,71	0,69	0,81

Na Tabela 21 consta a (re)descrição dos níveis de proficiência que representa o que os alunos brasileiros demonstraram saber fazer, com base no teste aplicado, e o percentual de alunos da 4ª série, que atingiu cada um desses níveis.

Tabela 21 – Descrição dos percentuais de alunos e dos níveis com base no teste de Língua Portuguesa - 4ª Série - 1997

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$\theta < 100$ 1,4%	Não houve itens que descrevessem esse nível.
$100 \leq \theta < 175$ 43,0%	Os alunos demonstraram ser capazes de comparar o que está dito em dois textos à vivência pessoal e localizar uma informação explícita em um texto informativo curto.
$175 \leq \theta < 250$ 45,9%	Alunos da 4ª série que atingiram esse nível de proficiência demonstraram ser capazes de: <ul style="list-style-type: none"> • localizar informações explícitas no início ou ao longo de textos que sejam curtos e com vocábulos de uso mais comum como alguns poemas infantis, receitas, lendas; • inferir uma informação implícita e o sentido conotativo e denotativo de uma palavra ou expressão em textos simples, curtos e diretos; • reconhecer: (1) características exclusivas do poema (repetição de palavras para dar idéia de ritmo ou musicalidade); (2) personagens comuns de livros infantis (bruxa e fada); (3) a consequência de uma instrução não seguida; (4) a idéia de dúvida expressa pelo tempo verbal futuro; (5) elementos de linguagem figurada (onomatopéia); (6) marcas linguísticas que caracterizam a fala de uma criança; • concluir uma idéia com base na informação de outro texto; • interpretar uma informação com base em uma ilustração gráfica (foto); • estabelecer relações entre partes do texto identificando substituições que contribuem para a continuidade do texto.
$250 \leq \theta < 325$ 9,4%	Neste nível de proficiência, os alunos de 4ª série demonstraram ser capazes de: <ul style="list-style-type: none"> • localizar uma informação explícita parafraseada e ao longo de textos mais complexos; • inferir uma informação implícita e o sentido de uma palavra ou expressão pelo contexto de textos de estrutura mais complexa e não diretamente relacionados ao universo infantil; e inferir ainda, o

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<p>sentido de um poema pelo tempo verbal e pontuação apresentadas no título;</p> <ul style="list-style-type: none"> interpretar texto com auxílio de material gráfico mais abstrato (quadrinhos); identificar: (1) marcas lingüísticas que denotam irritação de um personagem do texto; (2) o tema de um texto em quadrinhos, de um poema e de uma notícia; (3) palavras que com base no senso comum se condicionam a um tempo moderno; (4) uma notícia em um texto informativo; (5) uma forma de tratamento usada em carta comercial; (6) o referente que substitui uma palavra; (7) um modo de escrita diferente de uma palavra no texto para chamar a atenção do leitor; (8) o conflito gerador de uma narrativa; (9) o sentido de uma palavra intercalada por parênteses em uma frase; antecipar e inferir a respeito do conteúdo do texto com base no tema dado; reconhecer idéia do que ainda acontecerá com base em um advérbio (amanhã) e em um tempo verbal (vai).
$\theta > 325$ 0,4%	O teste não teve nenhum item que representasse esse nível de proficiência.

Nota: Não foi feita uma descrição em intervalos menores porque mudaram os intervalos arbitrados para os níveis de proficiência dos itens.

Da avaliação de 1997 foram produzidas escalas únicas para a educação básica, por disciplina, sendo desconsideradas as descrições por série. Assim, na Tabela 22 são apresentadas as descrições para Língua Portuguesa (SAEB 97 – Primeiros Resultados, p. 8):

Tabela 22 – Níveis de proficiência de Língua Portuguesa publicados e percentuais de alunos nos níveis - SAEB - 1997

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
100 1,4%	Os alunos localizam uma informação em um texto, recuperando-a, posteriormente, para continuar a leitura; identificam a pontuação expressiva e são capazes de analisar o efeito de sentido decorrente do seu uso; são capazes de consultar um pequeno texto informativo para resolver um problema localizado de leitura.
$100 \leq \theta < 175$ 43,0%	Os alunos são capazes de fazer uma leitura mais elaborada do texto, interpretando-o; identificam informações e reconhecem o tema central em textos curtos e simples como bilhetes, receitas, instruções, poemas, historinhas; relacionam informações contidas em outros textos; percebem o sentido da pontuação e das palavras que expressam sentimentos, por exemplo; identificam personagens a partir de sua fala (gírias e expressões típicas).
$175 \leq \theta < 250$ 45,9%	Os alunos percebem a organização interna de diferentes tipos de textos; reconhecem que o autor usa a linguagem para expressar determinados sentidos; reconhece o papel do contexto na identificação de sentidos explícitos ou implícitos presentes no texto.
$250 \leq \theta < 325$ 9,4%	Os alunos são capazes de, em textos de certa complexidade, articular informações implícitas e pressupostas; percebem que a maneira como o texto está organizado supõe o seu leitor; identificam diferentes pontos de vista no tratamento do assunto, compreendendo os argumentos utilizados.
$325 \leq \theta < 400$ 0,4%	Os alunos revelam um repertório significativo de leituras, uma vez que têm capacidade de compreender a paródia como uma referência a outros textos; reconhecem a estrutura do texto poético, bem como os recursos expressivos utilizados para criar efeitos de sentido nesse tipo de texto; reconhecem o papel das preposições, conjunções e advérbios na organização e desenvolvimento do texto.

Fonte: MEC/INEP

Comparando-se as duas descrições, pode-se observar que na escala publicada são apresentadas habilidades descritas no nível de proficiência 100. Na análise feita por essa

pesquisadora, e na (re)descrição, houve o indicativo de que os itens que representariam este nível de proficiência não apresentam qualidade suficiente para serem considerados.

5.2.2. AVALIAÇÃO DE MATEMÁTICA

O teste de Matemática do SAEB 1997 foi elaborado com base nas Matrizes Curriculares de Referência para o SAEB (1997). A Matriz abrangeu cinco áreas de conteúdos, com números diversos de descritores, conforme pode ser observado na Tabela 23, a seguir.

Tabela 23 – Estrutura da Matriz de Matemática - 1997

Áreas	Subáreas	Descritores
Geometria	Espaço – localização, movimentação e representação (pontos de referência – próprio corpo e objetos extra-corpóreos)	12
	Formas – bidimensionais e tridimensionais (elementos e propriedades)	19
	Curvas – região interior e exterior	6
Medidas	Significado e unidades de medida – comprimento, superfície, capacidade e massa	8
	Unidades de medida de tempo	5
	Unidades de medida de temperatura	2
	Unidades de medida do sistema monetário	4
Números	Números naturais	4
	Sistema de numeração	2
Operações	Números racionais positivos: representação decimal	4
	Significado das operações: adição, subtração, multiplicação e divisão	6
	Propriedades das operações: adição, subtração, multiplicação e divisão	11
	Cálculo mental	3
Estatística	Lista, tabela e gráfico	3

A Matriz de Matemática, como a de Língua Portuguesa, parece não ter sido submetida a professores de sala de aula. Observa-se que alguns conteúdos importantes não foram contemplados na matriz, como, cálculo de área e perímetro de figuras planas; representações fracionárias de números racionais; reconhecimento dos diversos significados associados à fração; porcentagem.

O teste foi constituído de 143 questões das quais 85 (59%) foram consideradas para proceder à descrição do que os alunos brasileiros sabem ou deveriam saber fazer. Em síntese, analisando-se a qualidade do instrumento, pôde-se verificar que o teste apresentou muito

poucas questões contextualizadas, além de apresentar várias questões com conteúdos repetidos e com o mesmo nível de dificuldade. O teste foi constituído basicamente de conteúdos representativos do que os alunos aprendem na sala de aula, ou seja, faltaram questões que abordassem aplicações práticas e significativas do cotidiano, onde o aluno teria oportunidade de mostrar o crescimento ocorrido após um período de escolarização.

Como nos outros testes analisados, houve várias questões inadequadas, tanto em relação ao conteúdo quanto ao formato, para o nível de escolaridade avaliado. E, como já comentado para o teste de língua portuguesa, houve mudança dos intervalos da escala e dos critérios para determinação dos níveis representativos dos itens, com isso, a posição de vários itens comuns passou para a seguinte, mostrando, nesse teste, que seria necessária uma proficiência maior para respondê-los, quando na verdade, a invariância dos parâmetros foi mantida.

Quanto à distribuição dos conteúdos, considerando-se a matriz de 1997, 32% dos itens do teste dizem respeito ao tema Geometria; 14% ao tema Medidas; 15% ao tema Números; 33% diz respeito ao tema Operações; e 6% ao tema Estatística. Dos 22 itens comuns do teste, 17 itens referiram-se ao tema operações, três itens de Número, dois itens do tema Medidas e nenhum item de Geometria.

De modo geral, retirando os itens que não foram adequados ou apropriados para avaliarem os alunos (apresentaram problemas pedagógicos ou psicométricos), pode-se dizer que o teste abrangeu conteúdos relacionados a todos os temas descritos na matriz, embora tenham sido contempladas poucas habilidades que exigissem raciocínio dos alunos ou que extrapolassem para situações de aplicação na vida real.

Na Tabela 24 são apresentadas as estatísticas geradas dos itens selecionados para a descrição da escala de proficiência dos alunos em Matemática, 4ª série, 1997. Essa seleção foi feita considerando-se os critérios já descritos na metodologia e citados nas tabelas anteriores.

Tabela 24 – Estatísticas – Teste Matemática – 4ª Série – SAEB – Edição 1997

	Parâmetro a	Parâmetro b	Parâmetro c	Discriminação Índice D	Correlação bisserial	Proporção de acerto
N	85	85	85	85	85	85
Média	1,63	-0,66	0,18	0,44	0,49	0,46
Desvio Padrão	0,54	0,94	0,19	0,13	0,09	0,17
Mínimo	0,60	-3,50	0,01	0,29	0,30	0,06
Máximo	3,51	1,21	0,50	0,75	0,66	0,87

Na Tabela 25 é apresentada a (re)descrição dos níveis de proficiência que representam o que os alunos brasileiros demonstraram saber fazer, com base no teste aplicado, e o percentual de alunos da 4ª série, que atingiu cada um desses níveis.

Tabela 25 – Descrição dos percentuais de alunos e dos níveis com base no teste de Matemática – 4ª Série – 1997

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$\theta < 175$ 40,2%	Neste nível não houve nenhum item que o representasse.
$175 \leq \theta < 250$ 49,8%	Alunos da 4ª série que atingiram esse nível de proficiência demonstraram ser capazes de: <ul style="list-style-type: none"> • resolver problemas simples envolvendo: (1) a facilitação de troco em situação de compra; (2) a adição de cédulas e moedas do Sistema Monetário Nacional; (3) uma subtração com o significado de alteração do estado inicial e com uma situação de invariância da diferença; e (5) multiplicação de números naturais e decimais (valor monetário); • identificar: (1) um círculo entre diversas formas arredondadas; (2) linhas poligonais abertas e fechadas; (3) a representação de figuras tridimensionais simples pelas formas arredondadas (cone e cilindro); (4) um objeto que se encontra entre duas posições; e (5) uma fração em desenho representando parte-todo; • calcular o tempo que falta para a realização de um evento com hora determinada; e o valor de uma incógnita em uma expressão, dadas algumas opções; • ler e interpretar horários em relógios de ponteiros e digitais; • decompor número e representá-lo por extenso; • inferir o que é menor pelo método de comparação e que para resolver um problema é necessário proceder a uma divisão; • ordenar numerais decimais em ordem crescente;
$250 \leq \theta < 325$ 9,6%	Neste nível de proficiência, os alunos de 4ª série demonstraram ser capazes de: <ul style="list-style-type: none"> • identificar: (1) o número de dias que tem uma semana e calcular o número de dias de 13 semanas; (2) quantas horas tem um dia, quantos dias tem a semana e calculam quantas horas tem uma semana; (3) o valor posicional de um algarismo na construção da escrita de um número; (4) a representação de um número decimal na reta numérica; (5) entre diversas medidas de mesma grandeza a mais apropriada para medir a altura de um prédio (m); (6) um objeto entre vários pela posição indicada no enunciado (esquerda e abaixo de um outro objeto); (7) a posição de um número dados dois referenciais; (8) a posição de uma pessoa em uma ilustração gráfica dada a referência (segunda criança da esquerda para a direita); • resolver uma situação-problema: (1) que utiliza cálculos simples com cédulas e moedas; (2) envolvendo relação de proporção, operações de multiplicação incluindo propriedade comutativa e do elemento neutro e cálculo de adição e de subtração; (3) utilizando procedimentos de cálculo mental aproximado e de uma expressão numérica; e, (4) aplicando a propriedade da subtração da invariância da diferença; • inferir que dois pontos em um círculo estão à mesma distância de um ponto central e que devem realizar uma operação de adição para solucionar um problema; • estabelecer: (1) troca de moedas em função do valor; (2) regra para seqüência de números correlacionados em duas colunas; (3) relação de correspondência entre os dados de uma tabela e um gráfico; • interpretar dados representados em gráficos de setores e ler um gráfico de colunas, interpretá-lo e calcular o resultado da soma de todas as colunas; • decompor número natural e escrito por extenso em sua forma polinomial (dezena de milhar) e formar o maior número considerando-se a ordem de grandeza e o valor posicional, dados quatro algarismos; • calcular o resultado: (1) de uma subtração com reserva e com um zero intercalado no numeral referente ao subtraendo; (2) de uma divisão de números naturais e o resto; (3) de multiplicações por 10 e por 100; (4) de uma subtração de frações com o mesmo denominador; (5) de uma expressão numérica, em que se tem que substituir os valores representados por figuras, envolvendo adição, subtração e multiplicação; (6) do número de minutos de um dia;

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<ul style="list-style-type: none"> representar um número natural na reta numérica e relações espaciais utilizando elementos posicionais (para frente / para baixo); comparar leituras realizadas em termômetros situados em ambientes diferentes, reconhecendo as diferenças indicadas;
$\theta > 325$ 0,4%	<p>Alunos de 4ª série que alcançaram esse nível de proficiência demonstraram ser capazes de</p> <ul style="list-style-type: none"> comparar: (1) grandezas de mesma natureza utilizando cálculos com unidades convencionais e não convencionais; (2) unidades de tempo (dias, semanas e meses); (3) e ordenar números racionais na forma decimal; (4) e converter horas e minutos em minutos; identificar: (1) a forma de um cilindro; (2) características próprias do retângulo e do quadrado; (3) semelhanças entre duas figuras tridimensionais pelo número de faces; (4) o vértice de uma figura plana e o número de vértices de uma figura tridimensional; (5) a posição representada de uma imagem espelhada; (6) dois segmentos de retas como paralelos; (7) o significado de frações como quocientes; (8) a escrita de uma fração na forma de um número decimal; interpretar: (1) um croqui utilizando elementos posicionais (frente e esquerda); (2) a representação da movimentação do corpo girando para frente de quem está olhando e na direção esquerda da figura; (3) dados representados em um gráfico de colunas; transformar medida de quilômetros para metros; resolver problema envolvendo situação comparativa com existência de um elemento neutro e aplicando a propriedade distributiva da multiplicação para resolver problema; efetuar uma multiplicação de números da ordem das centenas, sendo que um deles possui um zero intermediário;

A escala única produzida e publicada para a educação básica, para a disciplina Matemática, encontra-se transcrita na Tabela 26 (SAEB 97 – Primeiros Resultados, p. 7).

Tabela 26 – Níveis de proficiência de Matemática publicados e percentuais de alunos nos níveis – SAEB – 1997

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$\theta < 100$ 0,2%	Não houve itens que descrevessem este nível.
$X \leq \theta < 175$ 39,9%	Os alunos demonstram conhecimentos elementares de geometria; são capazes de localizar a posição dos objetos tendo como referência o próprio corpo e de reconhecer figuras geométricas simples, desde que representadas na forma usual; reconhecem o valor de cédulas e moedas; lêem horas em relógios digitais e analógicos e sabem que a hora tem 60 minutos; lêem e escrevem números de poucos dígitos e resolvem problemas simples de adição e subtração com números naturais.
$175 \leq \theta < 250$ 49,8%	Os alunos possuem conhecimento de geometria que possibilita a descrição da movimentação de objetos tendo como referência o próprio corpo; reconhecem polígonos e quadriláteros; estabelecem relações entre os valores de cédulas e moedas e resolvem situações de pagamento e troco, embora ainda não saibam operar com decimais; relacionam diferentes unidades de medida de tempo e sabem selecionar a unidade de medida mais adequada para fazer determinada medição; são capazes de multiplicar e dividir e de identificar unidades, dezenas, centenas etc.; resolvem problemas envolvendo mais de uma operação; adicionam e subtraem frações de mesmo denominador e conhecem números naturais na forma fracionária; interpretam gráfico de barras e de setor e identificam o gráfico mais adequado para representar uma dada situação; expressam generalizações observadas em seqüências numéricas por meio de representações algébricas.
$250 \leq \theta < 325$ 9,6%	Os alunos apresentam noções de paralelismo, perpendicularismo e ângulo e descrevem a posição e o movimento de um objeto utilizando diferentes pontos de referência; classificam sólidos geométricos em corpos redondos e poliedros; interpretam resultados de medidas de comprimento, massa, tempo e capacidade; estabelecem relações entre unidades de medida relacionadas a uma mesma grandeza e resolvem problemas envolvendo essas medidas; identificam, comparam e ordenam números racionais (nas formas fracionárias e decimais) e números inteiros; interpretam informações apresentadas em gráficos e tabelas, incluindo aquelas apresentadas na forma de porcentagem, sendo capazes de fazer prognósticos a partir dessas informações; interpretam escritas algébricas e resolvem equações de primeiro grau;

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$325 \leq \theta < 400$ 0,4%	Os alunos resolvem problemas de Geometria Euclidiana, empregando relações algébricas e trigonométricas, utilizando as propriedades e características das principais figuras espaciais; resolvem problemas envolvendo ponto, reta, circunferência e suas relações; solucionam problemas que envolvem cálculo de comprimento, área, volume, temperatura e tempo, utilizando fórmulas e as unidades de medidas usuais e suas inter-relações; resolvem problemas simples em porcentagens, proporcionalidade direta e inversa e probabilidades; interpretam dados em tabelas representando-os em gráficos de barras; estabelecem relações e fazem conversões entre frações ordinárias e números decimais; resolvem problemas que envolvem equações e inequações simples de 1º e 2º graus e sistemas de 1º grau; conhecem os princípios básicos de polinômios e efetuam operações elementares entre eles; resolvem problemas simples que empreguem conceitos de matrizes, determinantes, identificação de números complexos e sua representação e que envolvam a interpretação geométrica do conceito de módulo; conhecem as propriedades básicas de exponenciais e logaritmos.

Fonte: MEC/INEP

A descrição da escala no formato único, considerando-se o ciclo completo da educação básica, tira a oportunidade de o professor de todos os níveis se identificar no processo avaliativo. Essa fenda que se abre entre o que de fato ocorre nas salas de aula do Brasil e o que é relatado enquanto resultado dos alunos, torna os resultados da avaliação difíceis de serem compreendidos. Pois, mesmo os elaboradores de políticas e avaliadores de programas, que na área da educação geralmente tiveram sua origem como educadores, se vêem impotentes a indicarem, por exemplo, cursos de formação em áreas de maior defasagem de conhecimento.

5.3 SAEB – EDIÇÃO 1999

O documento Matrizes Curriculares de Referência para o SAEB 1997, passou por um processo de adequação e pertinência das matrizes e desse processo resultou uma segunda edição revisada e ampliada. A Matriz 1999 manteve a mesma concepção, mas mudou o objetivo central, que nesta edição, ficou explicitado que seria *identificar o conjunto de descritores do desempenho desejável dos alunos em cada disciplina ao longo da educação básica, para orientar a elaboração de itens que compõem as provas do SAEB* (p. 9).

Em relação aos conteúdos abrangidos, a proposta continuou a mesma, estando eles associados às competências cognitivas e às habilidades instrumentais. As competências, da mesma forma, continuaram categorizadas em três níveis de ações e de operações mentais que se estabelecem conforme as relações entre o sujeito e o objeto do conhecimento: básico ou presentativo, operacional ou procedural e global ou operatório.

5.3.1. AVALIAÇÃO DE LÍNGUA PORTUGUESA

A matriz curricular de Língua Portuguesa continuou tendo por base o novo olhar sobre o ensino da língua, que tem por orientação *criar situações nas quais o aluno amplie o domínio ativo do discurso nas diversas situações comunicativas, sobretudo nas instâncias públicas de uso da linguagem, de modo a possibilitar sua inserção efetiva no mundo da escrita, ampliando suas possibilidades de participação social no exercício da cidadania* (p. 13).

Uma mudança estabelecida para essa Matriz foi a junção das áreas Práticas de leitura de textos e Análise e reflexão lingüística, que segundo especialistas, não faz sentido a separação, considerando-se que os conhecimentos lingüísticos a serem contemplados nos testes são operacionais. Assim, mantiveram-se as Práticas de leitura de textos e a Produção de textos. Os descritores referentes à leitura foram reduzidos e modificados, numa tentativa, ainda que distante, de aproximar da prática real da sala de aula.

As mudanças incorporadas à área de Práticas de leitura relativas aos tópicos, muito se aproximaram da matriz atual (2001). Essa área ficou subdividida em cinco subáreas, totalizando 28 descritores, conforme apresentado na Tabela 27.

Tabela 27 – Estrutura da Matriz de Língua Portuguesa – 1999

Área	Subáreas	Descritores
Práticas de leitura de textos	Procedimentos de Leitura	8
	Implicações do suporte, do gênero e/ou enunciador na compreensão do texto	2
	Coerência e coesão no processamento do texto	5
	Relações entre recursos expressivos e efeitos de sentido	9
	Variação lingüística	4
Práticas de produção de textos	Tipos de textos	0
	Habilidades	0

Embora as subdivisões tenham ficado mais claras e compreensíveis, os descritores associados a cada subdivisão permaneceram ainda obscuros a quem não tem formação na área, a qual não é exigida para as séries iniciais. A título de curiosidade, foi feita uma pesquisa entre seis professores que estão atuando em 4ª séries de escolas públicas e particulares, três com formação em nível superior (magistério), solicitando que informassem o que compreendiam da matriz de referência de Língua Portuguesa (1999) para 4ª série. O resultado foi que, com exceção dos descritores de Procedimentos de leitura, que foram compreendidos por todas as professoras, todas as outras subáreas não foram bem

compreendidas. E a informação mais importante foi que todas concordaram que uma avaliação com essas características não reflete o que é praticado na sala de aula.

O teste de Língua Portuguesa, 4ª série, foi constituído de 169 questões das quais 105 (68%) foram aproveitadas para proceder à descrição do que os alunos brasileiros sabem ou deveriam saber fazer.

Em síntese, analisando-se a qualidade do instrumento, pôde-se verificar que houve vários textos com temas inadequados, entre esses, poemas abstratos, crônicas, quadrinhos do Hagar, charges, que exigiam um conhecimento de mundo e uma capacidade de abstrair muito maior do que a faixa etária desses alunos permite. O principal problema do teste referiu-se ao formato dos itens. Houve 17 itens que solicitavam relacionar colunas, quatro itens de completar frases, seis itens que não se referiam a nenhum texto, mas a frases sem contexto. Esses formatos comprometem a validade dos itens. Não se pode garantir até que ponto o aluno não acertou o item porque não sabia, ou porque não conseguiu relacionar tantas informações. Itens desse tipo devem ser evitados. Outro problema sério que apareceu com frequência neste teste foi a plausibilidade das alternativas dos itens. Alguns itens apresentaram opções que nem fazem parte do texto, o que aumenta a chance de acerto do aluno, mesmo sem saber o conteúdo avaliado.

O teste manteve a tendência de abranger conteúdos representativos apenas do que os alunos aprendem na sala de aula, sem contemplar aplicações práticas do seu cotidiano. Houve também muitos enunciados de difícil compreensão com informações acessórias que não tinham nada a ver com o que de fato estava sendo solicitado;

Mais uma vez houve mudança dos intervalos considerados para a descrição da escala e dos critérios para determinação dos níveis representativos dos itens. Com isso, a posição de vários itens comuns passou para a seguinte ou para a antecedente, mostrando que aparentemente, seria necessária uma proficiência maior ou menor para respondê-los, quando na verdade, relembrando, a invariância dos parâmetros foi e é sempre mantida, seguindo-se os pressupostos da Teoria de Resposta ao Item. Além disso, dos itens comuns, vários tinham sido desconsiderados na descrição da escala passada por terem apresentado problemas pedagógicos ou estatísticas ruins, e agora vieram compor o teste e a escala de 1999.

Na Tabela 28 são apresentadas as estatísticas geradas dos itens selecionados para a descrição da escala de proficiência dos alunos em Língua Portuguesa, 4ª série, 1999. Essa seleção foi feita considerando-se os critérios já descritos anteriormente.

Tabela 28 – Estatísticas – Teste Língua Portuguesa – 4ª Série – SAEB – Edição 1999

	Parâmetro a	Parâmetro b	Parâmetro c	Discriminação Índice D	Correlação bisserial	Proporção de acerto
N	105	105	105	105	105	105
Média	1.08	-1.07	0.19	0.48	0.52	0.46
Desvio Padrão	0.36	0.69	0.09	0.09	0.08	0.14
Mínimo	0.60	-2.88	0.01	0.26	0.29	0.18
Máximo	2.59	0.44	0.43	0.68	0.71	0.80

Na Tabela 29, a (re)descrição dos níveis de proficiência que representam o que os alunos brasileiros demonstraram saber fazer, com base no teste aplicado, e o percentual de alunos da 4ª série, que atingiu cada um desses níveis.

Tabela 29 – Descrição dos percentuais de alunos e dos níveis com base no teste de Língua Portuguesa – 4ª Série – 1999

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$\theta < 150$ 35,6%	Neste nível não houve nenhum item que pudesse representá-lo.
$150 \leq \theta < 175$ 22,0%	Os alunos que atingiram este nível de proficiência demonstraram ser capazes de: <ul style="list-style-type: none"> • localizar uma informação explícita em textos em quadrinhos; • reconhecer: (1) o efeito de sentido de uma onomatopéia comum (imitação de sons representados por palavras); (2) o locutor e interlocutor com base nas marcas linguísticas características de cada um deles; (3) o narrador da história; • inferir: (1) o sentido de uma expressão comum; (2) uma informação implícita em uma frase e em um texto com base no assunto principal; • relacionar informações do texto com conhecimentos do senso comum; • interpretar texto observando diretamente figuras que compõem uma ilustração gráfica.
$175 \leq \theta < 200$ 17,6%	Neste nível de proficiência, os alunos demonstraram ser capazes de: <ul style="list-style-type: none"> • localizar informação explícita no início de uma lenda; • inferir uma informação implícita com auxílio de ilustração gráfica e que se encontram na superfície do texto, e o sentido de uma expressão comum em um texto; • reconhecer: (1) o efeito de sentido de uma onomatopéia menos comum; (2) uma frase, entre outras que denota linguagem formal; (3) o efeito de sentido produzido pela repetição de palavras em uma frase; (4) personagens comuns de livros infantis (bruxa e fada); • estabelecer relação causa/conseqüência entre partes e elementos do texto; • interpretar texto com auxílio de material gráfico (quadrinhos).
$200 \leq \theta < 225$ 12,2%	Alunos que atingiram este nível de proficiência demonstraram ser capazes de: <ul style="list-style-type: none"> • localizar uma informação explícita no meio de um texto e algumas características do personagem principal de um texto; • inferir: (1) o sentido de uma expressão comum no texto ou aquela utilizada repetidas vezes ao longo do texto; (2) uma informação implícita em um texto curto, ou com base em um trocadilho com palavras, ou com auxílio de material gráfico, ou com base no contexto apresentado no texto; • utilizar o apoio de uma foto para compreenderem o motivo de uma propaganda ou a origem de um nome; • estabelecer: (1) relações entre partes de um texto a partir da substituição de um termo ou de substituições que contribuem para a continuidade do texto; (2) relação causa/conseqüência entre

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<p>partes e elementos do texto; (3) relação lógico-discursiva marcada por advérbio de tempo;</p> <ul style="list-style-type: none"> identificar: (1) marcas lingüísticas denotadas por sentimento de impaciência; (2) o efeito de sentido produzido pelo uso de uma rima e de uma onomatopéia (chamado efeito prosódico); (3) características do personagem principal de uma história narrada; (4) marcas lingüísticas características de pessoas que vivem na zona rural; (5) características de escrita de uma bula de remédios em comparação a outros textos; e (6) o sentido da repetição de um verbo ao longo de um poema;
<p>225 ≤ θ < 250 7,4%</p>	<p>Alunos que alcançaram esse nível de proficiência demonstraram ser capazes de:</p> <ul style="list-style-type: none"> localizar uma informação explícita entre várias similares, no meio do texto; inferir: 1) uma informação implícita com apoio de uma ilustração gráfica ou com base na compreensão da solução de um problema explicitado no texto ou no título de um texto, ou implícita em um poema; (2) o sentido de uma expressão ou de palavras que compõem o título do texto, ou expresso pelo contexto do texto; identificar: (1) o tema central de um texto informativo; (2) o efeito de sentido de súplica decorrente do uso de um ponto de exclamação; (3) marcas lingüísticas que evidenciam o personagem não comum de um texto, outras de um personagem de história em quadrinhos, e outras que denotam irritação de um personagem do texto; (4) o gênero informativo de um texto; (5) o efeito de sentido do uso de duas interrogações com sentidos diferentes no texto e o efeito de sentido de palavras produzido pelo uso de aliteração (repetição de fonemas) no meio das frases; (6) a condição de continuidade de um texto pela repetição de palavras; (7) o personagem responsável por um fato engraçado descrito; (8) uma expressão que representa um ditado popular; (9) o significado de uma palavra pouco comum; (10) a quem é dirigida uma propaganda; e (11) palavras que com base no senso comum se relacionam a um tempo moderno; estabelecer relação entre o problema apresentado no texto e a solução oferecida e uma relação lógico-semântica (comparação) entre características comuns de dois personagens; associar características dadas pelo autor a um dos personagens; interpretar texto com auxílio de material gráfico, inferindo uma informação implícita dele;
<p>250 ≤ θ < 275 3,5%</p>	<p>Os alunos que atingiram este nível de proficiência demonstraram ser capazes de:</p> <ul style="list-style-type: none"> localizar resposta a um dos problemas impostos em texto longo e com muitas informações. inferir: (1) uma informação implícita a partir do estabelecimento de uma relação causal no texto e com base na opinião do personagem principal de um texto e no efeito de sentido produzido pelo uso de um recurso sintático (repetição de palavras); (2) o sentido de expressões em textos longos e colocadas com base no contexto apresentado e com muitas informações, ou pelo sinônimo correspondente e por um sentido não comum ao universo infantil e, ainda, (3) o sentido do tema de um poema pelo tempo verbal e pontuação apresentadas no título. identificar: (1) entre diversas falas a que reproduz um diálogo; (2) o efeito de sentido produzido pelo uso de recursos sintáticos (repetição de um verbo) e do uso de uma interrogação em um texto; (3) marcas lingüísticas próprias de jovens evidenciadas por interlocutores em um texto; (4) o referente que substitui uma palavra; (5) uma característica (rima entre duas palavras) de um poema complexo; (6) o significado de uma palavra ainda pouco usual no vocabulário infantil; (7) a função do negrito em um texto; (8) a fala de um dos personagens secundários do texto. relacionar: (1) informações textuais com conhecimento de mundo para identificarem a parte faltante de um texto; (2) um fato de uma opinião relativa a esse fato em um texto complexo com muitas informações; (3) diferentes formas de tratar uma informação na comparação de textos que tratam do mesmo tema; estabelecer relação entre partes de um texto por meio de mecanismos de concordância nominal; antecipar e inferir a respeito do conteúdo do texto com base no tema dado.
<p>275 ≤ θ < 300 1,2%</p>	<p>Neste nível de proficiência, os alunos demonstraram ser capazes de:</p> <ul style="list-style-type: none"> inferir o sentido de uma expressão não comum ao jargão infantil e o sentido do apelido de um personagem pelas suas características e o sentido de uma expressão com base no uso de um recurso sintático (repetição) utilizado no texto; identificar: (1) opiniões expressas por personagens de uma história; (2) o efeito de sentido de uma onomatopéia no texto, que poderia exprimir outros sentidos; (3) o efeito de sentido de chamamento expresso por um ponto de exclamação em um quadrinho; (4) personagem e elemento da história por meio do efeito de sentido do uso de linguagem figurada; (5) um tipo de texto e suas características; (6) a função da notação de números entre colchetes, indicando nota de rodapé; (7) uma das características de um poema (estrutura);

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<ul style="list-style-type: none"> estabelecer relações entre partes de um texto por meio da identificação do referente de uma expressão.
> 300 0,5%	Neste nível, os alunos demonstraram ser capazes de identificar características típicas, mas não comuns, da fala de uma região do país.

Da avaliação de 1997 foram produzidas escalas únicas para a educação básica, por disciplina, sendo desconsideradas as descrições por série. Na Tabela 30, as descrições para Língua Portuguesa, educação básica (Resultados dos Níveis de Desempenho – Língua Portuguesa e Matemática, retirado do site http://www.inep.gov.br/basica/saeb/saeb_99.htm, pp. 61 – 69).

Tabela 30 – Níveis de proficiência de Língua Portuguesa publicados e percentuais de alunos de 4ª série nos níveis – SAEB – 1999

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
0 < 150 35,6%	Neste nível não houve nenhum item que pudesse representá-lo.
150 ≤ θ < 200 39,6%	Neste nível os alunos operam preferencialmente com estratégias locais de leitura, identificam informações cruciais/centrais, em posição destacada, e ainda a finalidade ou tema de um texto. Usam, também, conhecimento de mundo na percepção do sentido de um texto.
200 ≤ θ < 250 19,6%	Os alunos resolvem problemas de leitura a partir da compreensão global do texto, incluindo inferências, localizam informações secundárias, reconstroem uma narrativa, encadeando vários fatos na ordem de aparição, e reconhecem efeitos de sentido de recursos variados (repetição, substituição, onomatopéia).
250 ≤ θ < 300 4,7%	Os alunos estabelecem relações coesivas entre partes do texto, inclusive pelo reconhecimento de tópico e comentário, distinguem “fato” de “opinião”; problema de solução; tese de argumento; causa de efeito, fazem transformações estruturais e estabelecem relações de correspondência, e ainda compreendem explicações mais abstratas, metalingüísticas.
300 ≤ θ < 350 0,5%	Neste nível, os alunos, na comparação de textos afins, identificam e avaliam as estratégias argumentativas e a finalidade de cada um, estabelecem relações sintático-semânticas na progressão temática, mostram conhecimento da estrutura e funcionamento dos gêneros textuais e, também, apresentam boa noção da relação linguagem e sociedade.
350 ≤ θ < 400 0,0%	Neste nível, os alunos trabalham com linguagem figurada/conotativa em nível global, articulado, identificam diferentes níveis de tratamento temático, reconhecendo tópicos e subtópicos, analisam o efeito da seleção lexical em uma argumentação, e aplicam com propriedade conhecimentos metalingüísticos e literários.

Fonte: MEC/INEP

5.3.2. AVALIAÇÃO DE MATEMÁTICA

O teste de Matemática do SAEB 1999 foi elaborado com base nas Matrizes Curriculares de Referência, em documento reeditado em 1999, como as matrizes de Língua Portuguesa. A Matriz de Matemática continuou abrangendo as mesmas cinco áreas de conteúdos, mas modificando o número de descritores por tema e algumas subáreas. Na Tabela 31 é apresentada essa estrutura.

Tabela 31 – Estrutura da Matriz de Matemática – 1999

Áreas	Subáreas	Descritores
Geometria	Espaço – localização, movimentação e representação (pontos de referência – próprio corpo e objetos extra-corpóreos)	7
	Formas – bidimensionais e tridimensionais (elementos e propriedades)	23
Medidas	Significado e unidades de medida – comprimento, superfície, capacidade e massa	6
	Unidades de medida de tempo	5
	Unidades de medida de temperatura	2
	Significado e unidades de medida do sistema monetário	4
Números	Números naturais e Sistema de numeração decimal	4
	Números racionais positivos: representação fracionária e decimal	8
Operações	Significado das operações: adição, subtração, multiplicação e divisão	4
	Propriedades das operações	5
	Cálculo	2
Estatística	Lista, tabela simples e de dupla entrada e gráfico	4

Alguns problemas detectados na primeira versão da matriz e já discutidos anteriormente permaneceram sem ser resolvidos.

O teste foi constituído de 169 questões das quais 117 (69%) foram consideradas na descrição do que os alunos brasileiros sabem ou deveriam saber fazer. Em síntese, analisando-se a qualidade do instrumento, pôde-se verificar que em relação ao teste anterior (1997), este apresentou mais questões contextualizadas. Verifica-se que ainda houve privilégio de conteúdos representativos do que os alunos aprendem na sala de aula, ou seja, faltaram questões que abordassem aplicações práticas e significativas do cotidiano. Houve muitas questões inadequadas ao nível de desenvolvimento cognitivo, que normalmente os alunos dessa fase escolar se encontram;

Houve, novamente, mudança da métrica da escala e dos critérios para determinação dos níveis representativos dos itens, com isso, a posição de vários itens comuns passou para a seguinte, mostrando, nesse teste, que seria necessária uma proficiência maior para respondê-los, quando na verdade, a invariância dos parâmetros foi mantida;

Quanto à distribuição de conteúdos, verifica-se que a área de geometria foi privilegiada. Isso chama a atenção pelo fato de ser comum os professores das séries iniciais deixarem para contemplar os conteúdos relacionados a esse tema no final do ano, e muitas vezes não haver tempo para contemplá-los. A distribuição dos conteúdos, considerando-se a matriz de 1999, ficou da seguinte forma: 32% dos itens do teste dizem respeito ao tema

Geometria; 22% ao tema Medidas; 15% ao tema Números; 22% diz respeito ao tema Operações; e 9% ao tema Estatística.

Na Tabela 32, a seguir, são apresentadas as estatísticas geradas dos itens selecionados para a descrição da escala de proficiência dos alunos em Matemática, 4ª série, 1999. Essa seleção foi feita considerando-se os critérios descritos na metodologia e o atual foco da avaliação do SAEB.

Tabela 32 – Estatísticas – Teste Matemática – 4ª Série – SAEB – Edição 1999

	Parâmetro a	Parâmetro b	Parâmetro c	Discriminação Índice D	Correlação bisserial	Proporção de acerto
N	117	117	117	117	117	117
Média	1,18	-0,69	0,19	0,44	0,49	0,43
Desvio Padrão	0,43	0,69	0,10	0,12	0,09	0,14
Mínimo	0,55	-2,16	0,01	0,10	0,28	0,05
Máximo	2,64	0,81	0,48	0,69	0,68	0,70

A Tabela 33 apresenta a (re)descrição dos níveis de proficiência que representam o que os alunos brasileiros demonstraram saber fazer, com base no teste aplicado, e o percentual de alunos da 4ª série, que atingiu cada um desses níveis.

Tabela 33 – Descrição dos percentuais de alunos e dos níveis com base no teste de Matemática – 4ª Série – 1999

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$\theta < 150$ 23,9%	Neste nível não houve nenhum item que o representasse.
$150 \leq \theta < 175$ 23,9%	Alunos da 4ª série que atingiram esse nível de proficiência demonstraram ser capazes de: <ul style="list-style-type: none"> Identificar a localização de um objeto que se encontra atrás de outro em uma representação gráfica.
$175 \leq \theta < 200$ 22,4%	Neste nível de proficiência, os alunos de 4ª série demonstraram ser capazes de: <ul style="list-style-type: none"> Identificar: (1) elemento posicional (ao lado e mais distante); (2) a figura que possui maior área em relação a outras três figuras planas, sem apoio da malha quadriculada; (3) figuras planas representadas (losango, retângulo e quadrado); (4) a localização de objeto que se encontra entre outros dois; (5) a configuração de objetos conforme posição relacionada no enunciado; Resolver problemas por meio da decomposição de números naturais e por meio da escolha de uma expressão numérica que indica a sua solução; Ler e interpretar: (1) dados de uma tabela com três colunas e um gráfico de colunas; (2) o símbolo de litro (<i>l</i>) e grama (<i>g</i>); (3) calendário que informa dia do mês, ano e horas, marcada em relógio de ponteiros; (4) um gráfico de colunas, além de calcular o resultado da soma de todas essas colunas; (5) a representação da movimentação de um objeto, identificando a descrição da movimentação.
$200 \leq \theta < 225$ 15,3%	Alunos de 4ª série que alcançaram este nível de proficiência demonstraram ser capazes de <ul style="list-style-type: none"> Identificar: (1) em um mapa, entre quais cidades uma outra está localizada e a que está localizada à direita de um referencial; (2) figuras tridimensionais (cone e esfera) com superfícies arredondadas; (3) entre dois desenhos feitos sobre uma malha quadriculada de mesmas proporções, aquele que possui a menor área em relação ao outro; (3) a representação numérica de um número racional escrito por extenso; (4) o registro de temperatura em graus centígrados envolvendo decimais em um termômetro; (5) o símbolo que representa unidade de medida de massa (kg); (6) a forma escrita da fração $\frac{1}{2}$; (7) o hexágono como figura com seis vértices, entre outras figuras; (8) figuras originadas da decomposição de um retângulo, pelo nome;

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<ul style="list-style-type: none"> • Estimar medidas de grandezas proporcionais, utilizando unidades de medida não convencionais; transformar diferentes unidades de medida de tempo (horas/dias, semanas/dias, décadas/anos); comparar diversas temperaturas, ordenando do mais frio para o mais quente; estabelecer relação entre duração de um evento, dados horários de início e término e trocas entre cédulas e moedas em função de seus valores; • Compor números naturais representados em ábacos; • Resolver problemas envolvendo multiplicação e quantias em dinheiro e efetuar multiplicação de decimal por 100, envolvendo valor monetário (0,01 x 100) e facilitação de troco em situação de pagamento; • Ler e interpretar: tabelas com dados dispostos em duas colunas; gráficos de setor e de colunas, indicando a maior e a menor proporção, além da soma dos dados de duas entre cinco colunas representadas e associar dados de um gráfico a uma tabela.
<p>225 ≤ θ < 250 8,2%</p>	<p>Neste nível de proficiência, os alunos de 4ª série demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Identificar: (1) o mesmo número de lados em duas figuras diferentes; (2) figuras tridimensionais representadas graficamente; (3) a descrição da movimentação de um objeto, entre outras representações com elementos direcionais diferentes; • Comparar leituras realizadas em termômetros, reconhecendo as diferenças de temperatura que eles indicam; Converter unidades de medida de tempo (meses em semanas) e comparar grandezas de mesma natureza, utilizando unidades de medidas não convencionais; • Decompor número escrito por extenso na forma numérica polinomial e classificar em ordem crescente números inteiros e decimais com uma e três casas decimais; • Resolver problemas envolvendo: (1) soma de valores monetários com decimais e com somas simples de unidades, tendo números representados por figuras (2) subtração, abordando a invariância da diferença; e, outro, no qual, para chegar à solução, é necessário seguir a ordem inversa dos dados; (3) situações de compra, venda e devolução de troco, com quantias não exatas em dinheiro; (4) mais de uma operação, e envolvendo multiplicações por 10, com sentido comparativo proporcional e aplicando a propriedade do elemento neutro e a relação de proporção; (5) o reconhecimento de uma expressão numérica; e (6) procedimentos de cálculo mental aproximado, por meio de estratégias pessoais; • Ler tabela para comparar medidas de grandeza de mesma natureza (peso e altura) e estabelecer relação de correspondência entre os dados de uma tabela e um gráfico.
<p>250 ≤ θ < 275 4,2%</p>	<p>Alunos da 4ª série que atingiram esse nível de proficiência demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Identificar: (1) o número de faces de um cubo; (2) o valor posicional do algarismo nove como o de maior valor em um numeral formado por quatro algarismos e o de um algarismo na construção da escrita de um número pelo sistema de numeração; (3) o número de anos que tem um século e o símbolo da unidade de medida adequada para medir superfície da área (m²); (3) que diferentes problemas podem ser resolvidos utilizando uma mesma operação; (4) a fração correspondente a um desenho representando parte-todo; (5) diferentes formas de resolver um mesmo problema; (6) as propriedades do elemento neutro na adição em expressão representada por desenho e da invariância da diferença na subtração; • Ler horas e minutos em relógios digital e analógico e fazer comparação entre seus horários; estabelecer relações entre unidades de medida de comprimento (m/cm) e realizar trocas entre cédula e moedas; • Comparar e ordenar números racionais na forma decimal, envolvendo décimo, centésimo e milésimo; • Efetuar uma adição de decimais envolvendo numerais inteiros e milésimos; o cálculo de subtração de decimais e o cálculo para chegar ao resultado e ao resto de uma divisão de um algarismo da ordem das centenas por um da ordem das dezenas; • Resolver problema envolvendo: (1) soma de quantias em dinheiro, combinando dois estados para chegar a um terceiro; (2) a elaboração de uma expressão numérica; (3) situação de compra, devolução e facilitação de troco; (4) multiplicação comparativa; (5) a adição de número decomposto na forma polinomial; (6) a adição de números racionais, com inteiros e decimais; (7) a comparação e a conversão de horas e minutos em minutos; • Ler e interpretar dados de uma tabela com até cinco colunas e dados representados em um gráfico de setor, cujos valores encontram-se muito próximos; representar dados em um gráfico de setores; e associar dados representados em uma tabela a um gráfico correspondente.
<p>275 ≤ θ < 300</p>	<p>Alunos de 4ª série que alcançaram esse nível de proficiência demonstraram ser capazes de</p>

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
1,6%	<ul style="list-style-type: none"> • Identificar: (1) figuras planas (quadriláteros, triângulos e círculos), entre outras figuras; (2) que todos os pontos de uma linha circular encontram-se à mesma distância de um ponto central; (3) a conservação da área na decomposição de um quadrado em dois triângulos; • Ler e interpretar horas e minutos em relógio de ponteiros e comparar grandezas de mesma natureza (capacidade) usando cálculos com medidas convencionais e não convencionais; • Localizar na reta numérica, números da ordem das unidades de milhar (intervalos de 50 unidades); • Estabelecer relação entre o resto e o divisor em uma operação; • Resolver problemas que envolvam: (1) operação de divisão com o significado de comparação entre razões; (2) resolução de um cálculo de multiplicação comparativa e envolvendo a idéia de proporcionalidade e de multiplicações sucessivas; (3) adição e subtração de números decimais com combinação de dois estados; (4) subtração envolvendo a invariância da diferença em um problema complexo; (5) estimativa por meio de procedimentos de cálculo mental aproximado; • Ler e interpretar um gráfico e realizar operações com base nele.
$\theta > 300$ 0,5%	<p>Neste nível de proficiência, os alunos de 4ª série demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Identificar e nomear cinco figuras planas (trapézio, círculo, trapézio, triângulo e losango); • Estabelecer trocas em moedas em função de seus valores; • Localizar na reta numérica, números da ordem das unidades de milhar (intervalos de 10 unidades); • Decompor numeral que envolve inteiros e milésimos; • Reconhecer que há diferentes formas de representar um mesmo número decimal e o significado de frações como quocientes; • Estabelecer relação entre o resto e o divisor em uma hipótese e aplicar a propriedade distributiva da multiplicação para resolver problema; • Indicar a construção correta de uma tabela entre vários modelos e considerando-se os dados apresentados.

A escala única produzida e publicada para a educação básica, para a disciplina Matemática, encontra-se transcrita na Tabela 34 (Resultados dos Níveis de Desempenho – Língua Portuguesa e Matemática, retirado do site: http://www.inep.gov.br/basica/saeb/saeb_99.htm pp. 74-84.

Tabela 34 – Níveis de proficiência de Matemática publicados e percentuais de alunos nos níveis – SAEB – 1999

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$\theta < 160$ 33,4%	Sem informação de descrição publicada.
$160 \leq \theta < 175$ 14,5%	Neste nível, os alunos são capazes de identificar a localização de objetos, reconhecer figuras geométricas simples (círculo e quadrado), compreendendo dados apresentados em gráficos de coluna e interpretar medidas simples como litro e grama, em situações do dia-a-dia.
$175 \leq \theta < 225$ 37,7%	Neste nível, os alunos reconhecem figuras geométricas simples e suas propriedades e uma figura após ter sido submetida a um giro de 180°. São capazes de resolver problemas envolvendo adição e subtração, e identificam a decomposição de um número em unidades, dezenas e centenas. Resolvem também problemas relativos a situações do dia a dia (ler horas, conhecer o valor das cédulas e moedas). Interpretam medidas expressas em unidades usuais e são capazes de analisar dados apresentados em gráficos de colunas.
$225 \leq \theta < 275$ 12,4%	Além de reconhecer as figuras planas e seus elementos, neste nível os alunos já distinguem alguns sólidos geométricos e retas paralelas. Identificam vértice e lados de polígonos simples como dos triângulos e quadriláteros. Reconhecem posições relativas de retas no plano. Em representações de figuras tridimensionais, distinguem prismas e pirâmides. São capazes de ler e interpretar medidas de temperatura e de estabelecer relações entre diversas unidades de medida de tempo. Analisam dados

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	apresentados sob forma de tabelas, pictogramas, gráficos de barras e setores.
$275 \leq \theta < 325$ 2,0%	<p>Neste nível, os alunos estabelecem propriedades das operações, efetuam adição e subtração de números decimais, comparam e somam números inteiros relativos, representando-os em uma reta numérica. Efetuam cálculos com números naturais envolvendo potenciação e radiciação. São capazes de utilizar frações para representar uma parte em relação ao todo e de aplicar propriedades da proporcionalidade direta na resolução de problemas. Identificam eixo de simetria em figuras bidimensionais, assim como identificam o número de faces de poliedro. Calculam a quantidade de cubos unitários em que se decompõe um paralelepípedo.</p> <p>Estabelecem relações entre metro e centímetro e reconhecem o metro quadrado como unidade de medida de área. Analisam dados apresentados em tabela de dupla entrada e identificam uma seqüência de dados em um gráfico de linhas.</p>
$325 \leq \theta < 375$ 0,1%	<p>Este nível se caracteriza pelo domínio das operações com números inteiros relativos, e também da comparação de números racionais tanto na forma decimal quanto fracionária. Calculam média aritmética de um conjunto de números. Além de reduzir termos semelhantes em expressões algébricas, os alunos resolvem equações e sistemas de equações de 1º grau. São também capazes de fazer prognósticos a partir de dados apresentados em gráficos, além de resolver problemas envolvendo o conceito de porcentagem. Reconhecem quadriláteros como trapézios e losangos e aplicam propriedades dos polígonos regulares e da esfera. Reconhecem ângulos agudos, retos e obtusos, identificando o grau como medida de ângulo.</p>
$375 \leq \theta < 425$ 0,0%	<p>Neste nível, os alunos trabalham com juros simples e estabelecem relações entre as formas decimal e fracionária de um número racional, incluindo as dízimas periódicas e as porcentagens. Utilizam aproximações para efetuar cálculos e a média aritmética para fazer inferências. Identificam uma função exponencial. Têm noções de trigonometria e geometria analítica e também são capazes de usar as relações trigonométricas no triângulo retângulo, o Teorema de Pitágoras e a Lei Angular de Tales na resolução de problemas. Aplicam a noção de perímetro e área na resolução de problemas. Calculam probabilidades simples.</p>
$425 \leq \theta < 475$ 0,0%	<p>Este nível se caracteriza pelo domínio da maior parte dos conteúdos do Ensino Médio, como progressões, matrizes, análise combinatória, trigonometria, geometria analítica, funções exponenciais e logarítmicas, sabendo resolver equações envolvendo estas funções. Aplicam as relações trigonométricas em triângulo qualquer. São capazes de localizar uma reta num plano cartesiano, dada sua equação e também de estabelecer se um ponto pertence a uma reta neste plano. Encontram interseções de uma reta com os eixos coordenados. Além do reconhecimento da existência dos números imaginários, os alunos resolvem problemas simples envolvendo contagem, calculam o comprimento de uma circunferência e a área de um círculo.</p>

Fonte: MEC/INEP

Cabem aqui os mesmos comentários feitos às edições 1995 e 1997, em relação à descrição da escala no formato único, considerando-se toda a educação básica.

Outra observação a ser feita e que também ocorreu, em menor escala nos testes anteriores, diz respeito à quantidade de itens para descrever a proficiência dos alunos nos níveis. Constata-se que há muito mais itens para descrever os níveis mais altos. Isso decorre do mau planejamento de montagem dos testes, contemplando um maior número de itens mais difíceis e que exigem uma proficiência maior. Este é um problema simples de se resolver, considerando-se que os itens que estão no banco contêm todas as informações dos parâmetros, bastando, portanto, ter o cuidado de verificar se há um número equilibrado de itens cobrindo todo o *continuum* da escala e os conteúdos a serem contemplados.

Outra observação, diz respeito ao cuidado com a elaboração dos itens. Neste teste, embora tenha havido vários itens que descrevessem os descritores relacionados à localização e movimentação de objetos no espaço, a grande maioria não conseguiu contemplar tarefas de

modo a se fazer entender o que era solicitado. Dessa forma, quase todos os itens tiveram um mau funcionamento pedagógico e psicométrico, pouco podendo se dizer do que de fato os alunos desse nível de escolarização construíram.

5.4. SAEB – EDIÇÃO 2001

A Edição 2001 indicou vários fatos marcantes que se relacionaram aos processos de elaboração dos itens, montagem dos testes e apresentação dos resultados. O primeiro fato marcante foi o desenvolvimento de outro documento para dar apoio à elaboração de itens e montagem dos testes. Trata-se do Documento SAEB 2001 – Novas Perspectivas (2002) que se encontra como referência até a data atual. Neste documento constam: os pressupostos teóricos da avaliação nacional, as matrizes de referência, a organização das provas e dos questionários e o plano amostral.

Os outros marcos da Edição 2001 relacionam-se à apresentação dos resultados. Cada um desses marcos esteve associado a uma diretoria distinta. A Diretoria que esteve no período de janeiro de 2001 a janeiro de 2003 foi a responsável pela primeira apresentação de resultados em escalas por série e pelo primeiro relatório voltado para educadores. A Diretoria seguinte que atuou no período de janeiro de 2003 a outubro de 2005, foi a responsável pelo novo formato da escala. Trata-se da nova escala para leitura do desempenho dos alunos: a escala por estágios de aquisição de conhecimentos, ou de construção de competências e desenvolvimento de habilidades, cujos qualificativos que foram arbitrados pela direção à época são: *muito crítico, crítico, intermediário, adequado e avançado* e cujos pontos de corte e descrições saíram também de dentro do gabinete, sem discussões com educadores.

Em relação ao primeiro documento citado (SAEB 2001 – Novas Perspectivas), o seu objetivo foi lançar novas perspectivas sob a avaliação. Quanto às matrizes apresentadas nesse documento, houve várias mudanças em relação às anteriores. Nelas, ficaram mais fortemente evidenciados os eixos norteadores que embasam os testes do SAEB: em Língua Portuguesa, a Leitura, e em Matemática, a Resolução de problemas.

Desde que o SAEB tornou-se mais evidente e que procurou manter uma maior proximidade com professores da base e com gestores das avaliações estaduais, as críticas às matrizes das duas disciplinas e de todos os níveis têm aumentado. Essas críticas têm vindo,

principalmente, por parte dos professores de sala de aula que não conseguem compreendê-las com clareza. A seguir, são apresentadas as novas mudanças.

5.4.1. AVALIAÇÃO DE LÍNGUA PORTUGUESA

A organização da matriz de Língua Portuguesa, segundo esse documento, é resultante da percepção de que ao realizar as práticas de compreensão e produção de textos e de análise lingüística, o aluno *modifica, reorganiza e reconstrói sua visão da língua em funcionamento, tornando-se capaz de estabelecer um número sempre maior de relações substantivas entre ele, como sujeito do conhecimento, e a língua e o objeto a serem conhecidos*. (pp. 17 e 18). Assim, o sujeito que tem domínio da língua é considerado como alguém que sabe fazer uso dos recursos expressivos e que *consegue cumprir melhor seus objetivos comunicativos e suas intenções interlocutórias*.

A avaliação, dessa forma, deve estar orientada para o alcance da competência da linguagem. Os tópicos da matriz foram fundamentados na manifestação da competência do domínio da linguagem pelo sujeito, ou seja, na capacidade expressa do sujeito compreender textos orais e escritos adequados às situações de comunicação em que atua; de posicionar-se criticamente diante do que lê ou ouve, produzindo sentido, formulando perguntas e respostas significativas em variadas situações e descobrindo as finalidades e intenções, os desejos e as convicções implícitas.

Na perspectiva de o texto ser o objeto de estudo, os tópicos da matriz apresentaram conjuntos de habilidades a ele associados, que sob a forma de descritores especificaram o que cada habilidade implica. Assim, puderam-se tomar como referência no processo avaliativo, algumas das competências tidas como essenciais nas situações de leitura (p. 18).

A matriz de Língua Portuguesa abrangeu seis tópicos e o número de descritores variou conforme a série considerada. Foram também atribuídas prioridades aos tópicos e a cada descritor dentro dos tópicos conforme abrangência prevista nas propostas curriculares e informações pelos professores sobre o grau de importância relacionado ao conjunto de tópicos e descritores. A nova estrutura está apresentada na Tabela 35.

Tabela 35 – Estrutura da Matriz de Língua Portuguesa 2001

Tópicos	4ª série EF		8ª série EF e 3ª EM	
	Descritores	Prioridade	Descritores	Prioridade
1) Procedimentos de Leitura	05	P1	05	P3
2) Implicações do suporte, do gênero e/ou enunciador na compreensão do texto	02	P3	02	P5
3) Relação entre textos	01	P5	02	P4
4) Coerência e coesão no processamento do texto	04	P2	07	P1
5) Relações entre recursos expressivos e efeitos de sentido	02	P4	04	P2
6) Variação lingüística	01	P5	01	P5

Embora tenha melhorado bastante a estrutura e o enfoque dado, a matriz continua merecendo um novo olhar no sentido de que ela tenha mais coerência com o eixo adotado (leitura) e no sentido que os professores vejam com mais clareza a sua prática refletida nas habilidades contempladas.

O teste de Língua Portuguesa, 4ª série, foi constituído de 169 questões das quais 157 (93%) foram aproveitadas na descrição do que os alunos brasileiros sabem ou deveriam saber fazer. Em síntese, analisando-se a qualidade do instrumento, pôde-se verificar que, em relação às avaliações anteriores, houve um salto de qualidade. Melhoraram os textos, o formato dos itens, a plausibilidade das alternativas, a abrangência de conteúdos e a correspondência dos itens à matriz.

O que se pôde observar em termos de problemas ao longo do teste, foi a escolha dos textos, que embora tenha melhorado, há ainda alguns inadequados para a fase de desenvolvimento cognitivo em que a maioria desses alunos se encontra e alguns problemas com a construção dos itens. Como exemplo de textos inadequados, pode-se citar algumas crônicas do Millôr Fernandes, do Carlos Drummond de Andrade, Ruben Alves, entre outros, que contêm temas e ironias difíceis dos alunos dessa faixa de idade compreender. Ainda em relação aos textos, constatou-se que há frases sem contexto e relacionadas a aspectos gramaticais sendo contempladas, o que foge ao eixo adotado pelo referencial teórico da avaliação.

Quanto à construção dos itens, percebem-se ainda alguns problemas de plausibilidade, de peguinhas e de dificuldades em expressar com clareza o enunciado. Há itens, por exemplo, que cobram a mesma habilidade (Anexo 3), mas com formatos dos enunciados diferentes. E os alunos têm melhor desempenho em um dos formatos. O que chama a atenção é que em uma avaliação em larga escala, não se deve, em hipótese nenhuma,

dificultar a compreensão do aluno. Por exemplo, neste caso, avaliava-se, por meio de duas questões, se o aluno sabia identificar a finalidade de um mesmo texto de um gênero específico. Na questão em que o enunciado foi claro e objetivo, os alunos tiveram muito melhor desempenho do que naquela em que era solicitada exatamente a mesma coisa, mas de forma indireta. Vejam que o objetivo de uma avaliação em larga escala é constatar o que os alunos sabem, e questões desse tipo podem falsear os resultados da avaliação. Na descrição dos itens, essas questões aparecem descritas como duas habilidades diferenciadas, uma com maior habilidade exigida do que a outra (níveis 200 e 250), quando na verdade o que determinaria um processo de construção desse conhecimento seria o tipo de texto e seu conteúdo, não o formato da formulação da questão.

Quanto aos aspectos psicométricos dos itens, ainda constatou-se alguns problemas já discutidos nas avaliações anteriores, em relação aos intervalos da métrica da escala e à variância dos parâmetros gerados pela TRI.

Na Tabela 36 são apresentadas as estatísticas geradas dos itens selecionados para a descrição da escala de proficiência dos alunos em Língua Portuguesa, 4ª série, 2001. Essa seleção foi feita considerando-se os critérios descritos na metodologia e o atual foco da avaliação do SAEB.

Tabela 36 – Estatísticas – Teste Língua Portuguesa – 4ª Série – SAEB – Edição 2001

	Parâmetro a	Parâmetro b	Parâmetro c	Discriminação Índice D	Correlação bisserial	Proporção de acerto
N	157	157	157	157	157	157
Média	1,61	-1,38	0,14	0,53	0,57	0,49
Desvio Padrão	0,46	0,75	0,09	0,11	0,09	0,14
Mínimo	0,69	-3,17	0,00	0,17	0,28	0,21
Máximo	2,89	0,22	0,37	0,76	0,74	0,83

São apresentadas na Tabela 37 a (re)descrição dos níveis de proficiência que representam o que os alunos brasileiros demonstraram saber fazer, com base no teste aplicado, e o percentual de alunos da 4ª série, que atingiu cada um desses níveis.

Tabela 37 – Descrição dos percentuais de alunos e dos níveis com base no teste de Língua Portuguesa – 4ª Série – 2001

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$\theta < 125$ 22,2%	Neste nível não houve nenhum item que pudesse representá-lo.
$125 \leq \theta < 150$ 17,8%	Os alunos que atingiram este nível de proficiência demonstraram ser capazes de: <ul style="list-style-type: none"> • Localizar informações explícitas e inferir o sentido de uma palavra com auxílio de material gráfico

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<p>(rótulo instrucional e quadrinhos);</p> <ul style="list-style-type: none"> • Identificar o tema central de um texto informativo e um dos elementos (espaço) que constroem a narrativa de um conto; • Estabelecer relação causa/conseqüência entre partes de elementos de um texto em quadrinhos com apoio de ilustração gráfica.
<p>150 ≤ θ < 175 19,0%</p>	<p>Neste nível de proficiência, os alunos demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Interpretar texto com auxílio de material gráfico (quadrinhos); • Inferir informações implícitas em um poema e sobre a condição explícita de um personagem ao longo da história; • Identificar: (1) o tema de uma música (rap); (2) o assunto principal, o narrador, a finalidade, o personagem principal e o significado literal de uma palavra comum do vocabulário de contos infantis; (3) a finalidade de um texto informativo simples; (4) o assunto principal de uma lenda; e (5) as marcas linguísticas que evidenciam o morador da zona rural; • Estabelecer relação causa/conseqüência entre partes e elementos de um texto na forma de poema infantil, fábula, continho infantil, e relações entre partes de uma história em quadrinhos, identificando repetições que contribuem para a continuidade do texto.
<p>175 ≤ θ < 200 17,3%</p>	<p>Alunos que atingiram este nível de proficiência demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Localizar informações explícitas que se encontram claras, no início de um texto informativo e em bilhete, ou parafraseadas de modo simples; • Inferir: (1) uma informação implícita que se encontra na superfície do texto e com auxílio de material gráfico; (2) o sentido de uma expressão comum no texto; (3) sobre o sentimento do personagem principal de um conto e sobre o sentimento desperto na população de referência de um texto informativo; • Identificar: (1) um dos elementos que constroem (localização) a narrativa de textos complexos; (2) o final de uma história; (3) os personagens principais de histórias; (4) o tema central de um texto informativo; (5) as marcas linguísticas que evidenciam o destinatário de uma carta comercial; (6) o significado de uma palavra contida em um conto; (7) o conflito gerador do enredo de uma fábula; (8) o efeito de sentido decorrente de uma onomatopéia; (9) o narrador de uma história; • Estabelecer relação causa/conseqüência entre partes e elementos de um texto informativo, simples e que diz respeito ao universo infantil, de um conto e de um texto em quadrinhos com apoio de ilustração gráfica; • Distinguir um fato de uma opinião relativa a esse fato; • Reconhecer personagens comuns de livros infantis (bruxa e fada), comparando a textos que tratam do mesmo tema.
<p>200 ≤ θ < 225 12,2%</p>	<p>Alunos que alcançaram esse nível de proficiência demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Localizar informações explícitas em contos e lendas de textos de estruturas mais complexas e informações parafraseadas no texto; • Inferir: (1) informações implícitas com base em ilustrações gráficas, ou trocadilhos com palavras, ou repetição de palavras ao longo do texto, ou pelo contexto dos textos de fábulas, poemas, contos e crônicas; (2) o sentido de palavras ou expressões pelo contexto do texto ou pelos recursos utilizados por autores nos diferentes formatos da estrutura de poemas; • Identificar: a finalidade de um texto expresso na forma de música (rap), de propaganda, de bilhete, de carta comercial, de um informativo complexo; (2) o tema central de uma crônica; (3) os vários personagens que compõem uma lenda e dois personagens principais de um poema; (4) um referente que contribui para a continuidade do texto; (5) o conflito gerador do enredo da história; (6) a função do negrito em uma frase; (7) a função de comparação do advérbio <i>como</i>, expresso em uma frase de um texto informativo; (8) o efeito de sentido decorrente do uso de exclamações sucessivas em uma frase e da repetição de um verbo ao longo de um poema; • Estabelecer: (1) relações entre partes de um texto, identificando repetições que contribuem para a continuidade de um texto; (2) relação causa/consequência entre partes de um conto, de um texto fabular, de uma frase isolada e com base no tema central do texto; • Selecionar e substituir uma conjunção causal (<i>porque</i>) por outra sem alterar o sentido da frase (<i>pois</i>); • Interpretar texto com auxílio de material gráfico (quadrinhos).

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
<p>225 ≤ θ < 250 6,7%</p>	<p>Os alunos que atingiram este nível de proficiência demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Localizar informações explícitas em locais diversos de textos instrucionais, com apoio de ilustração e parafraseadas em um texto de propaganda; • Inferir: (1) informações implícitas em crônicas, poemas, textos informativos, com base, principalmente, na compreensão da solução dos problemas explicitados nos textos; (2) o sentido de palavras e expressões comumente ditas em poema infantil e pelo contexto do texto; • Identificar: (1) a função do travessão nos diálogos em textos e em frases isoladas; (2) o motivo de ironias em anedotas; (3) conflitos geradores de enredos que constroem a narrativa de fábulas e histórias em quadrinhos; (4) o efeito de sentido do formato das palavras em um poema e do uso de duas interrogações com sentidos diferentes no texto; (5) características do personagem principal de uma história narrada; (6) marcas lingüísticas que denotam irritação de um personagem do texto; • Estabelecer: (1) relação causa/conseqüência entre partes e elementos do texto em uma fábula e com base nos elementos que compõem uma narrativa; (2) relações entre partes de um texto, identificando substituição que contribui para a continuidade de textos de alta complexidade; (3) relação lógico-semântica (comparação) entre características comuns de dois personagens; • Associar características de personalidade do personagem principal de um texto à ilustração do texto e interpretar texto com auxílio de material gráfico, inferindo uma informação implícita dele.
<p>250 ≤ θ < 275 3,2%</p>	<p>Neste nível de proficiência, os alunos demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Localizar uma informação explícita em um poema complexo e em texto com muitas informações relacionadas a datas; • Inferir informações implícitas com base na opinião do personagem principal de um texto, no efeito de sentido produzido pelo uso de um recurso sintático (repetição) e na repetição de palavras associadas ao sentimento do personagem; • Identificar: (1) o tema central de um texto em quadrinhos com apoio de material ilustrativo e de um poema que não faz parte do universo infantil; (2) efeitos de ironia em palavra com duplo sentido e em trocadilhos com os significados de algumas palavras; (3) marca lingüística que evidencia a fala do narrador da história; (4) a finalidade de um texto complexo na forma de uma música e de um informativo, fora do contexto infantil, e de uma anedota; (5) o sentido expresso pelos recursos utilizados pelo autor no formato diferente de uma das palavras com significado complexo; (6) o efeito de sentido decorrente de uma interrogação no início do texto e colocada sozinha em um balãozinho de história em quadrinhos e de um ponto final em uma frase no meio do texto; (7) um tempo passado identificado pelo tempo verbal; (8) elementos (personagens) que constroem uma narrativa; (9) entre trechos de diferentes gêneros textuais, o que se aproxima a uma fábula; (10) opiniões expressas por personagens de uma história; (11) a quem é dirigida uma propaganda; (12) palavras que com base no senso comum se relacionam a um tempo moderno; • Estabelecer relações entre partes de um texto longo e complexo, identificando substituição que contribuem para a continuidade de um texto; • Associar uma característica física de um animal pela foto expressa no texto e ler e interpretar os dados de um gráfico; • Antecipar e inferir a respeito do conteúdo que o texto trará, com base no tema dado.
<p>275 ≤ θ < 300 1,3%</p>	<p>Neste nível, os alunos demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Inferir uma informação implícita em texto descontextualizado, na forma de uma frase e fora do universo infantil e o sentido de uma expressão não comum ao jargão infantil; • Identificar um dos elementos que constroem a narrativa (localização) de um texto complexo; o tema de um texto informativo complexo com grande quantidade de informações e uma das características de estrutura de um poema; • Distinguir um fato da opinião relativa a esse fato em um texto informativo complexo;
<p>θ > 300 0,4%</p>	<p>Neste nível de proficiência, os alunos demonstraram ser capazes de estabelecer relação causa/consequência entre partes e elementos de um texto complexo, com muitas informações e totalmente fora do contexto infantil.</p>

Da avaliação de 2001 foram produzidas escalas com descrições únicas para a educação básica, escalas por disciplina e série e escalas com descrições dos níveis de

desempenho. Na Tabela 38 estão transcritas as escalas de 4ª série publicadas, com as descrições para Língua Portuguesa (SAEB 2001 – Relatório Nacional, pp. 37 e 38).

Tabela 38 – Níveis de proficiência de Língua Portuguesa publicados e percentuais de alunos de 4ª série nos níveis – SAEB – 2001

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$\theta < 125$ 22,2%	Não houve itens que representasse este nível da escala.
$125 \leq \theta < 150$ 17,8%	<ul style="list-style-type: none"> • Localizam informações explícitas em um texto; • Inferem o sentido de palavras em textos simples (histórias em quadrinhos); • Interpretam texto com auxílio de material gráfico (histórias em quadrinhos); • Identificam a finalidade de texto narrativo simples; • Identificam elementos que constroem uma narrativa simples (espaço); • Estabelecem relação causa/conseqüência entre partes e elementos do texto;
$150 \leq \theta < 175$ 19,0%	<ul style="list-style-type: none"> • Localizam informações explícitas em textos narrativos curtos como histórias infantis (contos de fada e fábulas) e em outros gêneros textuais, como comunicado aos pais e pequenos poemas descritivos e narrativos; • Identificam informação implícita acompanhada de ilustrações em histórias em quadrinhos e em textos narrativos simples (prosa ou em verso) e em texto descritivo (fábula); • Identificam o tema de texto narrativo simples (histórias infantis) com apoio de elementos explícitos (figuras); • Distinguem um fato da opinião relativa a esse fato em textos narrativos simples (histórias em quadrinhos e texto com ou sem ilustrações); • Identificam elementos constitutivos da narrativa (personagens); • Estabelecem relações de continuidade em histórias em quadrinhos • Estabelecem relação causa/conseqüência em poemas curtos; • Identificam marcas lingüísticas próprias do meio rural e urbano
$175 \leq \theta < 200$ 17,3%	<ul style="list-style-type: none"> • Localizam informações explícitas em textos mais complexos, mais longos e não narrativos (texto publicitário); • Inferem o sentido de palavra ou expressão em texto narrativo (história infantil); • Identificam o tema de um texto informativo (em revista); • Identificam a finalidade de textos de diferentes gêneros; • Identificam elementos que identificam o conflito gerador do enredo; • Estabelecem relação causa/conseqüência em textos poéticos e em textos de complexidade média (fábula); • Estabelecem relações lógico-discursivas presentes no texto, marcadas por conjunções; • Identificam marcas lingüísticas que caracterizam o locutor e o interlocutor do texto;
$200 \leq \theta < 250$ 18,9%	<ul style="list-style-type: none"> • Localizam informações explícitas em textos publicitários de revistas e jornais; • Inferem o sentido de expressão de uso corrente em textos informativos; • Identificam informação implícita em narrativa curta, com material ilustrativo; • Identificam o tema de texto poético de baixa complexidade (poema descritivo) e em narrativa curta (fábula) com base em material ilustrativo; • Identificam elementos espaço (em crônicas), personagens, conflito e desfecho (em histórias infantis); • Estabelecem relações de continuidade por meio de relações anafóricas em textos curtos e simples; • Identificam o efeito de sentido decorrente da disposição gráfica das palavras em um texto;

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<ul style="list-style-type: none"> • Identificam marcas lingüísticas próprias de textos comerciais;
$250 \leq \theta < 300$ 4,4%	<ul style="list-style-type: none"> • Localizam informações explícitas em textos poéticos não descritivos (poemas narrativos simples e canção popular); • Estabelecem relações de continuidade por meio de relações anafóricas e textos de complexidade média; • Identificam o efeito de sentido decorrente do uso expressivo da pontuação (ponto de interrogação) e da repetição de uma letra em uma palavra;

Fonte: MEC/INEP

5.4.2. AVALIAÇÃO DE MATEMÁTICA

O teste de Matemática do SAEB 2001 foi elaborado com base no Documento SAEB – Novas Perspectivas (2002), o qual já foi descrito no início deste tópico.

A matriz de Matemática, segundo esse documento, está fundamentada no princípio de que o ensino da matemática não deve estar limitado a um simples processo de memorização de regras e técnicas. A Matriz toma como eixo norteador da atividade matemática a resolução de problemas, pois, segundo o documento, essa orientação *possibilita o desenvolvimento de capacidades como: observação, estabelecimento de relações, comunicação (diferentes linguagens), argumentação e validação de processos, além de estimular formas de raciocínio como intuição, indução, dedução e estimativa.* (p. 23).

Para que a avaliação de matemática do SAEB possa refletir essa nova perspectiva do ensino, é preciso que o teste se constitua basicamente por situações em que a resolução de problemas seja significativa para o aluno. Dessa forma, o teste deverá contemplar tarefas que permitam verificar em que medida a aprendizagem foi efetivamente realizada. Trata-se de contemplar atividades que permitam avaliar o quanto e o como a escola preparou o aluno para enfrentar os desafios que o mundo lhe impõe e até que ponto ele está pronto para exercer com plenitude a sua cidadania.

A avaliação, dessa forma, deve estar orientada para a verificação da competência matemática vivenciada no cotidiano. Os temas da matriz foram fundamentados na manifestação da competência do domínio do saber matemático pelo sujeito, ou seja, na capacidade expressa do sujeito compreender as situações, lançar mão de estratégias, posicionar-se diante do problema e buscar a melhor solução. Dessa forma, o aluno se sentirá estimulado e motivado a dar respostas significativas às variadas situações e terá oportunidade de descobrir a finalidade e a importância do ensino da matemática em sua vida.

A matriz de Matemática abrangeu quatro temas e o número de descritores variou conforme a série considerada. Foram também atribuídas prioridades aos temas e a cada descritor dentro dos temas, conforme abrangência prevista nas propostas curriculares e informações pelos professores sobre o grau de importância relacionado ao conjunto de temas e descritores. A nova estrutura está representada na Tabela 39.

Tabela 39 – Estrutura das Matrizes de Matemática para a Educação Básica – SAEB 2001

Temas	4ª série		8ª série		3ª série	
	Descritores	Prioridade	Descritores	Prioridade	Descritores	Prioridade
Espaço e Forma	05	P3	11	P2	10	P3
Grandezas e Medidas	07	P2	04	P3	03	P4
Números e operações, Álgebra e funções	14	P1	20	P1	20	P1
Tratamento da Informação	02	P4	02	P2	02	P2

O teste foi constituído de 169 itens dos quais 155 (92%) foram considerados para proceder à descrição do que os alunos brasileiros sabem ou deveriam saber fazer. Em síntese, analisando-se a qualidade do instrumento, pôde-se verificar que, em relação aos testes de edições anteriores do SAEB, houve uma melhoria na qualidade. Muito mais itens foram aproveitados. Contudo, observa-se que ainda não se atingiu o padrão esperado de qualidade, posto que, há ainda uma distância entre o que deve compor um instrumento de avaliação em larga escala e o que está sendo contemplado neste teste avaliado.

Constata-se que a maioria das questões desse teste foram muito difíceis para a população alvo; vários estudiosos citados no marco teórico deste estudo (Bloom, 1971; Vianna, 1987, Pasquali, 1997; Anastasi e Urbina, 2000; McIntire e Miller, 2000;) recomendam que a prova seja constituída de uma maior parte de itens de dificuldade média e percentuais menores de itens com maior e menor dificuldade. Isso possibilita atingir o *continuum* da escala e minimizar outros fatores intervenientes no desempenho do aluno, como, por exemplo, a voluntariedade do aluno para fazer o teste; ao fazer um teste com questões muito difíceis (cinco blocos tiveram proporção de acerto menor que 0,37 – blocos 1, 2, 11, 12 e 13), ele pode se sentir desestimulado e realizá-lo sem empenho. Dessa forma, o teste deveria ser mais bem planejado.

Em relação aos itens, observa-se que a maior parte deles se constituiu de itens contextualizados. Mas os contextos não foram adequados aos alunos da faixa etária prevista. Houve também questões com peguinhas, que podem ser constatados, principalmente, pelos

coeficientes bisseriais positivos nas alternativas falsas, indicando que os alunos que, hipoteticamente sabem mais, escolhem uma alternativa falsa. Em relação aos enunciados, estes se apresentaram pouco claros e complicados.

Constata-se, novamente, o desajuste da escala em decorrência da mudança da métrica mostrando, mais uma vez, que aparentemente seria necessária mais proficiência para resolver os mesmos itens que anteriormente exigiam um quantitativo menor de habilidade.

Quanto à abrangência dos conteúdos, pontua-se que houve para os temas, considerando-se as prioridades e o número de descritores por tema, uma boa distribuição. Internamente, em relação aos descritores contemplados, parece ter havido certa descompensação entre eles. Constata-se que houve privilégios de alguns descritores não considerados prioritários na matriz de referência e menos atenção aos que são considerados mais importantes. A distribuição de itens por tema da matriz ficou da seguinte forma: Números e operações, com 57%; Grandezas e medidas, 21%; Espaço e forma, 14%; e Tratamento da informação, 8%.

Na Tabela 40 são apresentadas as estatísticas geradas dos itens selecionados para a descrição da escala de proficiência dos alunos em Matemática, 4ª série, 2001. Essa seleção foi feita considerando-se os critérios descritos na metodologia e o atual foco da avaliação do SAEB.

Tabela 40 – Estatísticas – Teste Matemática – 4ª Série – SAEB – Edição 2001

	Parâmetro a	Parâmetro b	Parâmetro c	Discriminação Índice D	Correlação bisserial	Proporção de acerto
N	155	155	155	155	155	155
Média	1,81	-0,85	0,15	0,51	0,55	0,44
Desvio Padrão	0,49	0,76	0,09	0,13	0,10	0,14
Mínimo	0,74	-2,96	0,00	0,10	0,31	0,06
Máximo	3,17	1,20	0,45	0,73	0,72	0,84

A Tabela 41 apresenta a (re)descrição dos níveis de proficiência que representam o que os alunos brasileiros demonstraram saber fazer, com base no teste aplicado, e o percentual de alunos da 4ª série, que atingiu cada um desses níveis.

Tabela 41 – Descrição dos percentuais de alunos e dos níveis s com base no teste de Matemática – 4ª Série – 2001

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$\theta < 125$ 12,5%	Neste nível não houve nenhum item que o representasse.
$125 \leq \theta < 150$ 18,6%	Alunos da 4ª série que atingiram esse nível de proficiência demonstraram ser capazes de: <ul style="list-style-type: none"> • Calcular a área de uma figura plana com o apoio da malha quadriculada;

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<ul style="list-style-type: none"> • Identificar fração com base em uma figura concreta (torta), representando parte – todo.
<p>150 ≤ θ < 175 20,3%</p>	<p>Neste nível de proficiência, os alunos da 4ª série demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Ler e interpretar dados apresentados em tabelas com poucos dados e gráficos de colunas.
<p>175 ≤ θ < 200 19,0%</p>	<p>Alunos de 4ª série que alcançaram este nível de proficiência demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Em Espaço e forma, <ul style="list-style-type: none"> ➢ identificar a movimentação e a localização de objeto por meio de mapa e croqui; • Em Grandezas e medidas, <ul style="list-style-type: none"> ➢ estimar medida de grandeza (capacidade) utilizando unidades de medida não convencionais; ➢ resolver problemas envolvendo situação simples de trocas entre cédulas do sistema monetário brasileiro, em função de seus valores; ➢ determinar tempo que falta para o início de um evento e o intervalo de tempo que durará um evento dadas as referências; • Em Números e operações, <ul style="list-style-type: none"> ➢ armar e efetuar adição de três parcelas de diferentes ordens e calcular o resultado de uma subtração com recursos, de uma multiplicação de um número da ordem das centenas por unidades; ➢ resolver problemas envolvendo: representação fracionária, com apoio de ilustração (pizza); adição de números racionais escritos na forma decimal e utilizando a escrita decimal de cédulas e moedas do sistema monetário brasileiro; • Em Tratamento da informação, <ul style="list-style-type: none"> ➢ ler e interpretar dados das colunas representados em tabelas e gráficos.
<p>200 ≤ θ < 225 13,8%</p>	<p>Neste nível de proficiência, os alunos de 4ª série demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Em Espaço e forma, <ul style="list-style-type: none"> ➢ identificar: entre figuras tridimensionais, as que possuem superfície arredondada (cone e esfera); a modificação de medidas dos lados pela ampliação de figuras poligonais com apoio de malhas quadriculadas; • Em Grandezas e medidas, <ul style="list-style-type: none"> ➢ identificar a operação que deve ser feita para calcular a área de uma figura plana com apoio da malha quadriculada; ➢ estabelecer intervalo de duração de tempo, dados horário de início e término das atividades; ➢ resolver problemas estabelecendo conversão entre número de dias e semanas, de horas em dias e envolvendo trocas entre cédulas e moedas do sistema monetário brasileiro, em função de seus valores; • Em Números e operações, <ul style="list-style-type: none"> ➢ identificar: fração representada em figura parte-todo; o maior número que se pode formar com base em três algarismos e o menor, com base em quatro, pelo princípio do valor posicional; o valor posicional de um algarismo em um número; a composição e decomposição de números naturais na forma polinomial; ➢ calcular: os resultados de subtrações de números naturais envolvendo números de diversas ordens, com zeros intermediários; os resultados de multiplicações simples e de divisões exatas de números naturais; ➢ resolver problemas envolvendo: subtrações, no sentido de alteração positiva de um estado inicial e utilizando a propriedade da invariância da diferença; divisão em quantias iguais e com resto; subtração de números racionais, com duas casas decimais; situação de compra com troco, utilizando a escrita decimal de cédulas e moedas do sistema monetário; • Em Tratamento da informação, <ul style="list-style-type: none"> ➢ ler e interpretar dados representados em calendário e relógio de ponteiros (horas, dia, mês e ano) e dados apresentados em tabelas e gráfico de colunas, dos quais é solicitado calcular o resultado da soma de colunas.
<p>225 ≤ θ < 250 8,1%</p>	<p>Alunos da 4ª série que atingiram esse nível de proficiência demonstraram ser capazes de:</p> <p>Em espaço e forma,</p> <ul style="list-style-type: none"> ➢ identificar, entre outras figuras planas, as que têm o mesmo número de lados e a movimentação de objeto com apoio de croqui;

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<p>Em Grandezas e medidas,</p> <ul style="list-style-type: none"> ➤ estabelecer relações entre o horário de início e término de um evento com e sem o apoio de um relógio de ponteiros; ➤ estimar medidas de grandezas utilizando unidades de medidas convencionais (kg/g, m); resolver problemas envolvendo a estimativa de áreas de figuras planas, desenhadas em malhas quadriculadas e trocas entre cédulas e moedas do Sistema Monetário Brasileiro em função dos seus valores ➤ identificar, por meio de comparação em termômetros, temperaturas indicativas de climas ambientais diferentes; <p>Em Números e operações,</p> <ul style="list-style-type: none"> ➤ identificar: o valor de determinados algarismos pelo princípio do valor posicional; a localização de números naturais e racionais, na forma decimal, na reta numérica em diferentes intervalos; a composição de números naturais decompostos da ordem das unidades de milhar e a composição e a decomposição na forma polinomial de números naturais de diversas ordens, com zeros intermediários; expressão numérica que resolve um problema e que um mesmo problema pode ser resolvido de diferentes formas; ➤ resolver problemas envolvendo: a propriedade da invariância da diferença e do elemento neutro na multiplicação; procedimentos de cálculo mental aproximado; relação de proporção; a adição e subtração; a divisão e adição de números naturais; a subtração de números racionais escritos na forma decimal; a compra com troco e a multiplicação e adição utilizando a escrita decimal de cédulas e moedas; ➤ calcular: o resultado de uma subtração, com zeros no minuendo; de uma divisão simples; <p>Em Tratamento da informação,</p> <ul style="list-style-type: none"> ➤ ler e interpretar dados representados em gráfico de setores e estabelecer correspondência entre os dados de uma tabela e um gráfico.
<p>250 ≤ θ < 275 4,2%</p>	<p>Alunos de 4ª série que alcançaram esse nível de proficiência demonstraram ser capazes de</p> <p>Em Espaço e forma,</p> <ul style="list-style-type: none"> ➤ identificar: localização de pessoas perto e longe de objeto representado em croqui; quadriláteros entre outras figuras pelo número de lados; diferentes figuras planas (quadriláteros, triângulos e círculos); propriedades dos quadriláteros relacionadas à medida dos lados e dos ângulos; <p>Em Grandezas e medidas,</p> <ul style="list-style-type: none"> ➤ identificar símbolo da unidade de medida adequada para medir área de superfície (m²); ➤ estabelecer período de tempo em meses, dados os referenciais; ➤ estimar a medida de superfície de uma área e a medida do comprimento de um objeto utilizando unidades de medida não convencionais; ➤ resolver problema envolvendo: conversão de unidades de medida padronizadas de mesma grandeza (ml/l); comparação de grandezas de mesma natureza, utilizando unidades de medida convencionais e não convencionais; trocas entre cédulas do sistema monetário brasileiro, em função de seus valores; ➤ ler e interpretar horas e minutos em relógio de ponteiros; <p>Em Números e operações,</p> <ul style="list-style-type: none"> ➤ identificar: a localização de números naturais na reta numérica, simulando a linha do tempo; o valor posicional de um algarismo na construção da escrita de um número pelo sistema de numeração; a composição e decomposição de números de diferentes ordens, com zeros intermediário; ➤ classificar em ordem crescente números inteiros e decimais com uma e três casas decimais; ➤ calcular os resultados e os restos de divisões de números naturais (centenas e unidades de milhar por dezena); ➤ resolver problemas envolvendo: multiplicação com idéia de proporcionalidade; noções de porcentagem (50%); a propriedade da invariância da diferença; operações de adição e subtração com sentido de juntar e comparar em uma transformação positiva e negativa; soma de quantias em dinheiro, combinando dois estados para chegar a um terceiro; <p>Em Tratamento da informação,</p> <ul style="list-style-type: none"> ➤ ler e interpretar tabela e comparar medidas de grandeza de mesma natureza (peso e altura).
<p>275 ≤ θ < 300</p>	<p>Neste nível de proficiência, os alunos de 4ª série demonstraram ser capazes de:</p> <p>Em Espaço e forma,</p> <ul style="list-style-type: none"> ➤ identificar: a movimentação de pessoas por meio de representação gráfica; características

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
1,8%	<p>comuns de quadriláteros (nº de lados); que os pontos de uma linha circular encontram-se à mesma distância de um ponto central;</p> <p>Em Grandezas e medidas,</p> <ul style="list-style-type: none"> ➤ identificar a modificação de medidas do perímetro por meio da redução de figura poligonal usando malha quadriculada; ➤ calcular o perímetro de uma figura plana desenhada em uma malha quadriculada; ➤ estabelecer relação entre unidades de tempo por meio do tempo de validade de produto (anos e meses); ➤ resolver problemas significativos utilizando unidades de medida padronizadas (kg/g); <p>Em Números e operações,</p> <ul style="list-style-type: none"> ➤ identificar: a localização de números racionais, na forma decimal, representados entre inteiros na reta numérica; o maior valor que um determinado algarismo pode representar em unidades, em função de sua posição no quadro valor de lugar, entre números da ordem das dezenas de milhar; a decomposição de números pelo princípio do valor posicional e na forma polinomial, de ordens diversas; a representação decimal de um número fracionário e a representação fracionária como representação associada ao significado parte-todo; ➤ resolver problema envolvendo: a propriedade da invariância da diferença; multiplicação com idéia de configuração retangular e com idéia de análise combinatória; noções de porcentagem (25% e 50%); operação de divisão com o significado de comparação entre razões; <p>Tratamento da informação,</p> <ul style="list-style-type: none"> ➤ ler e interpretar tabela com três colunas e seis linhas.
$300 \leq \theta < 325$ 0,6%	<p>Neste nível de proficiência, os alunos de 4ª série demonstraram ser capazes de:</p> <p>Em espaço e forma,</p> <ul style="list-style-type: none"> ➤ Reconhecer uma imagem espelhada; <p>Em Grandezas e medidas,</p> <ul style="list-style-type: none"> ➤ Determinar intervalo de tempo entre horário de saída e de retorno; ➤ Comparar grandezas de mesma natureza, utilizando cálculos com unidades convencionais (m/cm); <p>Em Números e operações,</p> <ul style="list-style-type: none"> ➤ Identificar: a composição de um número decomposto na forma polinomial; uma operação utilizada para resolver um problema e o significado de frações como quocientes; ➤ Resolver problema utilizando a propriedade distributiva da multiplicação.
$\theta > 325$ 0,8%	<p>Em Espaço e forma,</p> <ul style="list-style-type: none"> ➤ Distinguir figuras tridimensionais pelo número de vértices, faces e arestas; <p>Em Números e operações,</p> <ul style="list-style-type: none"> ➤ Comparar e ordenar seqüência de números racionais na forma decimal, envolvendo décimo, centésimo e milésimo.

Na Tabela 42 está transcrita a escala de 4ª série publicada, com as descrições para Matemática (SAEB 2001 – Relatório Nacional, pp. 108-111).

Tabela 42 – Níveis de proficiência de Matemática publicados e percentuais de alunos nos níveis – SAEB – 2001

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$125 \leq \theta < 150$ 18,6%	<p>Espaço e Forma:</p> <ul style="list-style-type: none"> ➤ Calculam áreas de figuras desenhadas em malhas quadriculadas; <p>Números e Operações:</p> <ul style="list-style-type: none"> ➤ Reconhecem partes de um todo em representações gráficas;
$150 \leq \theta < 175$ 21,1%	<p>Números e Operações:</p> <ul style="list-style-type: none"> ➤ Resolvem problemas do cotidiano envolvendo adições de pequenas quantias em dinheiro; <p>Tratamento da Informação:</p>

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
<p>175 ≤ θ < 200 19,0%</p>	<p>➤ Lêem informações em tabelas e gráficos de colunas;</p> <p>Espaço e Forma</p> <ul style="list-style-type: none"> ➤ Reconhecem a localização ou movimentação de objeto situado em seu referencial em representações gráficas; <p>Grandezas e medidas</p> <ul style="list-style-type: none"> ➤ Estimam valores de uma mesma medida; ➤ Lêem horas e minutos em relógio digital; ➤ Resolvem problemas envolvendo intervalos de tempo em situações cotidianas; <p>Números e Operações</p> <ul style="list-style-type: none"> ➤ Resolvem problemas do cotidiano envolvendo adição de números naturais e racionais com o mesmo número de casas decimais; ➤ Calculam o resultado de uma adição e de uma subtração envolvendo números de até três algarismos, inclusive com recurso e reserva; e de multiplicação com um algarismo; ➤ Identificam frações com apoio de representação gráfica; <p>Tratamento da informação</p> <ul style="list-style-type: none"> ➤ Interpretam gráfico de colunas através de leitura de valores no eixo vertical;
<p>200 ≤ θ < 250 21,9%</p>	<p>Espaço e Forma</p> <ul style="list-style-type: none"> ➤ Distinguem sólidos que têm superfície arredondada; <p>Grandezas e medidas</p> <ul style="list-style-type: none"> ➤ Lêem horas e minutos em relógios de ponteiros; ➤ Identificam moedas para trocar uma quantia pequena de dinheiro; ➤ Realizam conversão entre medidas de tempo (hora, dia, semana) <p>Números e Operações</p> <ul style="list-style-type: none"> ➤ Resolvem problemas envolvendo subtração de números racionais com o mesmo número de casas; ➤ Resolvem problemas simples com as quatro operações; ➤ Reconhecem o valor posicional do Sistema de Numeração Decimal; ➤ Decompõem um número natural em suas ordens e vice-versa; ➤ Calculam o resultado de subtrações mais complexas com números naturais; de multiplicações com números de dois algarismos; de divisões exatas por um algarismo; <p>Tratamento da Informação</p> <ul style="list-style-type: none"> ➤ Resolvem problemas simples, usando dados em gráficos de barras ou tabelas;
<p>250 ≤ θ < 300 6,0%</p>	<p>Espaço e Forma</p> <ul style="list-style-type: none"> ➤ Reconhecem a localização ou movimentação de objeto em referencial diferente do seu, em representações gráficas; ➤ Utilizam propriedades dos polígonos, identificando seus lados; <p>Grandezas e Medidas</p> <ul style="list-style-type: none"> ➤ Lêem horas e minutos, convertendo a duração de tempo; ➤ Resolvem problemas com intervalo de tempo (meses); ➤ Realizam conversão de kg para g, de l para ml; ➤ Calculam o perímetro de figura poligonal conhecendo a medida dos lados; ➤ Reconhecem a unidade de medida de área como o metro quadrado; ➤ Estimam comprimento usando medidas não-convencionais; <p>Números e Operações</p> <ul style="list-style-type: none"> ➤ Resolvem problemas: que utilizam multiplicação envolvendo a noção de proporcionalidade; envolvendo mais de uma operação, incluindo o sistema monetário; de composição e decomposição mais complexas (mais zeros e ordens); ➤ Decompõem um número natural em agrupamentos de 1000; ➤ Calculam o resultado de uma divisão por número de dois algarismos, inclusive com resto e de porcentagens simples; ➤ Identificam número em intervalo da reta numérica; ➤ Reconhecem a composição/decomposição na escrita decimal, em casos mais complexos;

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<ul style="list-style-type: none"> ➤ Comparam números racionais na forma decimal; Tratamento da Informação <ul style="list-style-type: none"> ➤ Interpretam gráfico de setores, associando-os a dados em uma tabela;
$300 \leq \theta < 350$ 0,8%	Espaço e Forma <ul style="list-style-type: none"> ➤ Distinguem sólidos com forma esférica, quadrado fora da posição usual e elementos de figuras tridimensionais (faces, vértices e arestas); Grandezas e Medidas <ul style="list-style-type: none"> ➤ Realizam conversão de m para cm; Números e Operações <ul style="list-style-type: none"> ➤ Resolvem problema envolvendo divisão com resto ou multiplicação combinatória e envolvendo proporcionalidade; Tratamento da Informação <ul style="list-style-type: none"> ➤ Interpretam gráfico de linhas correspondentes a uma seqüência de valores (positivos e negativos)

Fonte: MEC/INEP

Nota: Os percentuais referentes ao quantitativo de alunos nos níveis são os apresentados no relatório SAEB

Em relação à descrição das escalas, o SAEB 2001 constituiu-se um marco. Constatase um grande avanço decorrente das descrições das escalas por série e por nível de desempenho. Pode-se observar pelas descrições por série, que os alunos de 4ª série agora são identificados na avaliação e que o professor tem oportunidade de conferir sua prática com base nos resultados do desempenho dos alunos.

Em relação às escalas por níveis de desempenho, nas Tabelas 43 e 44 são apresentadas as de Língua Portuguesa e Matemática – 4ª série.

Tabela 43 – Construção de competências e desenvolvimento de habilidades de leitura de textos de gêneros variados em cada um dos estágios para a 4ª série do ensino fundamental

MUITO CRÍTICO 22,2%	Não desenvolveram habilidades de leitura. Não foram alfabetizados adequadamente. Não conseguem responder aos itens da prova. Os alunos neste estágio não alcançaram o Nível 1 da escala do SAEB.
CRÍTICO 36,8%	Não são leitores competentes, lêem de forma truncada, apenas frases simples. Os alunos neste estágio estão localizados nos Níveis 1 e 2 da escala do SAEB.
INTERMEDIÁRIO 36,2%	Começando a desenvolver as habilidades de leitura, mas ainda aquém do nível exigido para a 4ª série. Os alunos neste estágio estão localizados no Nível 5 da escala do SAEB.
ADEQUADO 4,4%	São leitores com nível de compreensão de textos adequados à 4ª série. Os alunos neste estágio estão localizados no Nível 5 da escala do SAEB.
AVANÇADO 0,4%	São leitores com habilidades consolidadas, algumas com nível além do esperado para a 4ª série. Os alunos neste estágio estão localizados no Nível 6 da escla do SAAEB.

Fonte: MEC/INEP

Tabela 44 – Construção de competências e desenvolvimento de habilidades na resolução de problemas em cada um dos estágios para a 4ª série do ensino fundamental

MUITO CRÍTICO 12,5%	Não conseguem transpor para uma linguagem matemática específica, comandos operacionais elementares compatíveis com a 4ª série. (Não identificam uma operação de soma ou subtração envolvida no problema ou não sabem o significado geométrico de figuras simples). Os alunos neste estágio não alcançaram o nível 1 da escala do SAEB.
CRÍTICO 39,8%	Desenvolvem algumas habilidades elementares de interpretação de problemas aquém das exigidas para a 4ª série. (Identificam uma operação envolvida no problema e nomeiam figuras geométricas planas mais conhecidas). Os alunos neste estágio alcançaram os Níveis 3 ou 4 da escala do SAEB.
INTERMEDIÁRIO 40,9%	Desenvolvem algumas habilidades de interpretação de problemas, porém insuficientes ao esperado para os alunos da 4ª série (Identificam, sem grande precisão, até duas operações e alguns elementos geométricos envolvidos no problema). Os alunos neste estágio alcançaram os níveis 3 ou 4 da escala do SAEB.
ADEQUADO 6,8%	Interpretam e sabem resolver problemas de forma competente. Apresentam as habilidades compatíveis com a 4ª série. (Reconhecem e resolvem operações com números racionais, de soma, subtração, multiplicação e divisão, bem como elementos e características próprias das figuras planas). Os alunos neste estágio alcançaram o Nível 5 ou 6 da escala do SAEB.
AVANÇADO 0,0%	São alunos maduros. Apresentam habilidades de interpretação de problemas num nível superior ao exigido para a 4ª série (Reconhecem, resolvem e sabem transpor para situações novas, todas as operações com números racionais envolvidas num problema, bem como elementos e características das figuras geométricas planas). Os alunos deste estágio alcançaram o Nível 7 da escala do SAEB.

Fonte: MEC/INEP

Discussões nacionais devem ser feitas para definir os qualificativos e para descrever o que os alunos são capazes. Tomando como referência essas escalas publicadas e aqui transcritas, observam-se várias inconsistências. Em primeiro lugar, no documento não são apresentados os intervalos dos níveis de proficiência relativos aos estágios apresentados. Em segundo lugar, as descrições não coadunam com o que de fato os alunos são capazes e, em terceiro lugar, não é recomendado fazer descrições do que os alunos não alcançaram, mas sim do que foram capazes.

Considerando-se, por exemplo, o que está descrito, e que foi publicado, no nível crítico dessa escala: *Não são leitores competentes, lêem de forma truncada, apenas frases simples*, e no qual consta que os alunos deste nível estariam nos níveis 1 e 2, supostamente, níveis 125 ao 175, observa-se que há uma inconsistência considerável. Comparando-se as duas escalas, a dos níveis de proficiência (descrita pelo SAEB) e dos estágios de competências, não dá para compreender a relação que deveria existir entre elas. Observem na Tabela 45 as duas descrições supostamente relacionadas.

Tabela 45 – Comparação entre a descrição do estágio crítico e as descrições dos níveis 125 ao 175

Estágio Crítico	$125 \leq \theta < 150$
<p>Não são leitores competentes, lêem de forma truncada, apenas frases simples.</p>	<ul style="list-style-type: none"> • Localizam informações explícitas em um texto; • Inferem o sentido de palavras em textos simples (histórias em quadrinhos); • Interpretam texto com auxílio de material gráfico (histórias em quadrinhos); • Identificam a finalidade de texto narrativo simples; • Identificam elementos que constroem uma narrativa simples (espaço); • Estabelecem relação causa/conseqüência entre partes e elementos do texto;
	<p style="text-align: center;">$150 \leq \theta < 175$</p> <ul style="list-style-type: none"> • Localizam informações explícitas em textos narrativos curtos como histórias infantis (contos de fada e fábulas) e em outros gêneros textuais, como comunicado aos pais e pequenos poemas descritivos e narrativos; • Identificam informação implícita acompanhada de ilustrações em histórias em quadrinhos e em textos narrativos simples (prosa ou em verso) e em texto descritivo (fábula); • Identificam o tema de texto narrativo simples (histórias infantis) com apoio de elementos explícitos (figuras); • Distinguem um fato da opinião relativa a esse fato em textos narrativos simples (histórias em quadrinhos e texto com ou sem ilustrações); • Identificam elementos constitutivos da narrativa (personagens); • Estabelecem relações de continuidade em histórias em quadrinhos • Estabelecem relação causa/conseqüência em poemas curtos; • Identificam marcas lingüísticas próprias do meio rural e urbano

A relação entre as duas descrições dispensa comentários.

Quanto à descrição das escalas por série, constata-se ainda que se faz necessário um grande cuidado com a linguagem usada para descrever os itens e apresentar os resultados. Como já comentado anteriormente, professores desse nível não têm formação especializada em nenhuma das áreas analisadas, portanto, deve-se procurar evitar termos que sejam pouco comuns e muito específicos das áreas, não com a pretensão de nivelar por baixo, mas de se fazer entender e, na medida do possível, aproveitar a oportunidade para ensinar. Pode-se, por exemplo, colocar entre parênteses o significado e o conceito de termos matemáticos e da língua portuguesa menos comuns, pois assim, o professor poderá se apropriar de novos conhecimentos.

Observa-se que é preciso um cuidado ainda maior em relação à elaboração dos itens e à montagem dos testes, pois à medida que se opta por esse tipo de descrição e apresentação dos resultados, mais sujeito a críticas estará o sistema. Torna-se necessário, dessa forma,

aperfeiçoar um pouco mais esses dois processos, pois embora tenha havido uma melhoria significativa, grande parte dos erros comentados sobre as edições anteriores repetem-se aqui.

Este foi o primeiro passo para o relato de resultados com um formato capaz de atingir diversos públicos. Comparando-o aos processos de elaboração de itens e da montagem dos testes que, como pode ser observado, a cada edição do SAEB tem se aprimorado mais, este deve ser um processo que, da mesma forma, se amadurecerá ao longo do tempo. O importante é que se mantenha o propósito de atingir os mais diversos públicos, de maneira clara e objetiva, buscando disseminar esse diálogo com os educadores para que melhorias possam ocorrer tendo por base os processos avaliativos.

5.5. SAEB – EDIÇÃO 2003

Em 2003, foi mantido o mesmo documento para dar apoio à elaboração de itens das duas disciplinas avaliadas (Documento SAEB 2001 – Novas Perspectivas, 2002), que é o que se encontra como referência até a data atual. Todos os referenciais se mantiveram. A seguir, as discussões sobre os dois testes.

5.5.1. AVALIAÇÃO DE LÍNGUA PORTUGUESA

O teste de Língua Portuguesa, 4ª série, foi constituído de 169 questões das quais 148 (88%) foram aproveitadas na descrição do que os alunos brasileiros sabem ou deveriam saber fazer. Em síntese, analisando-se a qualidade do instrumento, pôde-se verificar que, em relação às avaliações anteriores, o teste vem melhorando. Constata-se uma melhor escolha dos textos para a avaliação, constituindo-se agora de textos mais apropriados para os alunos desse nível avaliado, embora nem todos tenham esse caráter. Ainda se observa problemas em relação ao formato e este foi o problema principal de exclusão de itens. Foram detectados problemas com a falta de clareza dos enunciados e com a plausibilidade das alternativas. Estes são problemas recorrentes constatados e já discutidos em todas as avaliações analisadas até o momento.

Quanto ao problema da variância dos intervalos definidos para a escala a ser descrita pelos itens selecionados é um problema que permanece e que já foi chamada a atenção anteriormente.

Na Tabela 46 são apresentadas as estatísticas geradas dos itens selecionados para a descrição da escala de proficiência dos alunos em Língua Portuguesa, 4ª série, 2003. Essa seleção foi feita considerando-se os critérios descritos na metodologia e o atual foco da avaliação do SAEB.

Tabela 46 – Estatísticas – Teste Língua Portuguesa – 4ª Série – SAEB – Edição 2003

	Parâmetro a	Parâmetro b	Parâmetro c	Discriminação Índice D	Correlação bisserial	Proporção de acerto
N	148	148	148	148	148	148
Média	1,84	-1,38	0,15	0,51	0,56	0,52
Desvio Padrão	0,60	0,78	0,09	0,11	0,10	0,16
Mínimo	0,69	-3,15	0,00	0,20	0,30	0,17
Máximo	3,96	0,06	0,39	0,73	0,74	0,84

Na Tabela 47 está apresentada a descrição dos níveis de proficiência que representam o que os alunos brasileiros demonstraram saber fazer, com base no teste aplicado, e o percentual de alunos da 4ª série, que atingiu cada um desses níveis.

Tabela 47– Descrição dos percentuais de alunos e dos níveis com base no teste de Língua Portuguesa – 4ª Série – 2003

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$\theta < 125$ 18,7%	Neste nível não houve nenhum item que pudesse representá-lo;
$125 \leq \theta < 150$ 17,1%	Os alunos que atingiram este nível de proficiência demonstraram ser capazes de: <ul style="list-style-type: none"> • Localizar informações explícitas em textos, com apoio de ilustração; • Inferir informação implícita relacionada aos elementos que constroem uma narrativa; • Identificar: <ul style="list-style-type: none"> ➢ as marcas lingüísticas que evidenciam relação de amizade entre locutor e o interlocutor de uma história em quadrinhos ➢ os elementos que constroem a narrativa (personagem principal); ➢ a finalidade de um conto; • Estabelecer relação causa/conseqüência entre partes de elementos de um texto em quadrinhos com apoio de ilustração gráfica.
$150 \leq \theta < 175$ 19,6%	Neste nível de proficiência, os alunos demonstraram ser capazes de: <ul style="list-style-type: none"> • Localizar informações explícitas: <ul style="list-style-type: none"> ➢ no início de uma narrativa infantil e de um texto informativo; ➢ em uma tirinha, com auxílio da ilustração; ➢ em um texto, com apoio de ilustração; ➢ em pequenos anúncios; • Inferir uma informação implícita a respeito do sentimento despertado em relação às ações dos personagens e o sentido de uma palavra pelo verbo que a acompanha ou pelo senso comum; • Interpretar: <ul style="list-style-type: none"> ➢ texto de tirinha com auxílio de material gráfico; ➢ o sentimento da personagem com o desfecho da história pela ilustração; ➢ o tema de um texto informativo explicitado já no título;

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<ul style="list-style-type: none"> • Identificar: <ul style="list-style-type: none"> ➢ a finalidade de um texto informativo simples; ➢ o conflito gerador do enredo de uma narrativa infantil e um dos elementos que constroem a narrativa (personagens); ➢ o efeito de sentido decorrente de uma onomatopéia (imitação de sons por meio de palavras), auxiliados por ilustração gráfica; ➢ as marcas linguísticas que evidenciam fala entre mãe e filho; • Estabelecer relações lógico-discursivas em poema, marcadas explicitamente por advérbio de tempo e relação causa/conseqüência entre imagens do início da história e elementos que compõem a imagem de um texto imagético (de imagens) e entre partes e elementos de uma poesia.
<p>175 ≤ θ < 200 19,0%</p>	<p>Alunos que atingiram este nível de proficiência demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Localizar informações explícitas: <ul style="list-style-type: none"> ➢ com auxílio de ilustrações (histórias em quadrinhos, textos ilustrados); ➢ em textos informativos, poéticos e em um poema por meio da associação de dois versos; relacionadas ao enredo e ao desfecho da história contada em um poema; ➢ no início de um convite; • Inferir: <ul style="list-style-type: none"> ➢ informações implícitas que se encontram na superfície do texto e com base em expressões simples e nos contextos dos textos; ➢ uma expressão com base no contexto do texto; ➢ o sentimento do personagem principal de um conto; • Identificar: <ul style="list-style-type: none"> ➢ elementos que solucionam o conflito gerador da história e que constroem a narrativa (característica da personagem principal de uma história e o narrador de um conto) e o desfecho de uma história com base no enredo que constrói a narrativa; ➢ a mudança de humor de um dos personagens pela ilustração de uma história em quadrinhos; ➢ as marcas linguísticas que evidenciam linguagem própria de familiares, linguagem infantil e do narrador de uma história; ➢ o sentido de reforço de uma ordem decorrente de uma exclamação no final de uma frase; ➢ a idéia principal de um poema que está relacionada ao título; ➢ a finalidade de um texto; • Comparar dois textos do mesmo gênero e identificar finalidades diferentes para cada um deles.
<p>200 ≤ θ < 225 13,2%</p>	<p>Alunos que alcançaram esse nível de proficiência demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Localizar informações explícitas: <ul style="list-style-type: none"> ➢ parafraseadas no texto; ➢ em texto ilustrado com muitas informações; ➢ em um texto lendário complexo; • Inferir: <ul style="list-style-type: none"> ➢ informações implícitas em textos fabulares e relacionadas a características do personagem principal; ➢ informações relacionadas ao desfecho da história contada e com base no contexto do texto e em uma expressão de um poema; ➢ o sentido de palavras e expressões pelo contexto do texto; • Identificar: <ul style="list-style-type: none"> ➢ um nome falado no diminutivo como uma linguagem comum utilizada entre familiares; ➢ outra forma de tratar uma informação na comparação de textos que tratam do mesmo tema; ➢ o conflito gerador do enredo e a solução para o conflito de uma fábula; ➢ a finalidade de um anúncio e de um texto informativo complexo; ➢ efeitos de humor em tirinha com apoio de ilustração gráfica e com base nas condições de um conhecido personagem das histórias em quadrinhos; ➢ o efeito de sentido das reticências no final dos últimos versos das estrofes de uma poesia; ➢ o tempo em que se passa uma história;

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<ul style="list-style-type: none"> ➤ o assunto principal de um poema, de uma crônica e de um texto narrativo descritivo simples; ➤ o sentido expresso pelos recursos utilizados pelo autor no formato diferente de um poema; ➤ um referente que contribui para a continuidade do texto; ➤ uma opinião relativa a um fato; ● Estabelecer: <ul style="list-style-type: none"> ➤ relações causa/consequência entre partes e elementos de textos com base no tema central, com apoio das ilustrações de uma história em quadrinhos e entre imagens e elementos relacionados ao desfecho da história de um texto de imagens (imagético); ➤ relação entre partes de um texto, identificando substituição que contribui para a continuidade do texto.
<p>225 ≤ θ < 250 7,6%</p>	<p>Os alunos que atingiram este nível de proficiência demonstraram ser capazes de:</p> <ul style="list-style-type: none"> ● Localizar informações explícitas em locais diversos de um texto instrucional e com apoio de ilustração; ● Inferir: <ul style="list-style-type: none"> ➤ uma informação implícita relacionada ao conflito gerador do enredo da história; ➤ o sentido de expressões relacionadas ao tema do texto e expressões comuns entre amigos; ● Interpretar texto com auxílio de material gráfico; ● Identificar: <ul style="list-style-type: none"> ➤ o tema de um texto em quadrinhos com apoio de material ilustrativo; ➤ a finalidade de um texto ilustrado do gênero irônico; ➤ o conflito gerador do enredo de uma história infantil e um dos elementos que constroem a narrativa de uma fábula (narrador); ➤ um fato da narrativa e uma opinião entre vários fatos descritos em textos diversos; ➤ o referente que substitui uma expressão e contribui para a continuidade do texto; ➤ o sentido do uso das reticências ao final de uma frase como continuidade do texto, de uma onomatopéia pouco comum com apoio de outras ilustrações em uma tirinha e de conformidade denotado por uma expressão ao final de uma tirinha; ➤ a função de um travessão em um diálogo; ➤ traços de humor ao longo de um texto poético, o motivo da ironia em uma anedota e um trecho de humor em um texto ilustrado; ➤ marcas linguísticas características de linguagem cotidiana; ● Estabelecer <ul style="list-style-type: none"> ➤ relações entre partes de um texto, identificando substituição que contribui para a continuidade de um texto; ➤ relações lógico-discursivas marcadas, explicitamente por advérbio de lugar e por um advérbio de modo no sentido comparativo; ➤ relação causa/consequência entre partes de um texto fabular.
<p>250 ≤ θ < 275 3,3%</p>	<p>Neste nível de proficiência, os alunos demonstraram ser capazes de:</p> <ul style="list-style-type: none"> ● Localizar uma informação explícita em texto informativo, envolvendo vocábulos de difícil compreensão; ● Inferir o sentido de uma expressão comum entre população jovem; ● Identificar <ul style="list-style-type: none"> ➤ elementos que constroem uma narrativa (narrador) e uma condição da personagem principal ao dar desfecho à história; ➤ a finalidade de um texto fabular e de uma anedota; ➤ o efeito de sentido do uso de reticências decorrente da interrupção e continuidade do texto em dois quadrinhos de uma tirinha e de indignação decorrente do uso de uma interrogação; ➤ termos que denotam linguagem comum utilizada entre colegas; ➤ o sentido expresso pelos recursos utilizados pelo autor no formato diferente de uma das palavras com significado complexo; ➤ efeitos de humor em texto ilustrativo e por uma condição imposta e expressa entre parênteses em um convite;

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<ul style="list-style-type: none"> ➤ marcas lingüísticas que evidenciam apresentações em situações descontraídas; • Estabelecer <ul style="list-style-type: none"> ➤ relações lógico-discursivas marcadas por advérbio de tempo e por adjunto adverbial de lugar e por advérbio que denota comparação; ➤ relação causa/conseqüência entre partes e elementos de um continho infantil; ➤ relações entre partes do texto, identificando substituição que contribui para a continuidade do texto • Distinguir um fato de uma opinião relativa esse fato em um texto fabular mais longo.
<p style="text-align: center;">$275 \leq \theta < 300$ 1,1%</p>	<p>Neste nível, os alunos demonstraram ser capazes de:</p> <ul style="list-style-type: none"> • Inferir uma informação implícita em um texto fabular com base no desfecho do enredo; • Identificar <ul style="list-style-type: none"> ➤ o conflito gerador do enredo de uma fábula; ➤ o efeito de sentido decorrente de uma exclamação em um poema e do uso de reticências dando sentido de que o texto ainda continuaria; ➤ efeitos de ironia no duplo sentido de uma palavra em uma anedota; • Estabelecer relações lógico-discursivas marcadas por um advérbio de tempo e de lugar e por conjunção indicativa de negação; • Distinguir uma opinião entre vários fatos relacionados a um texto.
<p style="text-align: center;">$\theta > 300$ 0,5%</p>	<p>Neste nível de proficiência, os alunos demonstraram ser capazes de identificar o efeito de sentido de destaque ao uso de aspas em uma expressão.</p>

Não houve publicação de resultados dos testes de 2003, embora tenha havido dois painéis para descrição das escalas, com dois produtos. O primeiro painel foi parte do contrato com a Fundação Cesgranrio, realizado no período de 14 a 16/05/2004, e o outro, promovido pelo próprio INEP, no período de 03 a 06/03/2005. Desses dois painéis foram produzidos documentos que sustentariam os relatórios de resultados dessa Edição. No entanto, até a presente data (março/2007) não houve registro de resultados.

5.5.2. Avaliação de Matemática

O teste de Matemática do SAEB 2003, como o de Língua Portuguesa, foi elaborado com base no Documento SAEB – Novas Perspectivas (2002). Todos os referenciais foram mantidos, não tendo havido nenhuma novidade no processo.

O teste foi constituído de 169 questões das quais 148 (88%) foram aproveitadas na descrição do que os alunos brasileiros sabem ou deveriam saber fazer. Em síntese, analisando-se a qualidade do instrumento, pôde-se verificar que, em relação às avaliações anteriores, e como a de língua portuguesa, o teste vem melhorando. Constata-se que os itens estão, em grande parte, contextualizados; houve menos problema com o formato dos itens e uma melhor distribuição de conteúdos.

O principal problema continua sendo o nível de dificuldade e a não adequação de muitos itens à fase de desenvolvimento cognitivo em que os alunos se encontram. Como já foi discutido anteriormente, alunos dessa fase ainda têm muita dificuldade de abstrair e há muitos itens com essa característica.

A distribuição dos itens pelos temas contemplados pela Matriz 2001 ficou da seguinte maneira: *Números e informações*, prioridade 1, 50%; *Grandezas e medidas*, prioridade 2, 28%, *Espaço e forma*, prioridade 3, 14%; e *Tratamento da informação*, prioridade 4, 8%.

Quanto ao problema da variância da métrica dos itens selecionados para descreverem a escala é um problema que permanece e que já foi chamada a atenção anteriormente.

Na Tabela 48 são apresentadas as estatísticas geradas dos itens selecionados para a descrição da escala de proficiência dos alunos em Matemática, 4ª série, 2003. Essa seleção foi feita considerando-se os critérios descritos na metodologia e o atual foco da avaliação do SAEB.

Tabela 48 – Estatísticas – Teste Matemática – 4ª Série – SAEB – Edição 2003

	Parâmetro a	Parâmetro b	Parâmetro c	Discriminação Índice D	Correlação bisserial	Proporção de acerto
N	148	148	148	148	148	148
Média	1,92	-0,90	0,17	0,49	0,53	0,46
Desvio Padrão	0,64	0,08	0,08	0,14	0,11	0,15
Mínimo	0,66	-3,26	0,01	0,09	0,26	0,08
Máximo	3,90	0,97	0,39	0,77	0,75	0,85

A Tabela 49 apresenta a descrição dos níveis de proficiência que representam o que os alunos brasileiros demonstraram saber fazer, com base no teste aplicado, e o percentual de alunos da 4ª série, que atingiu cada um desses níveis.

Tabela 49 – Descrição dos percentuais de alunos e dos níveis com base no teste de Matemática – 4ª Série – 2003

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
$\theta < 125$ 11,5%	Neste nível não houve nenhum item que o representasse.
$125 \leq \theta < 150$ 18,4%	Alunos da 4ª série que atingiram esse nível de proficiência demonstraram ser capazes de: <ul style="list-style-type: none"> ➤ calcular a área de uma figura plana com o apoio da malha quadriculada, dada a unidade de superfície.
$150 \leq \theta < 175$ 21,8%	Neste nível de proficiência, os alunos da 4ª série demonstraram ser capazes de: <ul style="list-style-type: none"> Em Espaço e forma, <ul style="list-style-type: none"> ➤ identificar um quadrado entre outras formas bidimensionais apresentadas; Em Grandezas e medidas,

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<ul style="list-style-type: none"> ➤ resolver problema, estabelecendo trocas entre cédula de pequeno valor e moedas do sistema monetário brasileiro; <p>Em Números e operações,</p> <ul style="list-style-type: none"> ➤ armar e efetuar uma adição com três parcelas com números de diferentes ordens; ➤ resolver problema envolvendo quantia de dinheiro que falta para aquisição de um objeto; <p>Em Tratamento da informação,</p> <ul style="list-style-type: none"> ➤ ler e interpretar tabela de dupla entrada, mas com poucas e claras informações.
<p>175 ≤ θ < 200 19,4%</p>	<p>Alunos de 4ª série que alcançaram este nível de proficiência demonstraram ser capazes de:</p> <p>Em Espaço e forma,</p> <ul style="list-style-type: none"> ➤ identificar: <ul style="list-style-type: none"> - a localização de objeto pela posição indicada em um referencial; - a figura plana de menor área entre outras desenhadas na malha quadriculada e as que possuem o mesmo tipo de ângulo; <p>Em Grandezas e medidas:</p> <ul style="list-style-type: none"> ➤ identificar a medida de um objeto com régua milimetrada envolvendo decimais; ➤ resolver problema, envolvendo situação simples de trocas entre cédulas do sistema monetário brasileiro, em função de seus valores; ➤ determinar tempo que falta para o início de um evento e o intervalo de tempo que durará um evento dadas as referências; <p>Em Números e operações:</p> <ul style="list-style-type: none"> ➤ identificar: <ul style="list-style-type: none"> - a localização de números naturais em reta numérica, dispostos em intervalos de 10 unidades, com vários números como referência, e em intervalos de uma unidade, com duas e três referências; - números escritos em forma numérica a partir de número escrito por extenso, da ordem das dezenas de milhar e das centenas, com zeros intermediários; ➤ compor número da ordem das centenas, com zero intermediário; ➤ calcular o resultado de uma multiplicação de um número da ordem das unidades de milhar por unidades ➤ resolver problema envolvendo: <ul style="list-style-type: none"> - adição de valores, indicando o valor total de uma compra e com o sentido de alteração de um estado inicial (positiva); - o valor de um objeto indicado pela soma das cédulas e moedas dadas como pagamento; - números racionais expressos na forma decimal envolvendo adição; <p>Em Tratamento da informação:</p> <ul style="list-style-type: none"> ➤ Ler e interpretar dados representados em tabelas de dupla entrada com poucas informações e identificando em uma coluna, o maior valor monetário, escrito na forma decimal, indicado entre outros apresentados.
<p>200 ≤ θ < 225 13,9%</p>	<p>Neste nível de proficiência, os alunos de 4ª série demonstraram ser capazes de:</p> <p>Em Espaço e forma,</p> <ul style="list-style-type: none"> ➤ identificar: <ul style="list-style-type: none"> - uma das propriedades comuns dos quadriláteros (número de lados); - a modificação do perímetro em função da redução pela metade e das medidas dos lados pela ampliação, de figuras poligonais desenhadas na malha quadriculada; - a localização de objeto dada a posição, os referenciais e a orientação em um desenho; <p>Em Grandezas e medidas:</p> <ul style="list-style-type: none"> ➤ estabelecer intervalo de duração de tempo, dados horário de início e término das atividades; ➤ resolver problema, estabelecendo conversão entre número de dias e semanas, de horas em dias e envolvendo trocas entre cédulas e moedas do sistema monetário brasileiro; <p>Em Números e operações:</p> <ul style="list-style-type: none"> ➤ identificar: <ul style="list-style-type: none"> - a localização de números naturais na reta numérica e a movimentação sobre ela, seguindo

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<p>ordens;</p> <ul style="list-style-type: none"> - a composição e a decomposição de números naturais na forma polinomial, da ordem das centenas, com um zero intermediário; - o maior número que se poderia formar, dados três algarismos e o menor número formado por quatro algarismos, pelo princípio do valor posicional; <p>➤ calcular o resultado de uma adição, pelo reconhecimento de que a diferença de uma subtração somada ao subtraendo vai resultar no minuendo, no qual, um dos números aparece como incógnita, e de uma subtração de um número da ordem das unidades de milhar, com dois zeros intermediários, de um outro da ordem das dezenas;</p> <p>➤ resolver problemas envolvendo:</p> <ul style="list-style-type: none"> - subtração: abordando a propriedade da invariância da diferença; no sentido de alteração positiva de um estado inicial; e de números racionais, com duas casas decimais; - multiplicação com sentido de proporcionalidade; - divisão exata e com resto de números naturais; - situação de compra e pagamento com troco, utilizando a escrita decimal de cédulas e moedas; <p>Em Tratamento da informação:</p> <p>➤ ler e interpretar dados representados em tabelas e gráficos de colunas, dos quais é solicitado calcular o resultado da soma de colunas;</p>
<p>225 ≤ θ < 250 8,7%</p>	<p>Alunos da 4ª série que atingiram esse nível de proficiência demonstraram ser capazes de:</p> <p>Em Espaço e forma,</p> <ul style="list-style-type: none"> ➤ identificar, figuras tridimensionais com suas planificações com duas faces circulares; <p>Em grandezas e medidas,</p> <ul style="list-style-type: none"> ➤ estabelecer relações entre: o trimestre e o ano; anos e meses; o horário de início, término e o intervalo de duração de um evento; ➤ estimar comprimento de um objeto utilizando medida não convencional e distâncias percorridas tendo por base medidas referenciais não convencionais; ➤ resolver problema tendo por base trocas entre cédulas e moedas de valores mais altos e envolvendo situação de compra, com facilitação de troco e de multiplicação de unidade de medida padronizada (litro); ➤ calcular: <ul style="list-style-type: none"> - o perímetro de uma figura plana desenhada em uma malha quadriculadas, dadas as suas medidas dos lados; - a área de figura plana irregular, desenhada em malha quadriculada, envolvendo meia medida da unidade de superfície; <p>Em números e operações:</p> <ul style="list-style-type: none"> ➤ identificar: <ul style="list-style-type: none"> - e utilizar agrupamento e trocas na base 10 com base no princípio do valor posicional, com ordens representadas por figuras; - a localização de números naturais na reta numérica, com intervalos de 20 unidades, com apenas duas referências, do início e final; - a composição de números de quatro e cinco algarismos que, com base no princípio do valor posicional, formarão o maior e o menor número possível; - a composição e a decomposição na forma polinomial de números da ordem das unidades de milhar com zeros intermediários; - a posição relativa de um algarismo em um número, pelo princípio do valor posicional; - uma fração como representação de parte-todo, com apoio de ilustração; - cédulas do sistema monetário que compõem um valor que está entre dois limites, por meio da análise das alternativas de um problema; ➤ calcular o resultado: <ul style="list-style-type: none"> - de uma subtração de um número da ordem das dezenas de milhar, com quatro zeros no final, de um número da ordem das centenas; - de uma divisão exata de um número da ordem das unidades de milhar por dezena; - de uma multiplicação de números por 10, com apoio de outras multiplicações com resultados e com base nos resultados, substituindo figuras que representam algarismos; ➤ resolver problemas envolvendo:

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<ul style="list-style-type: none"> - adição e subtração de números naturais no sentido de alteração de um estado inicial (positiva e negativa) e de distâncias em metros; - subtração de número da ordem das centenas, com zero intercalado de outro da unidade das dezenas e de racionais escritos na forma decimal, com duas casas decimais; - multiplicação de números da ordem das dezenas; - situação de compra, na qual utiliza cédulas do Sistema Monetário Nacional para efetuar o pagamento e realiza trocas de cédulas, com facilitação de troco; - porcentagem (25%); <p>Em tratamento da informação:</p> <ul style="list-style-type: none"> ➤ ler e interpretar: dados representados em tabela de dupla entrada, com cálculo do número total de respondentes de uma das colunas e dados representados em gráficos de colunas, dos quais tem que fazer cálculos para chegar à solução.
<p style="text-align: center;">$250 \leq \theta < 275$ 4,1%</p>	<p>Alunos de 4ª série que alcançaram esse nível de proficiência demonstraram ser capazes de</p> <p>Em espaço e forma:</p> <ul style="list-style-type: none"> ➤ identificar: <ul style="list-style-type: none"> - a localização de objeto em croqui, no qual o aluno o encontra de frente a ele; - figuras planas desenhadas em malha quadriculada que possuem a mesma área e bidimensionais desenhadas no plano, pelo nome (losango, triângulo, hexágono, pentágono) e uma figura com formato de uma esfera; <p>Em grandezas e medidas:</p> <ul style="list-style-type: none"> ➤ identificar pelo cálculo do perímetro, a ampliação proporcional de uma figura plana, com o apoio da malha quadriculada; ➤ estabelecer: relações entre o horário de início e término e intervalo da duração de um evento; ➤ estimar a medida de superfície de uma área e a medida do comprimento de um objeto utilizando unidades de medida não convencionais; ➤ resolver problema envolvendo: conversão de horas e minutos em minutos e anos e alguns dias em dias; transformação de unidades de medida padronizadas de mesma grandeza (ml/l); <p>Em números e operações:</p> <ul style="list-style-type: none"> ➤ identificar: <ul style="list-style-type: none"> - a localização de números em sequência ordenada e de números racionais representados na forma decimal, com duas casas decimais, na reta numérica; - a decomposição de um número da ordem das unidades de milhar com dois zeros intermediários; ➤ calcular o resultado: <ul style="list-style-type: none"> - de uma expressão numérica complexa que envolve operações de adição e subtração; - de uma subtração de números naturais, envolvendo números da ordem das unidades de milhar, com recursos, e com resultado parcial; ➤ resolver problemas envolvendo: <ul style="list-style-type: none"> - divisão de número da ordem das centenas por 10, com resto; - subtração: de um número inteiro de outro racional, expresso na forma decimal, com uma casa decimal; no sentido de decompor, ou retirar uma quantidade que numa ação comparativa alcançará o valor inicial; e utilizando a propriedade da invariância da diferença; - estimativa de número de objetos que podem ser adquiridos, considerando-se um valor determinado para cada objeto e utilizando a escrita decimal de cédulas e moedas do sistema monetário brasileiro; - porcentagem (50%); - multiplicação em situação associada à configuração retangular; - compra, devolução e facilitação de troco, utilizando cédulas e moedas do sistema monetário brasileiro; ➤ -Correspondem uma representação fracionária a uma decimal, de ordem milésimal.
<p style="text-align: center;">$275 \leq \theta < 300$ 1,6%</p>	<p>Neste nível de proficiência, os alunos de 4ª série demonstraram ser capazes de:</p> <p>Em Espaço e forma,</p> <ul style="list-style-type: none"> ➤ Identificar: <ul style="list-style-type: none"> - quadrilátero pela posição relativa entre seus lados (dois lados paralelos e dois não paralelos);

NÍVEIS	DESCRIÇÃO DOS NÍVEIS
	<p>- uma das características do cilindro de possuir duas faces planas; figura de um cubo planificado aberto;</p> <p>Em Grandezas e medidas,</p> <ul style="list-style-type: none"> ➤ estimar a medida de comprimento usando medida não convencional, envolvendo números decimais; ➤ estabelecer relações entre uma data inicial e final e o intervalo de validade de um produto e entre o horário de início e término e intervalo de duração de um programa; ➤ resolver problemas envolvendo conversão de medida de grandeza de comprimento (<i>km em m</i>) e de capacidade, utilizando unidades de medida convencionais (<i>l/ml</i>) e não convencionais; <p>Em Números e operações,</p> <ul style="list-style-type: none"> ➤ identificar: <ul style="list-style-type: none"> - a localização de números racionais representados na forma decimal na reta numérica, com apenas dois números referenciais inteiros e com apoio de referenciais representados na malha quadriculada; - fração que representa parte de uma soma (parte-todo) em um problema; ➤ resolver problema envolvendo: <ul style="list-style-type: none"> - subtração de número inteiro de um racional, com uma casa decimal; - multiplicação: com sentido de análise combinatória; de números decimais, de medidas convencionais (km); e com divisão, dando idéia de proporcionalidade; - situações de compra com trocas entre cédulas e moedas, facilitação de troco e tendo como incógnita, o valor da mercadoria adquirida; ➤ reconhecer o conceito de porcentagem pela solução de um problema; <p>Tratamento da informação,</p> <ul style="list-style-type: none"> ➤ ler e interpretar gráfico de colunas e inferir uma informação implícita dos dados apresentados.
<p>300 ≤ θ < 325 0,6%</p>	<p>Neste nível de proficiência, os alunos de 4ª série demonstraram ser capazes de:</p> <p>Em espaço e forma,</p> <ul style="list-style-type: none"> ➤ identificar: <ul style="list-style-type: none"> - a localização de pessoas em representação gráfica indicada pela posição de uma em relação à outra; - números em ordem crescente/decrescente; direita/esquerda de um referencial e pares/ímpares com base em um mapa; - paralelogramos, dada a sua principal propriedade (quatro lados paralelos dois a dois) <p>Em Grandezas e medidas,</p> <ul style="list-style-type: none"> ➤ calcular o perímetro de uma figura plana irregular desenhada em malha quadriculada e o perímetro de figura retangular plana, dadas as medidas de dois lados e desconto de medida referente a porta, com apoio de ilustração; <p>Em Números e operações,</p> <ul style="list-style-type: none"> ➤ identificar: <ul style="list-style-type: none"> - a localização de números naturais na reta numérica, na qual os referenciais são os números representados pelo início e final da reta, sem muita lógica; - a composição de números naturais pela indicação de número grande centens somado a número pequeno de unidades; - fração como representação associada a figura ilustrada representando parte-todo; ➤ resolver problema envolvendo porcentagem (100% de um determinado valor).
<p>$\theta > 325$ 0,1%</p>	<p>Em Números e operações,</p> <ul style="list-style-type: none"> ➤ comparar e ordenar seqüência de números racionais na forma decimal, envolvendo décimo, centésimo e milésimo; ➤ reconhecer número fracionário escrito na forma decimal (ordem centesimal).

Em relação à descrição das escalas do SAEB 2003, como comentado anteriormente, não se pode tecer nenhum comentário, posto que nenhuma publicação foi feita.

Quanto aos testes, cabem as mesmas observações em relação aos cuidados que se deve ter com a escolha dos itens que os compõem. A exclusão de itens dos testes relativos a essa edição, ocorreu principalmente e recorrentemente por problemas com a inadequação ao nível avaliado e ao formato dos itens (falta de clareza no enunciado e falta de plausibilidade das alternativas). Esses problemas poderiam ser minimizados se houvesse comitês compostos por representantes das diversas categorias ligadas e comprometidas com a educação (especialistas em avaliação, em psicometria, nas áreas avaliadas, professores de sala de aula, pais, elaboradores de políticas públicas, representantes da secretaria de educação básica, entre outros) que se responsabilizassem pelo acompanhamento do processo de escolha dos itens e montagem dos testes. Isso dá transparência e confiabilidade ao processo.

Neste capítulo pretendeu-se chamar a atenção para os instrumentos de avaliação de desempenho dos alunos e para a descrição do que os alunos demonstram ser capazes de fazer com base nesses instrumentos.

As análises realizadas demonstraram que a cada edição do SAEB houve um crescente aproveitamento de itens nos testes e que a qualidade pedagógica e psicométrica melhoraram, decaindo um pouco em relação ao aproveitamento de itens, mas crescendo em qualidade, em 2003. As Tabelas 50, 51 e 52 mostram esse processo.

Tabela 50– Aproveitamento de itens por teste/edição

Edição	Aproveitamento de itens	
	Língua Portuguesa	Matemática
1995	68/130 (52%)	76/130 (58%)
1997	53/143 (37%)	85/143 (59%)
1999	105/169 (62%)	117/169 (69%)
2001	157/169 (93%)	155/169 (92%)
2003	148/169 (88%)	148/169 (88%)

Tabela 51 – Síntese dos principais problemas apontados por edição pela análise pedagógica

Edição	Análise Pedagógica dos Testes
	Problemas mais freqüentes apontados
1995	<ul style="list-style-type: none"> • Quantidade excessiva de textos inadequados, tanto em relação ao conteúdo quanto à forma (frases e fragmentos curtos, fora de contexto e sem referências); • Falta de plausibilidade das alternativas; • Linguagem inadequada dos enunciados e das alternativas; • Conteúdos de 8ª série contemplados na 4ª; • Questões descontextualizadas; • Distribuição ruim de conteúdos;
1997	<ul style="list-style-type: none"> • Textos inadequados ao nível de escolarização avaliado, com problemas de formato, referências etc.; • Conteúdos repetidos, com o mesmo nível de dificuldade (questões quase idênticas); • Distribuição dos conteúdos comprometida (muitos itens de geometria); • Linguagem inadequada; • Questões descontextualizadas; • Em matemática, poucas situações de aplicação;
1999	<ul style="list-style-type: none"> • Textos com temas inadequados (poemas abstratos, crônicas, charges); • Formato dos itens inadequado (associação de duas colunas, completar frases); • Plausibilidade das alternativas; • Linguagem dos enunciados; • Distribuição de conteúdos; • Questões descontextualizadas; • Falta de correspondência entre itens e descritores;
2001	<ul style="list-style-type: none"> • Textos inadequados à fase de escolarização; • Plausibilidade das alternativas; • Peguinhas; • Linguagem dos enunciados; • Itens difíceis de matemática no mesmo bloco;
2003	<ul style="list-style-type: none"> • Textos com temas inadequados; • Formato dos itens (falta de paralelismo sintático, alternativas excludentes); • Falta de objetividade dos enunciados; • Falta de plausibilidade das alternativas; • Nível alto de dificuldade dos itens de matemática; • Não adequação dos itens à fase de desenvolvimento cognitivo que os alunos se encontram (conteúdos abstratos).

Tabela 52 – Síntese dos parâmetros obtidos das análises da qualidade psicométrica dos testes do SAEB

Edição	Língua Portuguesa				Matemática			
	Discriminação		Dificuldade		Discriminação		Dificuldade	
	Param <i>a</i>	Bisserial	Param <i>b</i>	Prop. Acer	Param <i>a</i>	Bisserial	Param <i>b</i>	Prop. Acer.
1995	1,39 (0,49)	–	-1,65 (0,95)	0,61 (0,17)	1,67 (0,59)	–	-0,69 (0,94)	0,44 (0,16)
1997	1,46 (0,48)	0,55 (0,08)	-1,25 (0,84)	0,53 (0,14)	1,63 (0,54)	0,49 (0,09)	-0,66 (0,94)	0,46 (0,17)
1999	1,08 (0,36)	0,52 (0,08)	-1,07 (0,69)	0,46 (0,14)	1,18 (0,43)	0,49 (0,09)	-0,69 (0,69)	0,43 (0,14)
2001	1,61 (0,47)	0,57 (0,09)	-1,38 (0,75)	0,49 (0,14)	1,81 (0,49)	0,55 (0,10)	-0,85 (0,76)	0,44 (0,14)
2003	1,84 (0,60)	0,56 (0,10)	-1,38 (0,78)	0,52 (0,16)	1,92 (0,64)	0,53 (0,11)	-0,90 (0,79)	0,46 (0,15)

6. Entre parênteses, o desvio padrão.

Pela Tabela 50, constata-se que em aproveitamento de itens dos testes, a edição de 2001 foi a melhor em relação às outras. Em relação à qualidade pedagógica dos itens, observa-se pela tabela 51, uma recorrência de alguns problemas. Entre esses, os relacionados à escolha de textos, ao formato dos itens, à linguagem e objetividade dos enunciados, à contextualização das questões, à plausibilidade das alternativas e à dificuldade dos itens de matemática.

Em termos de qualidade psicométrica, verifica-se na tabela 52, que a avaliação de 2003 teve os melhores parâmetros. Podem-se observar também os níveis altos e constantes de dificuldade dos testes de matemática ao longo de todas as edições. Esse é um problema que deve ser revisto e sanado com um melhor planejamento para a montagem dos testes.

Pôde se verificar que de 1995 a 2003 houve um amadurecimento visível e natural do processo. Pelos resultados das análises dos testes de 4ª série, das duas disciplinas escolhidas para esse fim, constata-se que os testes hoje são mais robustos e consistentes com a proposta de avaliação em larga escala. Em 1995, 1997 e 1999 houve uma divagação entre a avaliação de sala de aula e a avaliação em larga escala. Mas foi um processo de aprimoramento. Percebe-se que a partir de 2001 houve uma maior conscientização do papel da avaliação na melhoria da educação e um maior empenho para que os resultados saíssem das gavetas, e no meio de inúmeros problemas encontrados, pode-se observar que a própria sociedade hoje cobra alguns resultados.

Em relação às escalas, é importante que elas reflitam o resultado do que os alunos realizaram, de forma clara e contundente, sem deixar margem a fantasias. Ela deve, em

primeiro lugar, demonstrar por meio das habilidades aferidas que o conhecimento vai sendo construído ao longo dos períodos de escolarização. Se o professor tem essa percepção, ele pode compreender que se houver maior empenho, ou uma mudança em sua prática pedagógica, o aluno poderá alcançar níveis mais altos de habilidades.

Outro problema detectado diz respeito à falta de estabilidade na determinação dos níveis das escalas. A cada avaliação foram arbitrados intervalos diferentes. Se o professor quiser tomar a escala como referência de comparação do desenvolvimento, de uma avaliação para outra, ele ficará desorientado, pois não poderá constatar, por exemplo, se aumentou ou diminuiu o percentual de alunos em cada nível.

O trabalho realizado de (re)descrever as escalas teve o intuito de mostrar que há a possibilidade de manter um diálogo com educadores da base. Todo o trabalho pode ser mais bem apresentado, em termos de *layout*. O fundamental é que os resultados sejam descritos de forma que o professor possa entendê-los e possa se identificar no processo. Além da descrição por série, considera-se também importante mostrar a descrição por intervalos menores, pois a estatística pode causar um impacto grande quando se reúnem os percentuais de alunos que alcançam os níveis de proficiência.

Fazer os resultados chegarem à base deve ser um papel importante da avaliação, considerando-se que entre os diversos atores envolvidos com a educação escolar, aquele que tem maior capacidade de promover mudanças é o professor que se encontra na ponta do processo. Além disso, o principal papel do SAEB tem sido o de servir de modelo para avaliações nos estados e municípios. Ao tirar o SAEB da esfera dos gabinetes, a possibilidade de alcançar melhorias na qualidade da educação básica brasileira aumentará.

7. ANÁLISE CRÍTICA DOS RELATOS DE RESULTADOS DO SAEB - 1995 – 2003

Neste capítulo são analisadas e discutidas, sob um ponto de vista crítico, as publicações relacionadas aos resultados das edições do SAEB, no período de 1995 a 2003.

Idéia do presidente do INEP, Reynaldo Fernandes, apresentada à 1ª Reunião Anual da Associação Brasileira de Avaliação Educacional (Abave), realizada nos dias 11 e 12 de maio de 2006, em Belo Horizonte, publicada no Boletim Informativo do INEP no dia 16 de maio de 2006, resume o que uma avaliação que tenha por base os resultados de um processo avaliativo como o do SAEB deve apresentar. Segundo ele,

“Avaliações em larga escala dos sistemas educacionais, devem apresentar resultados postos à sociedade como direito à informação e como instrumento democrático de controle social da qualidade na educação. As diferentes perspectivas e a experiência compartilhada ajudam a delinear sistemas de avaliação mais eficazes, não só no que diz respeito à concepção e operacionalização, mas também em relação ao desafio, por exemplo, de fazer com que os gestores locais e professores se apropriem dos resultados e desenvolvam ações concretas a partir deles”. ()

() *“A avaliação em larga escala, como diagnóstico do sistema educacional, deve subsidiar não apenas as políticas públicas, mas também os gestores educacionais e a própria sociedade; países que dão visibilidade aos resultados de avaliação tiveram melhorias no desempenho dos alunos”, conclui.*

A partir de 1995, com a adoção da nova metodologia de avaliação, o SAEB produz relatórios e patrocina publicações dos resultados com base, principalmente, nas escalas de proficiência e de desempenho geradas das análises da TRI e nas estatísticas dos dados coletados dos questionários de contexto associados ao desempenho dos alunos. Nesse tópico serão apresentadas, de forma resumida, as publicações dos resultados das edições supracitadas com análise dos seguintes pontos: (1) público alvo dos relatórios; (2) conteúdo dos relatórios; (3) nível de compreensão dos dados relatados; (4) adequação da publicação para agentes envolvidos com questões educacionais; (5) alcance das publicações dos resultados.

6.1. EDIÇÃO 1995

O relatório que é objeto de análise dessa edição é o “Resumo Executivo: SAEB/95. (1997). Brasília: INEP.

Trata-se de uma pequena publicação de onze páginas, em que, de forma sintética, são apresentados os resultados do SAEB. Enquanto resumo, foi uma publicação bem estruturada, com linguagem simples, que alcançaria diversos públicos da área educacional.

Na apresentação ou em qualquer outra parte da publicação, não há indicação de um destinatário específico. Mas pelo conteúdo, percebe-se que tem a intenção de atingir educadores.

Em relação ao conteúdo abordado, o relatório apresentou o seguinte formato:

- Introdução: neste tópico, foi abordado, de forma direta, o processo de elaboração das escalas de proficiência. Não é apresentado um diálogo preliminar contextualizando as escalas, ou introduzindo o papel das escalas no processo avaliativo, como um dos componentes dos resultados da avaliação. Considera-se a forma de expor o processo de elaboração das escalas muito boa. Foram feitos esquemas que tornaram a compreensão mais fácil;
- As habilidades e os conteúdos examinados: este foi o tópico seguinte apresentado. Nele foram expostos, em Língua Portuguesa, os tipos de textos contemplados nos testes e a conceituação das três categorias de habilidades consideradas na formulação dos itens. Em Matemática, foram apresentadas as cinco áreas de conteúdo e as três categorias de habilidades relacionadas. As categorias de habilidades abordadas nas

duas disciplinas avaliadas foram apresentadas de forma muito reduzida, mas compreensível.

- Em seguida a esse tópico, foram apresentadas as escalas com os níveis de proficiência nas duas disciplinas. As escalas foram apresentadas sobre uma métrica de quatro pontos (150, 225, 300 e 375), acima dos quais foram colocados quadros coloridos, em que foram apresentadas, de forma muito reduzida, o que os alunos brasileiros, na educação básica (1ª série do EF à 3ª série do EM), demonstraram ser capazes de fazer. Em baixo da métrica foram apresentadas as médias de proficiência para todas as regiões e o Brasil.
- O terceiro tópico foi chamado “Principais resultados do SAEB”. Nele são apresentados gráficos do tipo “*boxplot*” (gráfico tipo caixa, apropriado para apresentação de dados contínuos) com a distribuição das médias de proficiência do Brasil e Regiões, com cores representando as séries, nas disciplinas Língua Portuguesa e Matemática. São também apresentados gráficos de linha mostrando a proficiência média dos alunos na 3ª série EM, por região e Brasil, segundo a idade dos alunos.
- O quarto tópico, chamado “Arcabouço do SAEB”, traz uma tabela mostrando os enfoques, as dimensões abordadas e os indicadores produzidos pela avaliação. Foi um resumo bem feito desses dados.
- O quinto tópico denominado “Características do SAEB” mostra qual a abrangência do SAEB e quais informações são disponibilizadas. Quanto à abrangência, são mostrados quem são os participantes, as séries abrangidas e as disciplinas avaliadas. Quanto às informações, são citados como disponíveis o desempenho dos alunos, as matrizes curriculares e os fatores associados.
- No sexto e último tópico, chamado “Avaliação da educação básica”, são apresentadas as ações relacionadas à avaliação da educação básica, que são o SAEB, o banco de itens e o ENEM associadas ao foco da avaliação e ao uso do resultado.

O formato dessa publicação parece ter sido retirado de transparências de uma apresentação sobre o SAEB. São apresentados tópicos, muito bem dispostos, que mereceriam um aprofundamento maior, pois ao lê-lo não se tem o apoio de um palestrista. Considera-se esse tipo de publicação como ideal para ser distribuída em conferências, congressos, oficinas, reuniões, as quais tenham como foco Secretarias, escolas e educadores. Em se tratando de

relatório de resultados, embora este tenha o caráter de um resumo executivo, é necessário que mesmo sendo um resumo, apresente os dados de forma que possam ser compreendidos, aprofundados, discutidos e que sirvam de subsídios para melhorias no processo educacional.

7.2. EDIÇÃO 1997

O relatório que é objeto de análise é o “SAEB 97: Primeiros resultados. Maria Inês Gomes de Sá Pestana *et al.* – Brasília: INEP (1999)”.

Trata-se de uma publicação de 140 páginas, em que são apresentados os resultados do SAEB, edição 1997, em relação à proficiência atingida dos alunos, além de sua relação com algumas variáveis de contexto relacionadas aos alunos e aos professores.

Não há um destinatário específico. Mas o conteúdo é estritamente voltado para questões educacionais, tendo, certamente, a intenção de atingir educadores.

Em relação ao formato da publicação, ela foi distribuída em cinco tópicos e três anexos. Os conteúdos relacionados a cada um desses tópicos são apresentados a seguir.

- Visão geral do SAEB: é um tópico introdutório, onde o leitor é situado sobre o que é a avaliação do SAEB. São apresentados dados relativos ao delineamento da avaliação;
- O SAEB/1997 em números: neste tópico são apresentados os números de participantes dessa edição, em relação ao número de alunos, turmas, professores e diretores das três séries avaliadas. Além disso, são fornecidos dados sobre parcela de participantes relacionada às escolas particulares e o número de responsáveis pela aplicação nos estados em relação às diversas categorias participantes;
- As escalas de proficiência em Português, Matemática e Ciências: neste tópico são fornecidas explicações sobre o processo de construção das escalas. Em figuras à parte são apresentadas as escalas únicas com as descrições do que os alunos demonstraram ser capazes de fazer, por disciplina. As descrições são apresentadas em quadros associados a intervalos, os quais aparecem acima de uma métrica. Em Matemática, os intervalos foram arbitrados nos pontos 175, 250, 325 e 400; em Português e Ciências, os intervalos foram arbitrados em 100, 175, 250 e 325.

- Análise do desempenho dos alunos, segundo os níveis de proficiência: este é um subtópico do tópico anterior. Neste subtópico são feitos comentários sobre o desempenho dos alunos por região e no Brasil. São apresentados gráficos mostrando o percentual de alunos que se encontra acima dos níveis 175, 250 e 325 nas disciplinas avaliadas; em seguida são feitas discussões comparativas dos dados relacionados ao alcance de cada nível de proficiência.
- Principais resultados do SAEB/97: esse é o quarto tópico que tem agregado a ele outros nove subtópicos. Nele são apresentadas as médias de proficiência para o Brasil, Regiões e Unidades da Federação; os resultados pelas redes de ensino (municipal, estadual e particular), localização (capital e interior), gênero (masculino e feminino), idade, escolaridade dos pais e rede de ensino e características selecionadas dos alunos e professores.
- Neste último subtópico (características selecionadas dos alunos e professores) são apresentadas por série, uma disciplina (4ª série – Matemática; 8ª série – Ciências e 3ª série EM – Português), da qual foram consideradas as proficiências médias por categoria nas seguintes variáveis relacionadas ao aluno: idade; cor; nível de instrução do pai; nível de instrução da mãe; abandono escolar em algum momento da trajetória escolar; repetência; conversa em casa sobre o que ocorre na escola; sabe usar computador; possui computador em casa; usa o computador da escola; motivos pelos quais frequenta escola. Das variáveis relacionadas ao professor, foram selecionadas: nível de escolaridade completo; gênero; participação em cursos de capacitação, treinamento ou atualização na área específica de atuação no ano 1997; carga horária na escola; salário bruto do professor na escola; oferta pela escola de programa de recuperação paralela; disponibilização de material pedagógico adequado e em quantidade suficiente; estimativa percentual de alunos aprovados.
- Considerações Finais: neste último subtópico são feitos comentários sobre os níveis de desigualdade relacionados ao desempenho dos alunos por Região e Unidades da Federação e a efetividade do currículo. São levantadas algumas hipóteses explicativas dos fenômenos constatados, a maioria associada à efetividade do currículo. É pontuado que o que emerge de todas as ponderações é *o papel fundamental da prática pedagógica das escolas e, principalmente, dos professores, que são o elo de ligação entre o currículo proposto e os resultados alcançados pelos alunos.*

- São destinadas aos anexos, 89 páginas. O Anexo I, desse relatório, traz inúmeras tabelas juntando os dados relacionados às proficiências médias por disciplina e por Regiões e Unidades da Federação. No Anexo II, estão as proficiências médias por variável considerada em todas as disciplinas e séries apresentadas em tabelas e nos gráficos correspondentes. No Anexo III, são apresentados exemplos de itens por nível das escalas de Português, Matemática e Ciências.

Esse relatório exprime uma linguagem clara e acessível a diversos tipos de públicos. São apresentados dados que certamente despertam interesse, mas de forma bastante poluída. Teria valido mais a pena comentar os dados que realmente indicassem ser merecedores de políticas de superação. Quanto ao conteúdo das escalas, embora tenha sido um assunto discutido em outro capítulo, merece chamar a atenção novamente, para o excesso de redução das descrições. Passa-se a impressão, principalmente com os itens dados como exemplo, que os testes não precisariam ter mais do que aquelas questões apresentadas. Os anexos, com exceção do III, seriam dispensáveis, pois apresentam dados repetidos do corpo principal do relatório, além de apresentações redundantes (tabelas e gráficos correspondentes).

Segundo informações de funcionários do SAEB deste período, os relatórios foram enviados a todas as secretarias de estado de educação. Se as informações fossem mais objetivas, o tamanho seria reduzido e talvez pudesse ter um maior número de exemplares para serem distribuídos.

8. EDIÇÃO 1999

Nesta edição não houve publicação impressa. Os relatórios de resultados de 1999 encontram-se subdivididos em vários arquivos disponibilizados no *site* do INEP: http://www.inep.gov.br/basica/saeb/saeb_99.htm. São seis documentos, cada um com um propósito diferente, que estão listados a seguir.

- Doc. 1 – Escalas do SAEB: trata-se de um documento com três páginas que contém explicações técnicas do processo de construção das escalas;
- Doc. 2 – Resultados dos níveis de desempenho – Língua Portuguesa e Matemática – SAEB 99 – Relatório 1: documento de 86 páginas, denominado “*Informe de resultados comparativos do SAEB 1995, 1997 e 1999*”. O documento aborda três

tópicos: (1) o contexto educacional brasileiro: 1995, 1997 e 1999; (2) os resultados comparados do SAEB:1995, 1997 e 1999; (3) Considerações finais.

O tópico 1 aborda principalmente dados relacionados ao censo, abrangendo o período em questão: expansão da matrícula; municipalização do ensino; transformações da rede instalada; mudanças pedagógicas em relação aos ciclos, à seriação e aos programas de aceleração da aprendizagem; e qualificação dos professores.

No tópico 2, sobre os resultados comparados do SAEB, foram abordados os seguintes dados: as escalas utilizadas e suas interpretações (embora tenha esse título, foi dada apenas uma explicação sobre como as escalas são apresentadas); a evolução, com análises comparadas de desempenho por série e disciplina; as características dos alunos nas séries avaliadas (crescimento da matrícula, idade e nível de escolarização dos pais); e por fim, a influência da idade e do nível socioeconômico no desempenho dos alunos, onde é abordado como o atraso e a escolaridade dos pais afetam os resultados do SAEB.

No tópico 3 são feitas considerações, chamando-se a atenção para os principais problemas detectados. Em seguida há um anexo em que aparecem as descrições do níveis das escalas com exemplos de itens após cada descrição.

- Doc 3 – Resultados dos níveis de desempenho – Ciências, Física, Química, História e Geografia – SAEB 99 – Relatório 2: documento de 33 páginas, em que são apresentados os mesmos dados do relatório anterior, voltados para as disciplinas supracitadas.
- Doc. 4 – Tabelas de níveis de desempenho: documento apresentado em sete planilhas Excel, com dados da evolução das médias de desempenho por Unidade da Federação, Região, Brasil, nas disciplinas Língua Portuguesa e Matemática, por dependência administrativa, por faixa etária dos alunos por série avaliada e por nível de escolaridade dos pais.
- Doc. 5 – Dados de Contexto: Brasil, Regiões e Unidades da Federação: dados apresentados em arquivos Excel contendo, para Brasil, Regiões e para cada Unidade da Federação, mais de 350 tabelas com informações sobre evolução de matrículas, matrícula por dependência administrativa, fluxo escolar, taxa de distorção idade/série, número de escolas e matrículas por ciclo e por classe de aceleração,

educação de jovens e adultos e funções docentes por nível de ensino, por dependência administrativa e com formação superior.

- Doc. 6 – Sumário executivo do SAEB: trata-se de um documento de 500 páginas, que se encontra subdividido em seis partes. A primeira parte, composta de 10 páginas, contém uma apresentação do documento, uma introdução na qual é feita uma contextualização da avaliação abordando a frequência dos participantes dos SAEBs 1995, 1997 e 1999 (alunos, professores, diretores e escolas), os construtos investigados, o plano amostral e instruções para leitura das tabelas de médias de desempenho.
- A segunda parte, composta de 23 páginas, apresenta as médias de desempenho por Unidade da Federação, Região e Brasil, por disciplina e série avaliada.
- A terceira parte, composta de 159 páginas, apresenta dados dos construtos relacionados aos alunos por série, Unidade da Federação, Região e Brasil.
- A quarta parte, composta por 197 páginas, abordou os construtos relacionados ao professor por Unidade da Federação, Região e Brasil.
- A quinta parte, com 81 páginas, trouxe dados dos construtos relacionados ao diretor, por Unidade da Federação, Região e Brasil.
- A sexta e última parte, com 29 páginas, apresentou os dados dos construtos relacionados às escolas.

Esses dados disponibilizados pela Internet apresentaram uma dimensão completa da situação e evolução da educação básica no Brasil no período 1995 a 1999.

A edição do SAEB 1999 apresentou uma produção bastante profícua em termos de informações estatísticas a respeito do sistema de educação básica brasileira. Os documentos disponibilizados pela Internet, se bem trabalhados e divulgados poderiam dar origem a programas, políticas e planos de superação de diversos problemas apontados. No entanto, não foi impresso nenhum documento e nenhuma política de divulgação desses dados foi feita.

Embora tenham seis documentos registrados, nenhum deles tratou apenas dos resultados da avaliação do SAEB 1999. Todos os documentos apresentam dados comparativos ou percentuais, linguagem e formato adequados para serem analisados por especialistas em dados censitários, estatísticos ou psicometristas.

Enfim, constata-se um paradoxo nesta edição do SAEB, pois se produziram muitos dados e nenhuma informação atingiu um público específico. Portanto, pode-se inferir que esta avaliação tenha sido quase inócua, não tendo suscitado ações que trouxessem algum tipo de melhoria à educação.

6.4. EDIÇÃO 2001

Os documentos que foram objeto de análise desta edição do SAEB foram os seguintes: “SAEB 2001 – Relatório Nacional. Brasília: INEP (2002); SAEB 2001 – Relatório – Língua Portuguesa. Brasília: INEP (2002); SAEB 2001 – Relatório – Matemática. Brasília: INEP (2002); Qualidade da Educação: uma nova leitura do desempenho dos estudantes da 4ª série do Ensino Fundamental. Brasília: INEP (abril/2003).

O primeiro documento a ser discutido é o SAEB 2001 – Relatório Nacional. Trata-se de um relatório com 303 páginas, com 148 páginas de corpo principal e 155 páginas de anexos. Ele está dividido em cinco capítulos que abordam os seguintes temas: (1) Contexto educacional brasileiro na década de 90; (2) O Sistema Nacional de Avaliação da Educação Básica – SAEB; (3) Apresentando os resultados do desempenho dos alunos avaliados no SAEB 2001; (4) Em busca da eficácia e equidade no Sistema Educacional Brasileiro; e, (5) Considerações finais. Nos anexos constam tabelas, gráficos e quadros referentes aos tópicos discutidos no corpo do relatório.

- Vários tópicos foram abordados no primeiro capítulo, de forma clara e bastante genérica. Foi discutido a respeito do crescimento da matrícula, da distorção idade/série, da educação de jovens e adultos, da municipalização do ensino fundamental, da formação dos professores e do fluxo escolar.
- No segundo capítulo foi apresentado o SAEB, o que ele é, como é constituído e quais os seus propósitos. Foram abordados temas como a evolução da amostra efetiva do SAEB, a partir de 1995, em relação a escolas, diretores, professores, alunos, séries e disciplinas avaliadas; o que o SAEB avaliou em 2001; como são construídos os testes do SAEB; o que informam os questionários contextuais; e como são apresentados os resultados do SAEB.

- No terceiro capítulo foi feita uma discussão sobre a apresentação dos resultados do desempenho dos alunos avaliados no SAEB 2001. O processo de avaliação dos resultados foi explicado tecnicamente, com detalhes. Em seguida, foi apresentada a escala comum de Língua Portuguesa, de duas formas: a escala única e a escala por série. Na escala comum e única, as descrições foram feitas no sentido vertical, considerando-se os tópicos da matriz de referência; na escala por série, as descrições dos níveis de desempenho foram feitas no sentido horizontal, contemplando os tópicos e as habilidades avaliadas, demonstrando que à medida que os níveis são mais altos, habilidades mais complexas vão sendo acrescentadas, expressando a conotação de que o conhecimento vai sendo construído. Essa idéia vai ao encontro do que é anunciado no referencial teórico. Em seguida à apresentação das escalas, são apresentados exemplos de itens para todos os níveis da escala por série. Após a exemplificação, são feitas algumas considerações sobre o desempenho dos alunos por Unidade da Federação, região e Brasil, com gráficos e tabelas comparativos e com gráfico mostrando a significância estatística da diferença entre as médias de desempenho entre Unidades da Federação. São abordados ainda, comentários e gráficos e tabelas correspondentes às médias de desempenho por disciplina, série e localização, dependência administrativa. Esses mesmos procedimentos se repetem para todas as séries e disciplinas.
- No quarto capítulo, é feita uma discussão sobre os fatores que estão associados ao desempenho dos alunos e sobre a consolidação da eficácia e da equidade no sistema Educacional Brasileiro, enfocando sobre os principais problemas considerados: defasagem idade/série; variáveis intra e extra-escolares (quantidade de livros em casa e hábitos de leitura); variáveis relacionadas com a escola (formação diretor, salário e satisfação do diretor com o seu salário, formação do professor, salário e satisfação do professor com o seu salário, prática pedagógica dos professores); o efeito escola; o efeito da raça/cor; e do nível socioeconômico.
- Ao final são feitas considerações sobre as principais constatações e algumas orientações de procedimentos que ajudariam a superar os problemas.

Tratando-se de um relatório nacional, o conteúdo foi em grande parte adequado. Na apresentação fica implícito que o relatório se destina a gestores de políticas públicas, em todos os níveis, diretores e professores. Houve algumas explicações e gráficos excessivamente técnicos que seriam dispensados em um relatório a esse público. Esses adendos podem vir em

relatórios técnicos cujos leitores são especialistas no assunto. Ao final, os anexos acrescentam pouco em termos de informações, podendo, até quem sabe, ser dispensados.

Em relação aos relatórios anteriores, esse representou um avanço, tanto em abrangência dos conteúdos quanto na linguagem e no esforço para que este fosse um instrumento que subsidiasse principalmente políticas por elaboradores de políticas públicas.

Outras duas publicações relativas aos resultados dessa edição do SAEB são: SAEB 2001 – Relatório – Língua Portuguesa e SAEB 2001 – Relatório – Matemática. Brasília: INEP (2002). Essas publicações tiveram grande impacto na área educacional, pois foram as primeiras publicações destinadas aos professores, tratando dos resultados do SAEB. Os dois relatórios têm o mesmo formato, com diferencial do conteúdo abordado, um trata da Língua Portuguesa, com 106 páginas, e o outro da Matemática, com 71 páginas.

Esses documentos foram estruturados em sete capítulos dispostos da seguinte forma: (1) Introdução; (2) O Sistema Nacional de Avaliação da Educação Básica; (3) Apresentando os resultados do desempenho dos alunos avaliados no SAEB 2001; (4) Análise dos itens do teste de Língua Portuguesa / Matemática 4ª série do Ensino Fundamental; (5) Análise dos itens do teste de Língua Portuguesa / Matemática 8ª série do Ensino Fundamental; (6) Análise dos itens do teste de Língua Portuguesa / Matemática 3ª série do Ensino Médio; (7) Análise do desempenho dos alunos em Língua Portuguesa / Matemática.

- Os documentos contêm logo após a apresentação, uma carta dirigida ao professor. Na introdução há uma contextualização da avaliação em relação ao momento de sua ocorrência, com dados a respeito da situação em que se encontra a educação básica brasileira (crescimento da matrícula, fluxo escolar, defasagem idade/série, novos programas e desafios).
- No capítulo seguinte, é apresentado o SAEB. Essa apresentação foi feita de forma topicalizada, com uma linguagem mais clara, procurando facilitar a compreensão do que consiste esse sistema. Foram abordados os seguintes temas: o que o SAEB avaliou em 2001; como são construídos os testes; como são analisados os resultados do SAEB e construídas as escalas de desempenho; e como são apresentados os resultados.
- No terceiro capítulo, que trata da apresentação dos resultados do desempenho dos alunos avaliados, em sua introdução é feita referência à tendência de estabilidade dos resultados em relação às avaliações anteriores; em seguida, é mostrado como os

resultados serão apresentados e como devem ser compreendidos. O primeiro tópico desse capítulo tem por tema: “*O ensino da Língua Portuguesa / Matemática, uma reflexão sobre a prática*”. Nele são discutidas as atuais tendências do ensino da Língua Portuguesa e da Matemática. O tópico seguinte é “*Conversando sobre os resultados*”. Nele são discutidos o modo como se procedeu à análise dos itens e como os itens se relacionam à escala de desempenho.

- No quarto/quinto/sexto capítulos, referentes à análise dos itens do teste para 4^a/8^a EF e 3^a série EM, é apresentada, em primeiro lugar, a escala de desempenho, por série, onde as habilidades demonstradas pelos alunos são apresentadas associadas aos tópicos / temas das matrizes (Língua Portuguesa e Matemática), em um sentido crescente dos níveis de desempenho dos alunos. Em seguida são feitos comentários e apresentados itens ilustrando diferentes níveis da escala, por tópico / tema da matriz. Os itens apresentados são acompanhados das análises que mostram o comportamento do aluno, com seu provável processo mental utilizado para chegar à resposta ao item. Esses comentários são feitos em linguagem que o professor possa se identificar no processo e de modo que ele compreenda o processo de construção do conhecimento. Após a análise dos itens, abrangendo os diversos níveis da escala e todos os tópicos / temas, vem um tópico, cujo tema é: “Interpretação dos resultados do desempenho dos alunos da 4^a e 8^a séries EF e 3^asérie EM em Língua Portuguesa / Matemática – Brasil, Regiões e Unidades da Federação, em que é mostrado o desempenho dos alunos em forma gráfica e com comentários, considerando-se todas essas unidades de análise.
- No sétimo capítulo, a respeito da análise do desempenho dos alunos em Língua Portuguesa / Matemática, são feitos comentários sobre o que revela o desempenho dos alunos nos diversos tópicos/temas das matrizes, com um alerta sobre a importância de tomar os resultados como subsídios para transformar a realidade encontrada na educação básica brasileira, considerando-se a construção de competências que se esperava já estarem consolidadas nos níveis de escolarização avaliados. Neste capítulo final são abordados os tópicos “Considerações sobre o desempenho dos alunos da 4^a / 8^a EF e 3^a série EM, em Língua Portuguesa / Matemática” e “Conclusões” com comentários sobre o que foi constatado em cada uma dessas instâncias analisadas.

Publicações desse tipo deveriam fazer parte dos produtos de cada edição do SAEB, e como foi o caso dessas publicações, deveriam ser distribuídas, no mínimo, às escolas que

fizeram parte da amostra. Isso denota ao menos atenção com os participantes, que podem se sentir coadjuvantes na busca da melhoria da qualidade, além de procurar projetar e tornar transparente o processo.

Entre as publicações referentes aos resultados da edição SAEB 2001, a última a ser comentada é “Qualidade da Educação: uma nova leitura do desempenho dos estudantes da 4ª série do Ensino Fundamental. Brasília: INEP (abril/2003)”. Trata-se de uma publicação do início de uma nova gestão na DAEB.

Este documento contém 31 páginas e está estruturado da seguinte forma: (1) Introdução; (2) Os desempenhos e seus contextos; (3) Brasil e Regiões: desigualdades educacionais e sociais; (4) Evolução do desempenho dos alunos – 1995 a 2001.

- No capítulo introdutório é dada a tônica da publicação. Por meio dela pretende-se chamar a atenção para a análise do contexto na compreensão dos resultados obtidos. Inicia-se esse capítulo com um dado bastante chamativo: *59% dos alunos brasileiros chegam à 4ª série do ensino fundamental sem terem desenvolvido competências e habilidades elementares de leitura e 52% desses mesmos alunos demonstram profundas deficiências em Matemática.* A partir desses dados é feita uma reflexão a respeito de como deve ser feita essa leitura, a quem deve ser atribuída a culpa de resultados tão desastrosos. E vai-se contextualizando o problema ao longo do discurso. Chama-se a atenção sobre o papel da avaliação enquanto processo que deve apresentar conclusões e deve estar associada a *uma ação promotora de mudança destacando a necessidade de fortalecer sua articulação com as instâncias de formulação de políticas que possibilitem a elaboração de programas de intervenção mais ajustados às necessidades diagnosticadas.*
- O capítulo seguinte sobre os desempenhos e seus contextos, inicia com uma justificativa de se classificar a média de desempenhos em estágios de aquisição dos conhecimentos. Dessa forma, são adotadas e discutidas, sem conhecimento de causa, as cinco categorias de desempenho arbitradas para o processo de avaliação do SAEB, em Língua Portuguesa e Matemática: muito crítico, crítico, intermediário, adequado e avançado. O discurso utilizado ao longo desse capítulo é pessimista e procura ressaltar, de forma negativa os dados da avaliação. São apresentadas tabelas com as descrições do que os alunos em cada estágio são capazes, e tabelas com as populações e os percentuais de alunos representativos em cada estágio para o Brasil e Regiões.

- O terceiro capítulo fala sobre as desigualdades educacionais e sociais relativas ao Brasil e Regiões. São apresentados inicialmente, alguns indicadores sociais e educacionais, dos quais são feitos comentários, relacionados ao desempenho dos alunos e, em seguida, são abordadas algumas variáveis consideradas fundamentais nas análises da qualidade da educação básica. São elas: escolaridade e salário dos professores; trabalho infantil; mudanças de escola e conseqüências no desempenho; adoção de ações de melhoria da aprendizagem; condições das escolas da rede pública e privada. O formato desse capítulo se apresenta um pouco desorganizado, pois não há uma associação entre as variáveis. Cada uma tem uma análise isolada.
- O quarto e último capítulo aborda a evolução do desempenho dos alunos entre 1995 e 2001. Os comentários feitos neste capítulo são sob a perspectiva dos estágios de construção das competências. São apresentados tabelas e gráficos diversos acentuando a queda de desempenho dos alunos ao longo das edições do SAEB, relacionados ao Brasil e Regiões e por rede de ensino.

Procedendo-se a uma análise crítica dessa publicação, observa-se que ela tem um caráter sociológico, chamando a atenção basicamente para indicadores que se apresentam nessa circunstância. Não dá para distinguir o destinatário dessa publicação. A linguagem é fria e carregada de expressões negativas. Considera-se inadequada para os atores que compõem a área educacional.

6.5. EDIÇÃO 2003

Em relação a publicações do SAEB 2003, não houve nenhuma publicação impressa. Nesta edição, como na de 1999, houve também muitos paradoxos. Embora não tenha havido publicações impressas, essa foi a edição mais divulgada no Brasil. Houve um trabalho de divulgação dos resultados do SAEB em todas as regiões. Além disso, em todos os estados do Brasil houve um evento denominado “Oficina de elaboração de itens”, no qual, na primeira parte os dados do SAEB eram divulgados, mostrando os resultados de cada estado, sempre de forma estatística, por meio de gráficos. A segunda parte do evento, era destinada a um diálogo com professores exclusivamente de sala de aula, quando se tinha oportunidade de explicar as escalas de desempenho e trabalhar o processo de elaboração de itens. O material impresso

distribuído e trabalhado nesses eventos eram os relatórios pedagógicos relativo ao SAEB 2001, para cada disciplina avaliada.

6.6. EDIÇÃO 2005

Quanto à edição do SAEB 2005, neste momento, está sendo divulgado, pela Internet, o primeiro relatório, denominado “SAEB 2005 – Primeiros Resultados: Médias de desempenho do SAEB 2005 em perspectiva comparada”. Esses dados encontram-se no *site* do INEP e foram consultados em 03/03/2007. Embora não devesse ser objeto de análise deste estudo, mas vale a pena comentá-lo. Trata-se de uma publicação com 62 páginas, dividida em duas partes: (1) Introdução e (2) Dados gerais.

Na primeira parte, “Introdução”, são feitos comentários gerais sobre o principal papel do SAEB: *coletar informações sobre o desempenho acadêmico dos alunos brasileiros apontando o que sabem e o que são capazes de fazer em diversos momentos do seu percurso escolar* (p. 3). Logo em seguida, é ainda reiterado o objetivo do SAEB de *oferecer subsídios para a formulação, reformulação e monitoramento de políticas públicas contribuindo, dessa maneira, para a universalização do acesso e a ampliação da qualidade, da equidade e da eficiência da educação brasileira*. É comentado também o avanço das técnicas de análises, equalização, amostragem e construção dos itens. Ao final, são apresentados dados comparativos do quantitativo de escolas e alunos avaliados por edição do SAEB e informações sobre a caracterização das escolas pesquisadas ao longo de todas as edições do SAEB.

Na segunda parte, são apresentadas 58 tabelas informando as médias de desempenhos das Unidades da Federação, Região e Brasil por edições do SAEB e localização.

Observa-se uma grande inconsistência entre o que é apresentado na introdução e o que é apresentado como resultado. Logo no início da introdução está confirmado o papel do SAEB: *coletar informações sobre o desempenho acadêmico dos alunos brasileiros apontando o que sabem e o que são capazes de fazer em diversos momentos do seu percurso escolar*, e em seguida, reiterado o seu objetivo de *oferecer subsídios para a formulação, reformulação e monitoramento de políticas públicas contribuindo, dessa maneira, para a universalização do acesso e a ampliação da qualidade, da equidade e da eficiência da educação brasileira*. As

questões são: “São apresentados dados sobre o que os alunos brasileiros demonstraram saber e serem capazes de fazer em diversos momentos do seu percurso escolar”? “Essas informações oferecem algum subsídio para formular alguma política que ajude a superar os problemas detectados”? “Quais são os problemas detectados”?

Os dados mostram apenas os estados, regiões e localização (capital e interior) que apresentam melhor ou pior média de desempenho nas séries e disciplinas avaliadas. E esses dados serviram para causar um grande alarde por meio da imprensa escrita e falada, que apresentaram, na realidade, dados ínfimos sobre a situação real da educação básica brasileira.

Conclui-se, ao longo deste estudo, que apenas uma edição do SAEB, a de 2001, produziu relatórios capazes de indicar informações que subsidiassem a formulação de ações que buscassem superar o estado de coisas encontrado. Tomando-se por base os relatórios produzidos em cada uma das edições, e considerando-se a linguagem adotada, os dados disponibilizados e os atores capazes de promover mudanças, verifica-se que quase nada poderia ser feito em prol da melhoria da qualidade da educação básica brasileira, tendo por base os dados coletados pelo SAEB.

Tem havido uma prática, que já está se tornando quase uma cultura, de tomar os resultados do SAEB apenas em números e transformá-los em uma tragédia nacional, apontando denúncias falaciosas e inócuas do ponto de vista da busca de soluções. A apresentação dos resultados à imprensa tem sempre causado um grande ruído nacional. E este tem sido o principal papel do SAEB.

Para confirmar esses dados, no capítulo seguinte é apresentado um estudo sobre o alcance do principal objetivo do SAEB, previsto desde a sua criação, qual seja o de oferecer subsídios para a formulação, reformulação e monitoramento de políticas públicas que contribuam para a melhoria do processo educacional brasileiro.

7. PESQUISA DE POLÍTICAS PÚBLICAS IMPLEMENTADAS COM BASE NOS RESULTADOS DO SAEB

Em diversos documentos do SAEB e relacionados a ele, em sites e artigos científicos (SAEB 2001 – Novas Perspectivas, 2002; Relatórios SAEB 2001, 2002; SAEB 97: Primeiros Resultados, 1999; Matrizes Curriculares de Referência, 1997, 1997; Matrizes Curriculares de Referência, 1999; PCN, 1998), consta como um dos principais objetivos do SAEB oferecer subsídios à formulação, reformulação e monitoramento de políticas públicas e programas de intervenção voltados para a melhoria da qualidade da educação no Brasil.

Encontra-se no *site* do SAEB:

... as informações obtidas a partir dos levantamentos do Saeb permitem acompanhar a evolução da qualidade da educação ao longo dos anos, sendo utilizadas principalmente pelo Ministério da Educação e secretarias estaduais e municipais de educação na definição de ações voltadas para a solução dos problemas identificados, assim como, no direcionamento dos seus recursos técnicos e financeiros às áreas prioritárias, visando ao desenvolvimento do Sistema Educacional Brasileiro e à redução das desigualdades nele existentes (<http://www.inep.gov.br/basica/saeb/objetivos.htm>, em 20/03/2007).

Neste capítulo, discutem-se os resultados da pesquisa feita a respeito do alcance desse objetivo, em especial ao que tange às políticas públicas implementadas tendo por base os resultados do SAEB. É importante considerar que os respondentes dessa pesquisa foram,

em sua maioria, os responsáveis pelo SAEB nos Estados, podendo, portanto, as respostas não conterem todas as informações relacionadas às políticas públicas.

Políticas públicas são geralmente consideradas como o Estado em ação, ou seja, é o conjunto de orientações e ações de um governo que visa ao alcance de determinados objetivos (Kerbaux, 2006). Dessa forma, as soluções propostas por parte dos governos federal, estaduais e municipais podem ser mais bem dimensionadas pelo uso do conhecimento da realidade na qual se busca superar problemas.

Nesses âmbitos governamentais, o processo de formulação de políticas públicas depende de como os assuntos chamam a atenção dos elaboradores, de como são formulados e de como os dados possibilitam construir uma proposta. Dessa forma, em primeiro lugar, o elaborador deve estar munido de um conjunto de dados que aliados aos valores, ideais, princípios e vontade do governo superar problemas, o estimularão a transformá-los em informações relevantes que produzirão conhecimento de ações orientadas. Em seguida, com base no conhecimento empírico e normativo essas ações orientadas serão transformadas em ações públicas.

Assim, se entre os objetivos do SAEB está “*oferecer subsídios à formulação, reformulação e monitoramento de políticas públicas e programas de intervenção ajustados às necessidades diagnosticadas nas áreas e etapas de ensino avaliadas*”, em seus relatórios de resultados e nos contatos com secretarias municipais e estaduais deveriam ser oferecidas orientações de ações que pudessem efetivamente ser implementadas.

Sabe-se que o SAEB é uma avaliação sofisticada que produz muitos dados relevantes na formulação de políticas públicas. Mas, no entanto, o que se percebe é que não existe um trabalho efetivo de incentivo, ou orientação de tais políticas. Por meio deste estudo pretendeu-se chamar a atenção para este importante papel que o SAEB deveria estar desempenhando e que parece não estar alcançando, neste momento.

Para avaliar a consecução desse objetivo, foi realizada pesquisa com as secretarias de educação de todos os estados brasileiros e Distrito Federal e com o Ministério da Educação, no Departamento de Políticas de Educação Infantil e Ensino Fundamental e Ensino Médio da Secretaria de Educação Básica, questionando sobre eventuais políticas públicas que tivessem sido implementadas com base nos resultados do SAEB, no período de 1995 a 2003.

Como relatado no Método, em janeiro deste ano foi enviado e-mail (Apêndice A4) aos gestores de avaliação de todas as secretarias de estado. Responderam à consulta apenas

quatro gestores. Foi reenviada a consulta em fevereiro, da qual se obteve mais três respostas. Como a intenção era pesquisar todos os estados, a solução foi realizar ligação telefônica para todos os que não tinham enviado respostas. Fazendo-se um balanço das consultas, obteve-se que um estado, não respondeu à consulta, pois o responsável encontrava-se em férias e nenhuma outra pessoa sabia informar e outros dois estados disseram que iriam fazer uma pesquisa e não retornaram. A maioria das justificativas pelas não-respostas aos e-mails foi a mudança de governo, na qual muitos haviam colocado o cargo à disposição e nem receberam a mensagem, pois grande parte tinha como endereço eletrônico as próprias secretarias. Outra justificativa, foi que como a resposta era negativa, dispensava retorno. Quanto à pesquisa no MEC, ela foi feita pessoalmente com uma das técnicas do Departamento anteriormente citado.

Os resultados da pesquisa indicaram que:

- Os 24 respondentes dos estados mais a respondente do MEC à consulta informaram não ter notícia a respeito de qualquer política pública implementada com base nos resultados do SAEB, no período de 1995 a 2003 e que o monitoramento de políticas públicas e programas de intervenção não existe, pois nunca houve esse tipo de relação entre as secretarias e o SAEB.

Entretanto, tanto a consulta por e-mail como a entrevista por telefone foi muito esclarecedora. Houve muitos depoimentos que vêm ao encontro do que este estudo tem chamado a atenção. A seguir, são relatadas algumas declarações dos respondentes.

- Houve relato de 14 estados de que o maior feito do SAEB foi a implementação da cultura de avaliação, em especial, a partir de 2001, em que os resultados começaram a ser divulgados e compreendidos nos estados.
- Desses estados, gestores relataram que além de terem sido estimulados a criar seus próprios sistemas de avaliação, o SAEB forneceu itens para serem equalizados com o sistema nacional e alguns ainda relataram ter recebido apoio técnico;
- Alguns gestores citaram como políticas implementadas com base nos resultados do SAEB, o Programa Nacional do Livro Didático, o Fundo Escola, Biblioteca na escola, PROINFO, PCN, Programa de Aceleração da Aprendizagem, entre outros;
- Houve reclamações de vários gestores em relação à apresentação e compreensão dos resultados do SAEB; esses gestores sentem falta de uma linguagem em que tanto eles próprios, como todos os envolvidos com educação possam entender;

- Alguns reclamaram da falta de parâmetros estáveis em relação à escala de proficiência;
- Outros disseram ter reivindicado ao SAEB, mas não ter recebido resposta, a respeito de resultados para o estado com base no desempenho dos alunos nos descritores das matrizes;
- Muitos disseram que se sentem revoltados pelo discurso de diversos políticos que chamam a atenção apenas pelo fracasso e atribuem a culpa aos professores;
- Gestores de dois estados chamaram a atenção pela forma de apresentação dos resultados à imprensa, principalmente em noticiários estaduais, nos quais classificam o desempenho dos alunos sem considerar o que conquistaram, ou os fatores que justificam o resultado do estado;
- Foram feitos vários comentários sobre a “Prova Brasil” e a maioria afirma que faria uma avaliação mais ampla e efetiva se fossem enviados recursos aos estados e municípios;
- Grande parte dos respondentes fez comentários sobre os seus próprios sistemas de avaliação, ressaltando as várias políticas implementadas, em especial, a partir desse ano, com base nas suas avaliações. Entre esses, podem-se citar:
 - Pela Secretaria de Educação do Estado do Espírito Santo, a gestora de avaliação, Marluza M. Balarini, ressalta que *com base na avaliação estadual, no sentido de disseminar a cultura da avaliação, vem progressivamente, despertando o interesse e aumentando a compreensão das Secretarias municipais que têm manifestado o desejo de aderirem à avaliação estadual. Na própria Secretaria, a área pedagógica manifesta também um crescente interesse pelos resultados do SAEB e das demais avaliações. Tanto assim, que decisões ligadas à **ampliação da jornada escolar diária do aluno**, a ênfase à **leitura** e à **matemática** no currículo escolar em 2007, decorrem da análise dos resultados do PAEB-ES (Programa de Avaliação da Educação Básica do Espírito Santo), do SAEB e da Prova Brasil.*

A discussão dos resultados das avaliações nas escolas necessita de um mediador, no caso, a Secretaria de Educação, sem o qual qualquer material enviado não será apropriado pelos professores, diretores etc.

No Espírito Santo, há um Plano de Estado (até 2025) e um Plano de Governo (2007-2010) que estabeleceu metas a partir dos resultados das avaliações estadual e nacionais. Uma dessas metas diz respeito ao aumento da proficiência em leitura e matemática no ensino fundamental e médio.

*Em síntese, para que as avaliações nacionais alcancem o objetivo estabelecido, é preciso mais do que divulgar dados quantitativos situados em escalas de mensuração. É preciso **formar** equipes técnicas estaduais, preparar materiais que sejam compreensivos aos professores, e úteis ao trabalho que se desenvolve nas escolas, nas salas de aula.*

- Pela Secretaria do Estado de Tocantins, Magali L. Chaves, a respeito do seu sistema de avaliação, destaca que *o foco da política educacional do Estado do Tocantins, definida pelo Planejamento Estratégico da Secretaria – PES, está na melhoria da qualidade do Sistema Estadual de Ensino, por intermédio da implantação e implementação de Programas de Avaliação Permanente; Profissionalização, Responsabilização e Valorização dos Profissionais da Educação; Reestruturação, Modernização e Consolidação da Gestão Compartilhada.*

Segundo Magaly, considerando que as avaliações do SAEB eram por amostragem, o Tocantins em parceria com a Fundação CESGRANRIO, em 2001, realizou a primeira avaliação censitária de todas as escolas da Rede Estadual de Ensino, utilizando-se das matrizes de referências do SAEB.

Na atualidade, afirma Magaly, a Secretaria da Educação tocantinense passa a utilizar os resultados da PROVA BRASIL, ANEB, PISA, ENEM, ENCEJA e do Sistema de Acompanhamento do PES (Planejamento Estratégico da Secretaria), para nortear os planos de trabalho dos Programas de Avaliação e Acompanhamento do Desempenho dos Educandos, da Escola e da Avaliação Institucional, bem como para o Programa de Formação Continuada dos Professores com o objetivo de melhorar os níveis de desempenho apresentados na escala do SAEB.

- Pelo Estado de São Paulo, a gestora de avaliação, Prof^a Maria Conceição Conholato, em sua resposta ao e-mail destacou que a Secretaria de Educação *tem utilizado os resultados da avaliação estadual (SARESP) como instrumento*

orientador das tomadas de decisão em todas as instâncias centrais da SEE, bem como das Diretorias de Ensino e em suas respectivas escolas, visando sempre, ao aprimoramento da gestão educacional, ao melhor aproveitamento escolar e, portanto, à redução e conseqüente melhoria nos índices de evasão e retenção no Estado. Em nível central, os resultados têm subsidiado a definição e/ou reformulação da política educacional; o redirecionamento dos projetos pedagógicos, a definição do processo de formação dos educadores, a elaboração de materiais pedagógicos, entre outros. Em nível das escolas e Diretorias Regionais de Ensino os resultados têm favorecido a revisão da prática educativa e as demandas de capacitação dos professores.

- Pelo Estado de Alagoas, a gestora de avaliação, Maria de Fátima Lima, ressalta a importância de *implementar a rede nacional de avaliação da educação básica, mediante convênio de cooperação técnica e financeira entre o INEP/MEC e os estados da federação, sendo respeitada a autonomia dos estados.* Além disso, ela chama a atenção para que *os resultados das avaliações dos sistemas educacionais sejam direcionados e amplamente divulgados entre os gestores educacionais, professores, instituições de ensino superior e a sociedade em geral, mas, fundamentalmente, que sejam trabalhados durante a formação continuada dos professores, visando confrontar as dificuldades dos alunos constatadas na análise pedagógica com a prática docente.*
- Por fim, um depoimento mais extenso, feito pela gestora de avaliação do Estado do Maranhão, Prof^a. Silvana Machado, que confirma que o estado não implementou políticas públicas com base exclusivamente nos resultados do SAEB, contudo, o SAEB cedeu em 2001, 80 itens para comporem os testes de Língua Portuguesa e Matemática, permitindo a equalização da avaliação estadual com a nacional, além dos microdados do estado, que possibilitaram proceder a análises dos fatores associados ao desempenho. Um dos seus comentários, a seguir, resume o que muitos respondentes afirmaram:

O SAEB surge com o objetivo de

- *Monitorar o desempenho dos sistemas de ensino, oferecendo informações que permitissem aos gestores conhecer com maior profundidade os problemas e deficiências de nosso sistema*

educacional, orientando com maior precisão as políticas governamentais destinadas à melhoria da qualidade e à promoção da equidade, na oferta de educação básica.

No entanto, como agente promotor de uma política de avaliação para a nação, o MEC tem tomado diferentes rumos no decorrer dos sucessivos processos históricos e políticos.

Os esforços empreendidos, da legitimidade do sistema, dos consensos e parcerias estabelecidos com os sistemas de ensino estaduais e municipais, a avaliação nacional vem sendo progressivamente apropriada pelos sistemas de ensino no sentido de uma “cultura avaliativa”. Contudo, apesar do SAEB ter se tornado uma referência em avaliação, não poderia responder a todas as necessidades de informação, porque não foi construído para apreender toda a diversidade e singularidade das escolas.

*Com a tendência de queda dos resultados de 95 a 2001, o Governo Federal inicia um novo movimento descentralizador da política de avaliação nacional. Surge a discussão de criarmos uma “**rede de avaliação**” que possibilitasse aos estados o fortalecimento das iniciativas locais (que é muito divergente) num acordo conjunto acerca da padronização do trabalho de campo e das metodologias na análise da TRI. Neste período ocorreu:*

- o estímulo à criação e ao fortalecimento de sistemas próprios de avaliação (rede nacional de avaliação);*

- a ampliação do escopo da avaliação, a fim de possibilitar a obtenção de dados e construção de indicadores por unidades de ensino, impactando mais fortemente na gestão escolar e na condução das políticas municipais e estaduais de educação.*

Em 2005, surpreendentemente, decreta-se a Portaria Ministerial nº 931 de 21 de março de 2005 que institui o Saeb como Sistema que abriga as principais pesquisas sobre a Educação Básica: um de caráter sistêmico – geral, e um outro mais extenso e detalhado com foco em cada unidade escolar, ou seja, a ANEB (antigo SAEB) e a ANRESC.

Voltamos, então, à centralização da política de avaliação do MEC. As experiências dos estados que já haviam avançado na área de avaliação de

sistema não foi sequer considerada pelo INEP. No meu ponto de vista, um retrocesso, o que ocasionou em muitos estados uma “overdose” de avaliações. Uma mesma escola foi avaliada na ANRESC, na ANEB, na Avaliação Estadual, assim, por parte da escola houve muitas confusões sobre as avaliações, dificultando a disseminação de uma cultura de avaliação sistemática e organizada.

Considerando as idas e vindas nos objetivos e proposições do INEP/MEC para a implantação de uma Política de Avaliação, parece-nos que muito mais proveitoso para os sistemas de ensino seria o fortalecimento das equipes locais, com uma padronização consensual e análise de dados na mesma metria. Mesmo sabendo que os estados se encontram em estágios bem diferenciados em relação às suas experiências em Avaliação de Sistema.

Conclui-se que os subsídios à formulação, reformulação e monitoramento de políticas públicas e programas de intervenção que o SAEB oferece, não conseguem estimular os formuladores de políticas dos estados e do próprio MEC a proporem políticas. Esse episódio pode ser decorrente, em parte, pelo fato de os dados fornecidos estarem desconectados do público alvo.

Pelos depoimentos, observa-se o reconhecimento pela disseminação da cultura da avaliação, mas verifica-se também, grande descontentamento dos responsáveis pelas avaliações nos estados, pela forma de apresentação de resultados. O mesmo foi observado nas secretarias do MEC. Se o SAEB visa oferecer subsídios para a formulação e implementação de políticas públicas que objetivem melhorar a qualidade da educação básica brasileira, seus relatórios e suas ações devem estar voltados para esse fim. Se se detecta que o problema é a linguagem, que seja feita a tradução desses dados, que se mantenha diálogo com as secretarias, no sentido de esclarecer o que os dados apontam e de se cobrar ações mais efetivas.

Afinal, a pesquisa do SAEB é realizada de forma muito sofisticada, mas seus resultados, quando saem, acabam ficando engavetados em gabinetes, porque educadores que são os principais interessados, não se identificam nessa pesquisa. No Brasil, nas secretarias de educação, os formuladores de políticas são em sua maioria pessoas com formação em educação, que como se sabe não têm entendimento de estatísticas e quem produz os relatórios são pessoas que imaginam que os dados apresentados estejam claros e compreensíveis ao seu público alvo.

Comparando-se os fins convencionados para a formulação das políticas e dos programas com os resultados apresentados examina-se até que ponto vem sendo cumprida a meta de ação e quais as causas da não exeqüibilidade da implementação. Afinal, o processo de implementação de programas e políticas é um processo sistemático de análise de atividades que permite compreender, de forma contextualizada, todas as suas dimensões e implicações e que tem por objetivo orientar futuras ações.

3ª PARTE

CONSIDERAÇÕES FINAIS

CONSIDERAÇÕES FINAIS

Ao tomar o termo “avaliar” em seu sentido literal, o associamos a determinação de valor, estimativa de mérito, cálculo, apreciação, entre outros. Ao relacioná-lo ao contexto educacional, ele adquire inúmeros outros sentidos isolados, dependendo da análise que se pretenda fazer e do ponto de vista a se considerar. Assim, do ponto de vista político, avaliar tem um sentido, do estatístico, tem outro, bem como do filosófico e do social. Atualmente, tem-se observado que se desenvolveu o hábito de considerar o ato de avaliar de forma isolada, ou seja, em função da perspectiva que o avaliador esteja considerando, de modo a não relacioná-lo aos seus inúmeros sentidos.

Na introdução deste estudo foram abordados diversos significados que a avaliação assume em seu contexto educacional, dependendo da perspectiva em que ela esteja sendo considerada. Ao finalizar este estudo, pode-se concluir que a avaliação no contexto educacional sistêmico só pode ser considerada enquanto processo e que ela só será completa se forem assumidos todos os seus significados possíveis, pois *seu papel deve ser mais construtivo, até porque avaliar é muito mais do que medir* (Gatti, 2002).

Dessa forma, considera-se que a avaliação sistêmica deve ter por objetivo abordar e atingir os mais variados propósitos possíveis, desde diagnóstico do desempenho, monitoramento do progresso dos alunos, avaliação do currículo escolar, referência na prestação de contas, orientação de políticas públicas, planejamento de programas, até como suporte de uma reflexão da prática pedagógica. A avaliação, portanto, para atingir esses diversos propósitos, não faz sentido se não for implementada como um processo ativo para o qual é possibilitada a condição de emitir um julgamento de valor por meio da medida, da estimativa, da classificação e da prestação de contas.

Este estudo permitiu conhecer com maior profundidade o que constitui o Sistema Nacional de Avaliação da Educação Básica, quais são suas ações e o que poderia estar sendo feito para alcançar maior efetividade com base nos dados por ele produzidos.

Pôde-se constatar que o SAEB representou um grande avanço na cultura avaliativa da educação no Brasil, principalmente ao considerar que foi a partir de 1995 que o SAEB começou a obter projeção, e que hoje já conta com 19 sistemas estaduais de avaliação educacional instalados, tendo a maioria se inspirado no modelo do SAEB. Além desses sistemas implantados, o SAEB contribuiu para a formação e o aperfeiçoamento de inúmeros pesquisadores e técnicos em avaliação educacional. Hoje, inclusive, conta com uma Associação Brasileira de Avaliação Educacional – a ABAVE.

Além das já consagradas Fundação Cesgranrio e Fundação Carlos Chagas, que já trabalhavam com avaliações sistêmicas, surgiram, inspiradas no SAEB, várias instituições públicas e privadas que se consolidaram na especialização de avaliações educacionais em larga escala. Entre as instituições públicas, podem se citar o Centro de Políticas Públicas e Avaliação da Educação (CAEd/UFJF); o Centro de Pesquisa em Avaliação Educacional (CPAE/IP/UnB); o Centro de Seleção e de Promoção de Eventos CESPE/UnB); o Grupo de Avaliação e Medidas Educacionais (GAME/UFMG); o Centro de Estudos em Avaliação Educacional IE/UFRJ. Entre algumas instituições privadas, podem-se citar: o Instituto Movens; o Núcleo de Avaliação Educacional e Psicológica; o AVALIA Assessoria Educacional; o Abaquer Consultores e Editores Associados, entre outros.

Este estudo permitiu também levantar alguns problemas sobre as ações do SAEB, que, solucionados, assegurarão a ele a condição de instituição com maior poder para orientar políticas que melhorem o nível de qualidade da educação do nosso país. A maioria dos problemas apontados decorre do distanciamento entre o que é proposto como missão do SAEB e o que é praticado. A realização desta pesquisa permitiu elucidar as intenções, a importância e o conhecimento do funcionamento desse sistema. Puderam-se constatar as dimensões abrangidas pela avaliação e as outras várias que poderiam ser atingidas. Entre as referências consideradas estão os próprios objetivos do SAEB que constam na página do SAEB, no *site* do INEP: <http://www.inep.gov.br/basica/saeb/objetivos.htm> (última pesquisa em 20/03/2007), como segue:

- oferecer subsídios à formulação, reformulação e monitoramento de políticas públicas e programas de intervenção ajustados às necessidades diagnosticadas nas áreas e etapas de ensino avaliadas;

- identificar os problemas e as diferenças regionais do ensino;
- produzir informações sobre os fatores do contexto socioeconômico, cultural e escolar que influenciam o desempenho dos alunos;
- proporcionar aos agentes educacionais e à sociedade uma visão clara dos resultados dos processos de ensino e aprendizagem e das condições em que são desenvolvidos, e,
- desenvolver competência técnica e científica na área de avaliação educacional, ativando o intercâmbio entre instituições educacionais de ensino e pesquisa.

Dessa forma, levando em conta tais objetivos e a perspectiva considerada para o desenvolvimento deste estudo, que foi dos testes de desempenho e de seus resultados, várias considerações podem ser feitas fundamentadas nos resultados da pesquisa.

No estudo do processo de avaliação brasileiro sob a perspectiva do modelo norte-americano, referência assumida pela instituição brasileira, foram destacados alguns pontos de evidentes diferenças em relação aos propósitos educacionais e ao sistema de avaliação. Entre as mais visíveis estão: (1) a que diz respeito ao foco da educação americana – os resultados do desempenho dos alunos; (2) a estreita relação entre os propósitos do ensino e a avaliação da educação; (3) a preocupação americana em atualizar, de forma prática e regulamentada por lei, os objetivos instrucionais; (4) a descentralização da avaliação que tem por unidade de análise a escola; (5) o acompanhamento e a co-responsabilidade do processo avaliativo por representantes dos mais variados segmentos sociais; (6) alguns pontos do processo de determinação da amostra e das análises; (7) as publicações relacionadas à avaliação.

Embora o Brasil esteja exigindo medidas que impactem o sistema educacional brasileiro, o foco da educação americana nos resultados é um ponto crítico que não deve ser desconsiderado. A melhoria da educação deve estar associada à consequência dos propósitos educacionais e não deve ter por foco os resultados em si. Inúmeros fatores propiciam ou não melhores resultados. E melhores resultados do desempenho estão relacionados à superação desses fatores intervenientes no processo educacional.

A diferença dos dois países no que diz respeito à estreita relação entre os propósitos de ensino e a avaliação, a pesquisa mostrou que, ao contrário do sistema norte-americano, há um distanciamento entre o que é avaliado e o que é praticado nas salas de aula brasileiras. Percebe-se, inclusive, que o nosso sistema educacional anda em descompasso com as

diretrizes curriculares nacionais. Uma sugestão, que tem por base o sistema educacional norte-americano, é a adoção de padrões que sejam referência para o ensino e para a avaliação.

No Brasil é ressaltado no documento que contém as matrizes, que elas não devem ser tomadas como referência para o sistema de ensino. E da forma como se apresentam, nem é mesmo aconselhável. Em um processo de reformulação, elas poderiam representar padrões esperados de ensino, servindo assim para dar um direcionamento ao processo de ensino. Nos contatos desta pesquisadora com docentes de municípios brasileiros, ouviram-se os depoimentos de alguns deles, de séries iniciais, de que alguns municípios optaram por não contemplar a geometria. Logicamente, que isso afetará o resultado final daquele estado, pois nas questões relativas à geometria haverá um mau desempenho dos alunos, que incidirá sobre os resultados. Dessa forma, pensa-se que se houvesse uma orientação nacional, partindo do sistema de avaliação, problemas dessa natureza seriam evitados. É preciso haver consonância entre o que é aprendido e o que é avaliado.

Outra diferença evidente, diz respeito à atualização constante de objetivos instrucionais. Observa-se que nos Estados Unidos as metas educacionais acompanham as necessidades que o período histórico exige e são regulamentadas por lei. A cada re-autorização da lei nacional de educação, há uma atualização dos objetivos educacionais que é naturalmente acompanhada pelo processo de revisão das matrizes de referência da avaliação. Dessa forma, observa-se uma constante adaptação do sistema às novas exigências que cada momento impõe. No Brasil verifica-se que não tem havido um trabalho efetivo de trabalhar na escola o desenvolvimento de habilidades que promovam a inserção dos alunos às necessidades que o mercado de trabalho atualmente demanda.

Outra sugestão, que tem por base o sistema norte-americano, e que poderia ser adotada pelo nosso sistema educacional e de avaliação, é em relação à descentralização da avaliação que toma por unidade de análise a escola. Naquele país, são os próprios estados e municípios que administram os seus processos avaliativos. Isso dá a eles poder para realizá-los com maior amplitude e competência, fornecendo resultados mais pontuais, dando condições de serem implementadas políticas mais diretas e eficientes. Em nosso sistema, a mais recente avaliação, realizada no ano 2005, que toma por unidade de análise a escola (ANRESC), abrangeu 40.920 estabelecimentos escolares e 3.304.350 alunos de escolas de zona **urbana** e que tivessem **mais de 30 alunos** matriculados na 4ª série e/ou na 8ª série. Considerando-se que o louvável propósito dessa avaliação seria fornecer subsídios para que a escola promovesse mudanças, buscando a melhoria da qualidade da educação básica

brasileira, pode-se constatar que este seu propósito acabou ficando bastante restrito e não atingindo quem mais precisava ser atingido. Em síntese, a realidade⁹ que se apresentava naquele momento está descrita nas Tabelas 53 e 54.

Tabela 53 – Número de estabelecimentos de ensino da educação básica e ensino fundamental instalados na zona urbana e rural e proporção, em 2005

Regiões e Brasil	Educação Básica				Ensino Fundamental			
	Urbana	Prop. urbana	Rural	Prop. rural	Urbana	Prop. urbana	Rural	Prop. rural
Norte	7.025	0,27	18.850	0,73	5.160	0,22	18.167	0,78
Nordeste	31.540	0,36	55.366	0,64	23.984	0,32	51.624	0,68
Sudeste	46.181	0,80	11.745	0,20	26.820	0,71	10.763	0,29
Sul	18.220	0,69	8.148	0,31	10.658	0,59	7.483	0,41
Centro Oeste	7.711	0,77	2.248	0,23	5.702	0,71	2.376	0,29
Brasil	110.677	0,53	96.357	0,47	72.324	0,44	90.413	0,56

Tabela 54 – Número de estabelecimentos de educação do ensino fundamental com até 30 alunos e proporção por região

	Ensino Fundamental		
	Escolas com até 30 alunos	Total de escolas	Proporção na região
Norte	9.214	23.317	0,39
Nordeste	25.269	75.608	0,33
Sudeste	6.820	37.583	0,18
Sul	4.174	18.141	0,23
Centro Oeste	1.289	8.078	0,16
Brasil	46.766	162.727	0,29

Na Tabela 53 consta o número de estabelecimentos de ensino localizados na zona urbana e rural, referentes à educação básica e ao ensino fundamental. Na Tabela 54 está o número de estabelecimentos do ensino fundamental com até 30 alunos, por região. Esses números apresentados tendem a ser bem maiores, considerando-se que o ponto de corte de 30 alunos foi dado à 4ª e/ou 8ª séries e não à escola, podendo, portanto, englobar um número bem superior ao informado.

Diversas pesquisas indicam que os problemas maiores da educação básica brasileira se localizam nas Regiões Norte e Nordeste. Sabe-se que os índices de desigualdade considerando-se as cinco regiões são maiores entre essas regiões e as outras. Pode-se constatar, portanto, que essa desigualdade tende a se perpetuar até mesmo com a avaliação nacional. Pela tabela, pode-se observar que muitas escolas que deveriam estar sendo atingidas pela avaliação não se encontram na população definida. Na Região Norte observa-se que cerca de 78% das escolas que oferecem o ensino fundamental estão localizadas na zona rural. Na Região Nordeste, esse quantitativo é de 68%.

⁹ Dados extraídos do Censo Escolar 2005, tabelas 3.2; 3.9 e 3.13.

Creio que, se ao invés de destinar tantos recursos a uma avaliação nacional que não consegue atingir o seu universo, houvesse políticas que dessem suporte técnico e financeiro aos estados dessas regiões, resultados muito mais produtivos seriam alcançados.

No que diz respeito ao acompanhamento e à co-responsabilidade do processo avaliativo por representantes dos mais variados segmentos sociais, sugere-se que o SAEB constitua pequenos comitês e comissões, com diferentes representantes da área educacional, política e da sociedade em geral. Essas comissões e comitês teriam por missão acompanhar todo o processo avaliativo, ou seja, a elaboração e revisão dos itens, a constituição dos testes, as análises, as publicações e a divulgação dos resultados, pois dessa forma estariam sendo compartilhadas as responsabilidades e o SAEB estaria revelando todo o seu processo.

O estudo também permitiu constatar, comparando o sistema brasileiro ao modelo norte-americano, que o nosso sistema está em condições de igualdade com a metodologia sofisticada de análises utilizada. Observa-se que o SAEB tem exigido das empresas contratadas para realizarem as análises, método de ponta para que retrate com fidelidade a realidade do nosso sistema educacional.

Ainda em relação às análises, algumas diferenças são observadas e talvez possam ser repensadas pelo nosso sistema. A primeira diz respeito à determinação da amostra para os estados que queiram apresentar inferências relativas a eles. Nos Estados Unidos é feito um delineamento para cada estado participante, garantindo representatividade que permita inferir sobre o estado. De modo resumido, os procedimentos são os seguintes: são selecionadas 100 escolas por estado, dessas escolas é determinada uma amostra de 30 estudantes por disciplina e por série sendo que todos os estudantes, que caem na amostra, independente de serem portadores de necessidades especiais ou não, são avaliados. No Brasil isso representaria um aumento da amostra para 486.000 alunos.

Outro ponto referente às análises, e que se sugere adotar aqui, diz respeito às estimativas dos parâmetros que são gerados e separados para os contextos de leitura considerados na matriz de leitura do NAEP, e, em Matemática, aos cinco conteúdos contemplados na matriz. Nos Estados Unidos, este processo é feito após a calibração conjunta dos itens.

Outro aspecto referente às análises, diferente do que é feito aqui, é a atribuição de valores plausíveis que são extraídos da distribuição preditiva dos valores dos escores calibrados para cada estudante. Nos Estados Unidos, desde 1992, as variáveis usadas para

calcular os valores plausíveis para uma escala de avaliação nacional ou para os grupos de escalas incluem um amplo espectro de variáveis de contexto, de atitude e experimentais, e as composições de tais variáveis.

Em relação às publicações, constata-se que o sistema norte-americano preocupa em tornar público, para os diferentes segmentos, os relatos e os documentos que subsidiam o processo avaliativo nacional. Cada um com linguagem e referenciais adequados aos diferentes leitores. Constata-se que o Brasil encontra-se ainda incipiente nessa prática.

Outra pesquisa realizada neste estudo, foi uma análise crítica das avaliações do SAEB, considerando-se as habilidades aferidas e publicadas na forma de escalas e níveis de desempenho. A realização dessa pesquisa implicou a análise de cada matriz referencial para a construção dos testes; a verificação da qualidade pedagógica e psicométrica de cada item e de cada teste; a (re)descrição das escalas e a comparação com as escalas publicadas.

Em relação ao referencial adotado para a construção dos testes, foram analisadas as matrizes de referência de 1997, 1999 e 2001. Confirma-se, como já exposto anteriormente, que há necessidade urgente de uma revisão das matrizes de Língua Portuguesa (Leitura) e Matemática. Constatou-se que há uma distância entre o que é tomado como referencial teórico e o que está descrito na matriz. Os referenciais teóricos das matrizes norte-americana talvez pudessem se constituir em bons parâmetros tanto no que tange à leitura, pois estão consoantes com as capacidades leitoras exigidas nos momentos atuais, além de estar pautada na perspectiva do letramento. Assim, como com a matemática, que adota um referencial pautado na aprendizagem significativa.

Em relação aos itens que compõem os testes que avaliam o desempenho dos nossos alunos, constatam-se problemas recorrentes em todas as edições. Observa-se falta de rigor na construção dos itens. Os problemas mais comuns foram: textos inadequados, itens mais apropriados para situações de seleção, com enunciados não compreensíveis e sem objetividade, alternativas sem plausibilidade e excludentes entre si, conteúdos inadequados para o nível de idade previsto, peguinhas, entre outros. Recomenda-se maior rigor na construção dos itens e na montagem dos testes.

Deve haver, também, uma consonância entre o que está sendo contemplado no teste e o que é previsto como parâmetro curricular nacional e estadual. A questão dessa dissonância entre o currículo praticado e o previsto já foi detectada e tornada pública nos relatórios de resultados do SAEB 1997 e, ao que parece, até hoje nada foi feito. É necessário, dessa forma,

que os professores entendam o que é esperado dos alunos após determinados ciclos de escolarização. E os testes devem refletir esse padrão de aprendizagem esperado. Trata-se de colocar em sintonia a aprendizagem e a avaliação.

É preciso, ainda, ficar atento aos itens que apresentam boa qualidade estatística e que do ponto de vista pedagógico não são adequados. Eles poderão falsear os resultados da avaliação. No modelo norte-americano, além das matrizes de referência, há uma tabela de especificações em que são dadas orientações detalhadas daquilo que se pretende avaliar. Essa pode ser uma boa idéia no sentido de melhorar essa situação.

Em relação às escalas, concluiu-se que elas devem ser mais bem descritas. Elas devem refletir o que os alunos demonstraram ser capazes de fazer no teste como um todo. E por esse motivo, o teste deve ser construído com muito mais cuidado. Pois, quem pode fazer alguma coisa para melhorar o nosso sistema educacional tem que entender a real competência dos alunos, que deve ser retratada fielmente nas escalas de proficiência. Quanto à apresentação das escalas, há inúmeros *layouts* que favoreceriam a compreensão das descrições nas escalas de proficiência.

O trabalho de (re)descrição das escalas, a partir da edição do SAEB 1995, teve essa intenção: mostrar a redução apresentada na descrição das escalas publicadas. Em relação aos níveis de desempenho descritos, certamente a maioria dos educadores concorda que há necessidade urgente de se discutir toda a escala com diferentes atores da sociedade. Devem-se definir os pontos de corte; discutir os qualificativos dos intervalos determinados; demarcar, de acordo com o desempenho dos alunos, o que se pode esperar que eles sejam capazes em cada nível.

Por meio deste estudo, também foram pesquisados os relatórios de resultados do SAEB, referentes a todas as edições. Em relação à essas publicações, considera-se que este seja um dos pontos que deve receber maior atenção, pois elas são o principal produto das avaliações. Não adianta ter resultados válidos e fidedignos, se eles não atingem nenhum público-alvo. O problema das publicações seria resolvido se se adotassem diferentes modelos de apresentação de resultados para variados públicos, com linguagem adequada e resultados compreensíveis.

No período de 17 anos do SAEB, só houve uma publicação pedagógica e, mesmo assim, não exclusivamente voltada para a base, pois ainda conteve vários dados quantitativos, de difícil compreensão. Segundo opinião da pesquisadora, não há transformação que não

ocorra pela base. Portanto, o SAEB não deveria perder isso de vista, já que possui uma fonte de dados que, traduzida, poderia ser um grandioso instrumento de melhoria da qualidade da educação.

Outra pesquisa realizada neste estudo, foi a verificação das políticas públicas implementadas nos estados e no MEC, apoiadas nos resultados do SAEB. Essa pesquisa foi feita, nos estados, com responsáveis pela avaliação, que em sua maioria eram representantes do SAEB e, no MEC, com técnica do Departamento de Políticas de Educação Infantil e Ensino Fundamental, da Secretaria de Educação Básica. Os resultados indicaram que, fundamentadas nos resultados do SAEB, não há notícias de políticas públicas implementadas. Contudo, muitos dos respondentes ressaltaram que o SAEB trouxe grandes avanços na disseminação da cultura avaliativa e que há várias políticas implementadas com base em avaliações próprias dos estados.

Para que o SAEB cumpra o objetivo de oferecer subsídios apoiados em seus resultados para a implementação de políticas públicas nos estados e Brasil, duas ações parecem ser fundamentais: o diálogo com as secretarias de estado e com o MEC, e orientações de ações fundamentadas em seus resultados que possam se reverter em políticas públicas de melhoria da qualidade da educação. O SAEB muitas vezes fornece informações, principalmente as que correlacionam o desempenho com as variáveis de contexto, mas não dá orientações que redundariam em importantes políticas de superação de problemas. Observou-se, ainda, pelo histórico do SAEB que, embora tenha havido algumas iniciativas de proximidade com secretarias, em especial com educadores, o SAEB se mantém como uma instituição insular, que ocasionalmente busca diálogo, ou produz publicações claras a respeito do que permeia o sistema educacional. Sua função tem sido mais a de alardear os insucessos do que orientar ações que poderiam superar as dificuldades constatadas.

De modo geral, considerando-se o estudo dos testes que avaliam o desempenho dos alunos, as análises dos dados, a apresentação dos resultados e as publicações, pode-se concluir que a dimensão da mensuração é fortemente assumida na apresentação dos resultados e, aliada a essa dimensão, a classificação e a pura prestação de contas, sem expressão do julgamento de valor. O alcance de alguns objetivos propostos, como os que relacionam a oferta de subsídios à formulação de políticas públicas, a identificação dos problemas e das diferenças regionais do ensino, a visão clara dos resultados dos processos de ensino e aprendizagem para o SAEB ficam, portanto, comprometidos.

Quanto às questões que nortearam este estudo, considera-se que a pesquisa forneceu subsídios à pesquisadora para inspirar propostas, todas fundamentadas em sua ótica.

A primeira questão foi: “Como deve ser constituído um teste que pretende avaliar o processo de construção do conhecimento dos alunos brasileiros e que informações ele pode oferecer ao sistema?”

A resposta a essa questão fundamenta-se na crença de que esse teste deve ser constituído com base em uma matriz de referência que tenha um marco teórico sólido e adequado para as necessidades prementes do momento. Além disso, devem ser considerados a dimensão do país, os propósitos do ensino e a autonomia dada aos estados no que se refere às orientações curriculares. Levando em conta, ainda, que, por meio do teste pretende-se avaliar o processo de construção do conhecimento dos alunos brasileiros, deve-se considerar a amplitude necessária de conteúdos e de itens que cubram o máximo possível, o *continuum* da escala. O teste deve ser elaborado seguindo todo o rigor que o planejamento e as técnicas de elaboração exigem e, antes e após a pré-testagem, devem passar por processos de revisão e validação por comitês compostos por diferentes tipos de especialistas, que devem dar o aval final.

Garantindo-se a qualidade do teste de desempenho, haverá melhores descrições do que os alunos são capazes de fazer, com conseqüente melhoria do diagnóstico da realidade e orientação mais segura do que poderá ser feito para atingir proficiência adequada dos alunos.

A segunda questão norteadora do estudo relacionou-se à constituição e à descrição de uma escala de proficiência. Ao longo do estudo adotou-se o posicionamento de que a escala deve ser constituída das descrições do que os alunos demonstraram fazer com base no teste aplicado, em intervalos da escala não superiores a meio desvio padrão (25), tomando como referência a média e o desvio da distribuição do desempenho dos alunos da 8ª série, no ano de 1997, ou seja, o valor de 250 para a média e o desvio padrão de 50. Essa condição sendo adotada, as escalas poderão refletir de modo claro, o processo de construção do conhecimento dos alunos, pontuando o que podem ainda construir. A descrição dos níveis de desempenho feita com clareza e objetividade é indispensável porque sinaliza aos professores do que os alunos são capazes e o que ainda poderiam desenvolver ao final de um ciclo de escolarização.

A terceira, quarta e quinta questões dizem respeito à apresentação dos resultados do SAEB, às orientações de políticas públicas e à melhoria da qualidade da educação. O SAEB pode e deve apresentar resultados que sejam úteis aos sistemas educacionais. Como já foi

discutido anteriormente, ele deve produzir relatórios com características adequadas aos diferentes públicos, de forma que todos possam se apropriar dos dados e contribuir para a melhoria da qualidade da educação. No conteúdo de suas publicações deve ter orientações de ações que poderão se transformar em políticas que busquem superar os problemas encontrados.

Neste estudo foi assumida uma pequena parcela do que poderia ser estudado a respeito do SAEB, sob apenas um ponto de vista. Há uma gama de estudos que podem ser realizados, sob diferentes pontos de vista, em especial os que relacionam o desempenho com as variáveis de contexto.

Espera-se que este estudo traga à reflexão a premente necessidade de melhoria das matrizes referenciais, dos testes de desempenho, das descrições das escalas e da apresentação dos resultados. A verdadeira intenção deste estudo foi contribuir com a conseqüente melhoria da qualidade da educação de nosso país. Afinal, avaliar é conhecer e conhecer é transformar (Balarini, 2007).

REFERÊNCIAS BIBLIOGRÁFICAS

- Ackerman, T.A. (1992). A didatic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, 29, 67–91.
- American Psychologist (1992). Ethical Principles of Psychologists and Code of Conduct. *Joint Committee on Standards for Educational Evaluation*, Washington, DC
Vol 47(12) 1597-1611.
- Anastasi, A. e Urbina, S. (2000). *Testagem Psicológica*. Porto Alegre: Artmed
- Andrich, D. (1978) A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Angelo, T.A. (1995). Reassessing (and defining) assessment. *AAHE Bulletin*, 48 (3), 7-9.
- Balarini, M. M. (2007). Afirmção feita em resposta à consulta referente a essa pesquisa, em fevereiro de 2007.
- Barretto, E. S. de S. et al. *Avaliação na educação básica (1990 - 1998)*. Brasília: MEC/Inep, 2001. (Série Estado do Conhecimento, 4)
- Beaton, A. E. & Johnson, E. G. (1992). Overview of the Scaling Methodology Used in the National Assessment. *Journal of Educational Measurement*, V. 29, N. 2, 163-175.
- Beaton, A. E., Johnson, E.G. & Ferris, J. J. (1987). The assignment of exercises to students. In A.E. Beaton, *Implementing the new design: The NAEP 1983 – 1984 Technical Report* (p. 97-118). Princeton, NJ: Educational Testing Service.
- Bloom, B. S. (1956). *Taxonomy of educational objectives : The classification of educational goals. Handbook 1: Cognitive domain*. Nova York: David McKay.
- Bloom, B. S., Hastings, J. T. & Madaus, G. F. (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. McGraw-Hill, Inc. USA.
- Bock, R. D. & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, 35, 179-197.
- Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of items parameters: Aplicação of an EM algorithm. *Psychometrika*, 46, 443-459.
- Bock, R.D., Gibbons, R., e Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, 12, 261-280.
- Boletim informativo do INEP: www.inep.gov.br/imprensa/noticias/outras/news06_09.htm, publicado em 16/05/2006.
- Bonamino, A. & Martínez, S. A. (2002). Diretrizes e Parâmetros Curriculares Nacionais para o ensino fundamental: a participação das instâncias políticas do Estado. *Educação & Sociedade*, v.23, n.80, 371-388.

- Bonamino, A. (2001). *Tempos de avaliação educacional: o SAEB, seus agentes, referências e tendências*. Rio de Janeiro: Quarteto, 2001.
- Bourque, Mary Lin, Ed. (1998). Proceedings of Achievement Levels Workshop – National Assessment Governing Board, September.
- Bourque, Mary Lyn & Bird, Sheila (Eds). (2000). Student Performance Standards on the National Assessment of Educational Progress: Affirmation and Improvements – A study initiated to examine a decade of achievement level setting on NAEP. National Assessment Governing Board, Washington, DC.
- Brasil. Constituição (1988). *Constituição da Republica Federativa do Brasil*. Brasília,DF: Senado Federal.
- Brasil. Lei 9.394, de 20 de dezembro de 1996. Estabelece as Diretrizes e Bases da Educação Nacional. *Diário Oficial da União*. Brasília, DF, v. 134, n. 248, 833-841, 23 dez. 1996.
- Brasil. Secretaria de Educação Fundamental (1997). Parâmetros Curriculares Nacionais: (1ª 4ª Séries): Introdução; Ciências Naturais; Introdução aos Temas Transversais e Ética; Meio Ambiente e Saúde. Brasília: MEC/SEF.
- Cambron-McCabe, N. H. (2002). Educational accountability in the USA: focus on state testing. *Education and the Law*, v. 14, 1-2
- Carmines, E. G. & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.
- Carroll, J. B. (1945). The effect of difficulty and chance success on correlations between tests. *Psychometrika*, v. 10, 1-19.
- Carroll, J.B. (1983). The difficulty of a test and its factor composition revisited. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement: A Festschrift in honor of Frederic M. Lord*. (p. 257-282). Hillsdale, NJ: Erlbaum.
- Castro, M. H. G. (2000). *EFA 2000 Educação para todos: avaliação do ano 2000*. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais.
- Castro, M. H. G., (1998). Avaliação do sistema educacional brasileiro: Tendências e perspectivas. *Ensaio: Avaliação de Políticas Públicas em Educação*, v.6, 303-364.
- Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, v. 40, n.1, 5-32.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, L. R. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.
- Cooper, B. S. (1992). Standard Go Global-The War For Better Schools. *The College Board Review*. 164, 14-21.
- Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. United States of America, Orlando: Harcourt Brace Jovanovich College Publishers.

- Cronbach, L. J. (1963). Course Improvement through Evaluation. In: *Teachers College Record*, pp. 627-683.
- Cronbach, L. J. (1996) *Fundamentos da testagem psicológica*. Porto Alegre: Artes Médicas.
- Cunha, L. A. (1980). *A universidade temporã: o ensino superior da Colônia à era de Vargas*. Rio de Janeiro: Civilização Brasileira, 1980.
- Dietel, R.J., Herman, J.L.& Knuth, R.A. (1991). *What Does Research Say About Assessment?* North Central Regional Education Laboratory, Oak Brook
- Divgi, D.R. (1979). Calculation of the tetrachoric correlation coefficient. *The Psychometric Society*, 44, 169-172.
- Doppen, F. H. and Yeager, E. A. (1998). National versus state curriculum standards for history in the United States: where will the debate lead us? *The Curriculum Journal*, 9, 2, 165-75.
- Dorans, N.J., e Kingston, N.M. (1985). The effects of violations of unidimensionality on the estimation of item and ability parameters and on item response theory equating to the GRE verbal scale. *Journal of Educational Measurement*, 22, 249-262.
- Doyle W. (1977). Learning the classroom environment: An ecological analysis. *Journal of Teacher Education* v. 28, n. 6, 51-55
- Education Week (2002). 'Quality counts 2002. Building blocks for success: State efforts in early childhood education', 10 January. Retirado do site <http://www.edweek.com/> em janeiro/2007.
- Eisner, E. W. (1979) *The Educational Imagination*. New York: MacMillan.
- Eisner, E. W. (1985) *The Art of Educational Evaluation. A Personal View*. London: The Falmer Press.
- Eisner, E. W. (1986). *Procesos Cognitivos y Currículum*. Barcelona: Martinez Roca.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ferrer, A. T. Avaliação e mudança de sistemas educacionais: a interação que falta. In: *Seminário Internacional de Avaliação da Educação*. Rio de Janeiro, 1995. *Anais*. Brasília: MEC, 1996. p.33-54.
- Franco, C. O (2001). Saeb - Sistema de Avaliação da Educação Básica: potencialidades, problemas e desafios. *Revista Brasileira de Educação*, n.17, p.127-133, maio/ago.
- Franco, C. O. (Org) (2001). *Avaliação, Ciclos e Promoção na Educação*. Porto Alegre: Artmed.
- Frary R.B. (1988). Formula scoring of multiple choice tests (correction for guessing): Instructional Topics in *Educational Measurement Educational Measurement: Issues and Practices*, v. 7 n. 2

- Freitas, D.N. T. (2004). Avaliação da educação básica e ação normativa federal. *Cadernos de Pesquisa*, v. 34, n. 123, p. 663-689, set./dez.
- Freitas, L. C. de (org.). (2002). *Avaliação: construindo o campo e a crítica*. Florianópolis: Insular,
- Frisbie, D.A. (1988). Reliability of scores from teacher-made tests. *Instructional Topics in Educational Measurement: Issues and Practices*, v. 7, n.1, 25-35.
- Gatti, B. (2002) Avaliação educacional no Brasil: pontuando uma história de ações. *EccoS Revista Científica*, UNINOVE, São Paulo: v. 4, n. 1: 17-41
- Goldstein, H. (1980). Dimensionality, bias, independence and measurement scale problems in latent trait test score models. *British Journal of Mathematical and Statistical Psychology*, v.33, 234-46.
- Gomes, J. B. F. e Rosenberg, L. (1995). Indicadores de qualidade do ensino e seu papel no sistema nacional de avaliação In: Avaliação Educacional. *Em aberto* n. 66, Brasília: INEP.
- Haberman, S.J. (1977). Log-linear models and frequency tables with small expected cell counts. *Annals of Statistics*, 5, 1148-1169.
- Haladyna, T. M. (1999). *Developing and validating multiple-choice test items* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston: Kluwer-Nijhoff Publishing.
- Hambleton, R. K. e Jones, R. W. (1993): Comparison of Classical Test Theory and Item Response Theory and Their Application to Test Development. *Instructional Topics in Educational Measurement (ITEMS)*, Module 16. Madison, WI: National Council on Measurement in Education (NCME).
- Hambleton, R.K.& Swaminathan H. (1995). *Item Response Theory – Principles and Applications*. Boston: Kluwer . Nijhoff Publishing.
- Hambleton, R.K., e Rovinelli, R. (1986). Assessing the dimensionality of a set of test items. *Applied Psychological Measurement*, 10, 287-302.
- Hambleton, R.K., Swaminathan, H., Cook, L., Eignor, D.R. & Gifford, J.A. (1978). Developments in latent trait theory: Models, technical issues, and applications. *Review of Educational Research*, 48, 467–510.
- Hambleton, R.K., Swaminathan, H., Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, California: Sage Publications.
- Harris, T.L., and R.E. Hodges, eds. (1995). *The Literary Dictionary: The Vocabulary of Reading and Writing*. Newark, DE: International Reading Association.
- Hopkins, K., Stanley, J. & Hopkins, B. R. (1990). *Educational and psychological measurement and evaluation* (7^a ed.). Englewood Cliffs, NJ: Prentice Hall,

- Houston, W. Robert. (1979). *Exploring competency based education*. USA: Board of Regents of the University of Houston.
- White House:: Education: The Promise of America, retirado do *site*: <http://www.whitehouse.gov/infocus/earlychildhood>, acessado em março de 2007, Autor.
- International Association for the Evaluation of Educational Achievement (2000). *Framework and Specifications for PIRLS Assessment 2001*. Chestnut Hill, MA: International Study Center, Boston College.
- Joint Committee on Standards for Educational Evaluation (1981). *Standards for Evaluations of Educational Programs, Projects and Materials*. New York: McGraw-Hill.
- Kaplan, R.M. (1985). The controversy related to the use of psychological tests. In B.B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications* (pp. 465-504). New York: Wiley.
- Kelley, T. L. (1939). The selection of upper and lower groups for the validations of tests items. In: *Journal of Educational Psychology*, 30, pp. 17-24.
- Kerbauy, M. T. M. (2007) Acessado em março/2007, kerbauy@travelnet.com.br.
- Kirsch, I.S., e Jungeblut, A. (1986). *Literacy: Profiles of American's young adults*. Princeton, NJ: Educational Testing Service.
- Kizlik, R. (2006) Measurement, Assessment, and Evaluation in Education. Retirado do *site* <http://www.adprima.com/measurement.htm>, confirmado em março/2007.
- Klein, R. & Fontanive, N. S..(1995) Avaliação em larga escala: uma proposta inovadora. *Em Aberto*, Brasília: v. 15, n. 66, p. 29-34, abr./jun.
- Klein, R. & Ribeiro, S. C. (1991). O censo educacional e o modelo de fluxo: o problema da repetência. *Revista Brasileira de Estatística*, Rio de Janeiro, v. 52, n. 197/198, p. 5-45.
- Klein, R. (2000). *Relatório Saeb 1999, Teoria de Resposta ao Item – TRI*. Fundação Cesgranrio.
- Klein, R. (2002). *Relatório Saeb 2001, Teoria de Resposta ao Item – TRI*. Fundação Cesgranrio.
- Klein, R. (2003). Utilização da Teoria de Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB). *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, vol 11, n. 40, p. 283-296, jul/set 2003.
- Klein, R. (2004). *Relatório Saeb 2003, Teoria de Resposta ao Item – TRI*. Fundação Cesgranrio.
- Klein, R. (2006) Como está a educação no Brasil? O que fazer? *Ensaio: Avaliação de Políticas Públicas Educacionais*. Rio de Janeiro, v.14, n.51, p. 139-172, abr./jun.
- Klein, R. e Klein, T. S. (1998). Programa para extensão do teste Mantel-Haenzel para amostras complexas.

- Koretz, D., & Deibert, E. (1993). *Interpretations of National Assessment of Educational Progress (NAEP) Anchor Points and Achievement Levels by the Print Media in 1991*. Santa Monica, CA: RAND.
- Kratwohl D R, Bloom B S and Masia B B (1964) *Taxonomy of Educational Objectives, the classification of educational goals– Handbook II: Affective Domain* New York: McKay
- Laros, J. A. (2001) *Diferenças entre Estados em Escores Gerais e Sub-escores nos Temas e Competências Cognitivas das Provas do SAEB 1999 em Matemática e Português para a 4ª série do Ensino Fundamental – Relatório Técnico*. Brasília: Centro de Pesquisa em Avaliação Educacional – Universidade de Brasília.
- Laros, J. A., Pasquali, L. e Rodrigues, M. M. M. (2000). *Análise da Unidimensionalidade das provas do SAEB - Relatório Técnico*. Brasília: Centro de Pesquisa em Avaliação Educacional – Universidade de Brasília.
- Le Métails, J. (2003). International Trends in Primary Education, *INCA Thematic Study*, N. 9.
- Linn, R. L., Baker, E.L. e Dunbar, S. B, (1991). Complex, Performance-Based Assessment: Expectations and Validation Criteria. *Educational Researcher*, Vol. 20, No. 8 Nov 1991
- Linn, R. (2001) *The Design and Evaluation of Educational Assessment and Accountability Systems (CSE Technical Report 539)*. CRESST / University of Colorado at Boulder.
- Locatelli, I. (2002), "Construção de Instrumentos para a Avaliação de Larga Escala e Indicadores de Rendimento: O Modelo SAEB". *Estudos em Avaliação Educacional*, 25, pp. 3-21.
- Lord, F.M. (1980). *Applications of item response theory to practical test problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lyon, G. Reid, (1998). Overview of Reading and Literacy Research. In S. Patton & M. Holmes, eds., *The Keys to Literacy*: Washington, DC: Council for Basic Education.
- MacDonald, B. (1971). The Evaluation of the Humanities Curriculum Project : a holistic approach. In: *Theory into Practice*, 10, pp. 163-167.
- Madaus, G. F. & Raczek, A. E. (1996). A turning point for assessment: reform movements in the United States.' In: Little, A. & Wolf, A. (Eds) *Assessment in Transition: Learning, Monitoring and Selection in International Perspective*. Oxford: Elsevier Science Ltd.
- Mager, R. F. (1962). *Preparing objectives for programmed instruction*. San Francisco: Fearon.
- Manual de Elaboração de Itens – Língua Portuguesa (2006). Programa Nova Escola, Rio de Janeiro: CAEd/UFJF.
- Manual de Elaboração de Itens – Matemática (2006). Programa Nova Escola, Rio de Janeiro: CAEd/UFJF.
- Marshall, J. C. & Hales L. W. (1972). *Essentials of testing*. Reading, MA: Addison-Wesley.

- Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika*, v. 47, n. 2, 149-174.
- McDonald, R.P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, v. 34, 100-117.
- McIntire, S. A. & Miller, L. A. (2000). *Foundations of Psychological Testing*. McGraw-Hill: New York.
- Michael, W. B. & Metfessel, N. S. (1967). A Paradigm for Developing Valid Measurable Objectives in the Evaluation of Educational Programs in Colleges and Universities in *Educational and Psychological Measurement*, v. 27: 373-383.
- Mislevy, R. J. & Bock R. D. (1990). BILOG 3 – Item Analysis and Test Scoring with Binary Logistic Models. Scientific Software, Inc.
- Morris, L. L. & Fitzgibbon, C. T. (1978). *How to deal with goals and objectives*. Beverly Hills, CA: Sage.
- Mullis, I.V.S., M.O. Martin, T.A. Smith, R.A. Garden, K.D. Gregory, E.J. Gonzalez, S.J. Chrostowski, & K.M. O’Connor (2001). *TIMSS Assessment Frameworks and Specifications 2003*. Boston, MA: Boston College, The International Study Center.
- Muraki, E. & Bock, R. D. (1997). PARSCALE: IRT based Item Analysis and Test Scoring for Rating-scale Data. Chicago: Scientific Software International, Inc.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, v.16, n.2, 159-176.
- Muraki, E., & Engelhard, G. (1985). Full-information item factor analysis: Applications of EAP scores. *Applied Psychological Measurement*, v. 9, 417-430.
- Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, v.43, n.4, 551–560.
- National Assessment Governing Board (1992). *1994 National Assessment of Educational Progress: Mathematics Assessment Framework and Specifications*. Washington, DC: The College Board and The National Assessment Governing Board.
- National Assessment Governing Board (2003). *Mathematics Framework for the 2003 National Assessment of Educational Progress*. Washington, DC: National Assessment Governing Board.
- National Assessment Governing Board (2003). *Reading Framework for the 2003 National Assessment of Educational Progress*. Washington, DC: National Assessment Governing Board.
- National Assessment Governing Board (2005). *Mathematics Framework for the 2005 National Assessment of Educational Progress*. Washington, DC: National Assessment Governing Board.

- National Assessment Governing Board (2005). *Reading Framework for the 2005 National Assessment of Educational Progress*. Washington, DC: National Assessment Governing Board.
- National Assessment Governing Board (ed.) (2001). *The Nation's Report Card – Reading 2000 grade 4, 8 e 12*.
- National Assessment Governing Board (ed.) (2001). *The Nation's Report Card – Mathematics 2000 grade 4, 8 e 12*.
- National Assessment Governing Board (ed.) (2004). *The Nation's Report Card – Reading 2003 grade 4, 8 e 12*.
- National Assessment Governing Board (ed.) (2004). *The Nation's Report Card – Mathematics 2003 grade 4, 8 e 12*.
- National Assessment Governing Board (ed.) (2006). *The Nation's Report Card – Reading 2005 grade 4, 8 e 12*.
- National Assessment Governing Board (ed.) (2006). *The Nation's Report Card – Mathematics 2005 grade 4, 8 e 12*.
- National Assessment Governing Board (NAGB) (1999). *About NAEP*. Retirado do site: <http://www.nagb.org/>
- National Assessment Governing Board (NAGB) (1999a). *How Does NAEP Work?* Retirado do site: <http://www.nagb.org/>
- National Assessment Governing Board (NAGB) (1999b). *NAEP: Achievement Levels*. Retirado do site: <http://www.nagb.org/>
- National Assessment Governing Board (NAGB) (1999c). *How did NAEP Evolve?* Retirado do site: <http://www.nagb.org/>
- National Center for Education Statistics (NCES) (1996). *The Condition of Education 1995*. Washington, DC: NCES
- National Center for Education Statistics (NCES) (1999). *The Condition of Education 1998*. Washington: NCES.
- National Center for Education Statistics (NCES) (2003). *The Nation's Report Card*. Washington DC: NCES.
- National Center for Education Statistics (NCES) (2003). *The Nation's Report Card*. Washington DC: NCES
- National Center for Education Statistics (NCES)(2000). *The Condition of Education 1999*.
- National Council On Teacher Quality And US Department Of Education, International Affairs Office (2004). OECD: Paris. Retirado do site: <http://www.oecd.org/dataoecd/18/52/33947533.pdf>]

- National Institute for Child Health and Human Development (NICHD) (2000a). *Report of the National Reading Panel*. Washington, DC: Author.
- National Institute for Child Health and Human Development (NICHD) (2000b). *Report of the National Reading Panel: Report of the Subgroups*. Washington, DC: Author.
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric Theory*. 3ª ed. USA: McGraw-Hill.
- Organization for Economic Co-operation and Development (2000). *Measuring Student Knowledge and Skills: The PISA 2000 Assessment of Reading, Mathematical, and Scientific Literacy*. Paris, France: Author.
- Parlett, M. & Hamilton, D. (1972). *Evaluation as Illumination: a new approach to the study of innovatory programmes*. Centre for Research in the Educational Sciences. Edimburg University.
- Pasquali, L. & Primi, R. (2003). Fundamentos da Teoria de Resposta ao Item – TRI. *Avaliação Psicológica*, v.2, n.2, pp. 99-110.
- Pasquali, L. (1996). *Teoria e Métodos de Medida em Ciências do Comportamento*. Brasília: MEC, SEDIAE/INEP e LabPAM.
- Pasquali, L. (1997). *Psicometria: teoria e aplicações*. Brasília: Editora Universidade de Brasília.
- Pasquali, L. (1999). *Instrumentos psicológicos: Manual prático de elaboração*. Brasília: LabPAM / IBAPP
- Pasquali, L. (2001). *Técnicas do exame psicológico: Fundamentos das técnicas psicológicas*. Brasília: LabPAM, Conselho Federal de Psicologia & Casa do Psicólogo.
- Pasquali, L. (2003). *Psicometria: Teoria dos testes na Psicologia e na Educação*. Petrópolis: Vozes.
- Pasquali, L. (2005). *Análise Fatorial para pesquisadores*. Brasília: UnB / Instituto de Psicologia / LabPAM.
- Pasquali, L. (2006). Validade dos testes psicológicos: será possível reencontrar o caminho? *Psicologia: Teoria e Pesquisa*, Edição Especial (Aceito para publicação).
- Pasquali, L. (2007). *Teoria de Resposta ao Item (TRI) Teoria, Procedimentos e Aplicações*. Brasília: LabPAM.
- Perrenoud, Philippe. (1993). *Práticas pedagógicas, profissão docente e formação: perspectivas sociológicas*. Lisboa: Dom Quixote.
- Pestana, M. I. (1998). O sistema de avaliação brasileiro, in *Revista Brasileira de Estudos Pedagógicos - RBEP*, Vol. 79.
- Pestana, M. I. G. S. (1998) O sistema de avaliação brasileiro. In: *Anais do Seminário Internacional de Avaliação Educacional*. Brasília, DF: INEP, 1998. p. 15-30.

- Pestana, M. I. G. S. et al. (1997). *Matrizes Curriculares de Referência para o SAEB*. Brasília: Instituto Nacional de Estudos e Pesquisas.
- Pestana, M. I. G. S. et al. (1999). *Matrizes Curriculares de Referência para o SAEB*. 2. ed. Rev. Ampl. Brasília: Instituto Nacional de Estudos e Pesquisas.
- Pestana, M. I. G. S. et al. (1999). *SAEB 97: Primeiros resultados*. Brasília: INEP
- PISA para Docentes. (2005). *La Evaluación como oportunidad de aprendizaje*. México: Instituto Nacional para la Evaluación de la Educación.
- Popham, J. W. (1969). Objectives instruction. *AERA Monograph Series on Curriculum Evaluation*. Vol. 3
- Popham, J. W. (1975) *Educational Evaluation*. USA: New Jersey: Prentice Hall.
- Qualidade da educação: uma nova leitura do desempenho dos estudantes da 4ª série do Ensino Fundamental. (2003). Brasília: INEP.
- Ribeiro, S. C. (1991). A pedagogia da repetência. *Estudos Avançados*, São Paulo, v. 5, n. 12, maio/ago.
- Rodrigues, M. M. M. (2002). *Instrumentos de Avaliação Educacional: Uma Visão Pedagógica e Psicométrica Integradas – Estudo das provas de Matemática – 8ª série – 1997 e 1999*. Dissertação de Mestrado, Instituto de Psicologia, Universidade de Brasília.
- Rodrigues, M. M. M. (2006). Proposta de análise de itens das provas do SAEB sob a perspectiva pedagógica e psicométrica. *Estudos em Avaliação Educacional*, v. 17, n. 34, mai/ago 2006.
- SAEB 1995 – Resumo Executivo: SAEB/95. (1997). Brasília: INEP.
- SAEB 1999 – Relatório Nacional SAEB/99. Retirado do site: http://www.inep.gov.br/basica/saeb/saeb_99.htm.
- SAEB 2001 – *Novas Perspectivas (2002)* Brasília: Ministério da Educação / Instituto Nacional de Estudos e Pesquisas / Diretoria de Avaliação da Educação Básica Brasileira.
- SAEB 2001 – *Relatório Língua Portuguesa: Saeb 2001 (2002)*. Brasília: Ministério da Educação/ Instituto Nacional de Estudos e Pesquisas / Diretoria de Avaliação da Educação Básica Brasileira.
- SAEB 2001 – *Relatório Matemática: Saeb 2001 (2002)*. Brasília: Ministério da Educação/ Instituto Nacional de Estudos e Pesquisas / Diretoria de Avaliação da Educação Básica Brasileira.
- SAEB 2001 – *Relatório Nacional: Saeb 2001 (2002)*. Brasília: Ministério da Educação/ Instituto Nacional de Estudos e Pesquisas / Diretoria de Avaliação da Educação Básica Brasileira.
- SAEB 2003 (2004). *Guia de Elaboração de Itens: Língua Portuguesa*. Brasília: MEC/INEP.

- SAEB 2003 (2004). *Guia de Elaboração de Itens: Matemática*. Brasília: MEC/INEP.
- Samejima, F. (1997). Graded response model. In W. J. Van der Linders & R. K. Hambleton (Orgs.), *Handbook of Modern Item Response Theory* (pp. 85-100). New York: Springer
- Schwartz, Steven. (1984). *Measuring Reading Competence – A Theoretical-Prescriptive Approach*. USA: New York: Plenum Press.
- Scriven, M. (1967). The methodology of evaluation. *AERA Monograph Series on Curriculum Evaluation*. Vol. 1. Rand McNally, Chicago
- Scriven, M. (1986). Evaluation as a paradigm for education research. In E. House (Ed.), *New directions in education evaluation* (pp. 53 - 67). London: Falmer Press.
- Snow, C.E., Burns S. & Griffin, P., eds. (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.
- Souza, S. Z. L. (1997). Avaliação do rendimento escolar como instrumento de gestão educacional. In: OLIVEIRA, D. A. (org.) *Gestão democrática da educação: desafios contemporâneos*. Petrópolis: Vozes, pp.264-283.
- Stake, R. E. (1975). *Evaluating the arts in Education: A Responsiveness Approach*. Ohio: Merrill, Columbus.
- Standards for Evaluations of Educational Programs, Projects, and Materials (Paperback). (1981). USA: McGraw-Hill Book Company.
- Taba, H. (1962) *Curriculum Development: Theory and practice*. New York: Harcourt Brace and World.
- Talmage, H. (1982). Evaluation of Programs. In H. E. Mitzel (ed) *Encyclopedia of Educational Research*, 5th ed. New York: The Free Press.
- Technical Report: The NAEP 1998 (2001, August), edited by National Assessment Governing Board.
- The Education (National Curriculum) (Foundation Stage Early Learning Goals) (England) Order 2003 (*Statutory Instrument 2003 No.391*)
- The Education (National Curriculum) (Foundation Stage Profile Assessment Arrangements) (England) Order 2003 (*Statutory Instrument 2003 No.1327*)
- Thurstone, L.L. (1947). *Multiple factor analysis: A development and expansion of The Vectors of Mind*. Chicago, IL: University of Chicago Press.
- Travers, R. M. W. (1955). *Educational Measurement*. New York: The McMillan Company,
- Travers, R. M. W. (1983). *How research has changed American schools*. Kalamazoo, MI: Mythos Press.
- Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.

- Tyler, R. W. (1969). *The Year Book Committee and Associated Contributors*. USA, Chicago, Illinois: National Society for the Study of Education.
- Tyler, R. W. (1975). Specific Approaches to Curriculum Development. In: Giroux, H., Penna, A. e Pinar, W.: *Curriculum and Instruction*. Berkeley: McCutchan.
- Tyler, R. W. (1981). *Princípios básicos de currículo e ensino* (Trad.). Porto Alegre: Editora Globo.
- United States Department Of Education (1999). *Charter Schools Taking Root as Public School Choice Continues to Expand*. Retirado do site <http://www.ed.gov/PressReleases/>, em jan/2007
- United States Department of Education (1999). *Educational Excellence for All Children Act of 1999: Fact Sheet*.
- United States Department Of Education (1999a). *The Re-authorisation of the Elementary and Secondary Education Act (ESEA)*.
- United States Department of Education (1999b). *Educational Excellence for All Children Act of 1999: Fact Sheet*.
- United States Department of Education (2000). *Class-Size Reduction Programme*. Retirado do site: <http://www.ed.gov/offices/OESE/ClassSize>, em jan/2007
- United States Department of Education (2003). *Paige Announces Grant to Improve Rural Education* (Press Release, 13 November). Retirado do site: <http://www.ed.gov/> em dezembro de 2006.
- United States Department of Education (2003). *Testing: Standardised Admissions Tests*.
- United States Department of Education (2004). *New Website will Help Communities Develop Character Education Programs* (Press Release, 17 February). Retirado do site: <http://www.ed.gov/>, em março de 2007.
- United States Department Of Education, National Center For Education Statistics (NCES) (2003). *Teaching Mathematics in Seven Countries: Results from the TIMSS 1999 Video Study*. Washington DC: NCES.
- United States Government. Federal Statutes (1994). *The Goals 2000: Educate America Act*. (The 1994 re-authorisation of *ESEA*.)
- United States Government. Federal Statutes (1999). *The Educational Excellence for All Children Act*. (The 1999 re-authorisation of *ESEA*.)
- United States Government. Federal Statutes (2002). *No Child Left Behind*. (The 2001 re-authorisation of *ESEA*.)
- United States Government. Federal Statutes. *Individuals with Disabilities Education Act*.
- United States Government. Federal Statutes. *The Elementary and Secondary Education Act (ESEA)*.

- United States Of America. National Education Goals Panel (1995). *Education Goals Report: Executive Summary. Improving Education Through Family-School-Community Partnerships*. Washington: National Education Goals Panel.
- USA. National Education Goals Panel (1995). *Education Goals Report: Executive Summary. Improving Education Through Family-School-Community Partnerships*. Washington: National Education Goals Panel.
- van Batenburg, T. A. e Laros, J. A. (2001). “Graphical Analysis of Test Items”. In: *Educational Research and Evaluation (An International Journal on Theory and Practice)*. Lisse: Swets e Zeitlinger.
- van der Linden, W. J. e Hambleton, R.K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer –Verlag.
- Vianna, H. M. (1982) *Testes em Educação*. São Paulo: Ibrasa.
- Vianna, H.M. (1980) A perspectiva das medidas referenciadas a critério. *Educação e Seleção*, São Paulo, Fundação Carlos Chagas, (2): 5-14, dez., 1980.
- Vianna, H.M. (1984) Qualificação técnica e construção de instrumentos de medida. *Educação e Seleção*, São Paulo. Fundação Carlos Chagas, (10): 43-49, jul./dez., 1984.
- Vianna, H.M. (1985). Programa de medidas para fins de ensino e orientação. *Educação e Seleção*, São Paulo. Fundação Carlos Chagas, (11): 87-95, jan./jun., 1985.
- Vianna, H.M. (1987). *Testes em Educação*. São Paulo: Ibrasa.
- Vianna, H.M. (1989). *Introdução à avaliação educacional*. São Paulo, IBRASA.
- Wang, X.B.(2006). An introduction to the system and culture of the college entrance examinations of China. *The College Board in Quarterly Newsletter – Vol 14 n. 3* National Council on Measurement in Education - NCME.
- Wiersma W & Jurs S. (1990). *Educational measurement and testing*. 2nd ed. Boston: Allyn and Bacon.
- Wilson, D.T., Wood, R. & Gibbons, R. (1991). *TESTFACT: Test scoring, item statistics, and item factor analysis*. Chicago: Scientific Software International.
- Worthen, B. R. & Sanders, J. M. (1987). *Educational Evaluation: A Alternative Approaches and Practical Guidelines*. New York and London: Longman.
- Young, M. F. D.. (2002). Durkheim, Vygotsky e o currículo do futuro. *Cadernos de Pesquisa*, São Paulo, n. 117.
- Zimowski, M. F., Muraki, E., Mislevy, R. J. & Bock, R. D.(1996). BILOG-MG - Multiple-group IRT Analysis and Test Maintenance for binary items. Scientific Software International, Inc.
- Zwick, R. (1985). *Assessing the dimensionality of NAEP year 15 reading data* (ETS Res. Rep. No. 86-4). Princeton: Educational Testing Service.

Zwick, R. (1986). *Assessment of the dimensionality of NAEP year 15 reading data*. Princeton, NJ: Educational Testing Service (ERIC Document Reproduction No. ED269420 TS).

Zwick, R. (1987). Assessing the dimensionality of NAEP reading data. *Journal of Educational Measurement*, 24, 293-308.

APÊNDICE

A1 – Modelo de análise pedagógica dos itens

Quadro 1 – Análise pedagógica de item

Disciplina: Língua Portuguesa – 4ª série – Ensino Fundamental – SAEB – Edição 2001													
Item 01 Tópico: Coerência e Coesão no Processamento do Texto													
Descritor avaliado: Estabelecer relação causa/conseqüência entre partes e elementos do texto.													
Aspectos de forma													
Texto			Enunciado						Alternativas				
Adequação ao ciclo	Adequação do tema	Erros digitação	Redação	Clareza	Adequação ciclo	Expressões	Termos	Gramática	Gabarito correto	Coer. alt e enunciado	Redação	Atração	Clareza e objetividade
Sim	Sim	Não	Boa	Sim	Sim	Boas	Bons	Coerente	Sim	Sim	Boa	Não	Sim
Aspectos do conteúdo													
Texto	Enunciado					Alternativas							
Pertinência	Correspondência enunciado / competência		Contextualização	Situação factível	Raciocínio	Conceitos universais	Pegadinha		Coerência com enunciado		Plausibilidade		
Sim	Sim		Sim	Poema	Sim	Sim	Não		Sim		Sim		
Operação mental possível													
Altern. A: A opção por essa resposta pode ter sido feita pelos seguintes motivos: falta de atenção; pelo fato de o aluno não ter nenhum domínio da capacidade leitora; por ter sido a primeira informação que apareceu no poema.													
Altern. B: um dos motivos da escolha por essa alternativa pode ter sido, também, a falta de atenção; outro motivo pode ter sido o de aluno ter pensado que ao dar susto nos outros, seria incoerente dar um abraço apertado nos outros. E, ainda pode ter optado por essa informação estar também explícita, como da alternativa anterior.													
Altern. C: Essa opção pode ter sido feita pelos alunos que imaginam que o fato de o porquinho ter namorada o impedia de dar abraço apertado nos outros.													
Altern. D: A opção correta foi escolhida por 71% dos alunos, podendo-se inferir que esses alunos conseguem estabelecer uma relação de causa / conseqüência entre partes do texto. Para chegar à resposta correta seria necessário que o aluno lesse e compreendesse o poema para chegar à conclusão de que o porquinho não pode dar um abraço apertadinho porque ele é um porco-espinho.													
Observações													

A2 – Modelo de análise da qualidade psicométrica (itens selecionados)

Tabela 1 - Parâmetros psicométricos dos itens selecionados para descrição do Teste LP/1997

Língua Portuguesa – Edição 1997						
Item	Param a	Param b	Param c	Dificuld.	Discrim.	Bisserial
3	1.52	-2.49	0.12	0.79	0.48	0.65
5	1.59	-0.99	0.2	0.47	0.5	0.49
6	0.91	-2.14	0.01	0.63	0.53	0.57
10	1.39	-1.13	0.05	0.44	0.61	0.61
13	1.49	-0.34	0.17	0.29	0.31	0.38
17	1.01	-1.27	0.04	0.49	0.56	0.56
25	1.1	-0.9	0.01	0.34	0.47	0.53
29	1.64	-1.84	0.23	0.69	0.56	0.62
30	1.53	-0.32	0.16	0.36	0.44	0.51
32	1.04	-0.05	0.2	0.38	0.43	0.48
34	1.63	-0.99	0.18	0.46	0.63	0.62
35	0.77	-0.92	0	0.44	0.52	0.51
36	0.58	-3.24	0.03	0.75	0.35	0.45
37	0.71	-2.79	0.01	0.71	0.45	0.52
38	1.48	-1.43	0.04	0.5	0.71	0.68
39	1.24	-1.08	0.07	0.44	0.64	0.63
41	1.32	-1.42	0.26	0.6	0.57	0.56
42	0.61	-3.19	0.07	0.77	0.37	0.52
43	1.26	-1.28	0.16	0.46	0.53	0.54
44	1.63	-1.62	0.17	0.55	0.63	0.61
45	1.35	-2.93	0.01	0.81	0.43	0.6
47	2.48	-0.59	0.26	0.44	0.57	0.59
48	1.17	-2.47	0.01	0.71	0.49	0.58
49	1.82	-2.17	0.08	0.72	0.58	0.69
50	1.53	-0.15	0.25	0.46	0.48	0.48
51	1.37	-2.37	0.09	0.76	0.53	0.66
52	1.66	-1.3	0.18	0.53	0.64	0.63
53	1.17	-1.92	0.08	0.65	0.63	0.66
54	1.38	-1.82	0.12	0.64	0.59	0.63
55	0.85	-0.77	0.02	0.38	0.47	0.53
57	1.11	-0.49	0.28	0.49	0.41	0.42
60	1.45	-1.59	0.28	0.66	0.52	0.56
63	1.94	-0.56	0.14	0.35	0.54	0.58
68	2.13	-1.5	0.2	0.61	0.63	0.65
69	1.9	-0.34	0.19	0.36	0.47	0.52
73	2.56	-0.29	0.1	0.25	0.48	0.62
76	2.18	-1.46	0.34	0.65	0.56	0.59
87	1.79	-0.22	0.27	0.41	0.44	0.44
89	2.59	-0.91	0.4	0.6	0.51	0.51
90	2.18	-0.49	0.3	0.46	0.49	0.48
91	2.28	-0.24	0.22	0.36	0.44	0.51
92	1.77	-0.16	0.21	0.36	0.43	0.48
93	1.3	-1.32	0.13	0.54	0.59	0.57
94	0.88	-1.16	0.38	0.65	0.36	0.38
95	1.68	-0.37	0.19	0.38	0.48	0.49
97	1.05	-1.98	0.13	0.65	0.48	0.52
101	1.29	-1.92	0.05	0.62	0.57	0.58
103	1.16	-1.01	0.09	0.45	0.54	0.56
111	1.17	-0.54	0.18	0.42	0.47	0.49
116	0.97	-0.86	0.19	0.48	0.44	0.44
125	1.26	-1.22	0.19	0.55	0.54	0.54
130	1.32	-1.76	0.04	0.58	0.63	0.62
143	2.02	-0.09	0.23	0.34	0.37	0.44

A3 – Modelo de descrição das habilidades dos itens por nível, por disciplina e por edição

SAEB – Edição 1997 - Matemática

Nível 175

- 034 – Indicam o valor de uma incógnita em uma expressão, dadas algumas opções;
- 046 – Reconhecem fração em desenho representando parte-todo;
- 057 – Resolvem um problema simples, envolvendo uma multiplicação;
- 064 – Identificam um círculo entre diversas formas arredondadas;
- 072 – Identificam linhas poligonais abertas e fechadas;
- 076 – Realizam operação de adição envolvendo cédulas e moedas do Sistema Monetário Nacional
- 077 – Identificam a representação de figuras tridimensionais simples pelas formas arredondadas (cone e cilindro);
- 081 – Resolvem problema envolvendo multiplicação de decimais (valor monetário 0,10 por 25)
- 082 – Resolvem um problema envolvendo uma subtração com o significado de alteração do estado inicial;
- 083 – Resolve problema aplicando a invariância da diferença;
- 096 – Calculam o tempo que falta para a realização de um evento com hora determinada;
- 097 – Resolvem problema envolvendo subtração;
- 100 – Lêem e interpretam horários em relógios de ponteiros e digitais;
- 104 – Identificam objeto que se encontra entre duas posições;
- 112 – Resolvem problema envolvendo a facilitação de troco em situação de compra;
- 113 – Decompõem número e o representam através da escrita;
- 121 – Inferem que para resolver um problema é necessário proceder a uma divisão;
- 135 – Decompõem número e o representam através da escrita;
- 137 – Ordenam em ordem crescente numerais decimais;
- 143 – Inferem o que é menor pelo método de comparação;

Nível 250

- 007 – Identificam número de dias que tem uma semana e calculam o número de dias de 13 semanas;
- 010 – Aplicam a propriedade do elemento neutro na multiplicação para resolver uma situação- problema;
- 019 – Resolvem uma situação-problema envolvendo relação de proporção;
- 020 – Identificam, pelo sistema de numeração, o valor posicional de um algarismo na construção da escrita de um número;
- 027 – Inferem que dois pontos em um círculo estão à mesma distância de um ponto central;
- 029 – Resolvem situações-problema utilizando procedimentos de cálculo mental aproximado, por meio de estratégias pessoais;
- 031 – Interpretam dados representados em um gráfico de setores;
- 033 – Resolvem uma situação-problema envolvendo relação de proporção;
- 035 – Identificam quantas horas tem um dia, quantos dias tem a semana e calculam quantas horas tem uma semana para solucionarem um problema;
- 036 – Estabelecem regras para seqüência de números correlacionados em duas colunas;
- 038 – Resolvem uma expressão numérica, em que se tem que substituir os valores representados por figuras, envolvendo adição, subtração e multiplicação;
- 039 – Estabelecem relação de correspondência entre os dados de uma tabela e um gráfico;
- 040 – Dados quatro algarismos, formam o maior número considerando-se a ordem de grandeza e o valor posicional;
- 041 – Lêem um gráfico de colunas, interpretam e calculam o resultado da soma de todas as colunas;
- 045 – Decompõem número natural em sua forma polinomial (dezena de milhar);
- 047 – Armam e efetuam uma subtração, com reserva;
- 048 – Calculam o resultado de uma divisão de números naturais;
- 049 – Armam e efetuam uma subtração com reserva e com um zero intercalado no algarismo referente ao subtraendo;
- 050 – Calculam o resultado e o resto de uma divisão de um algarismo da ordem das centenas por um da ordem das dezenas;
- 051 – Realizam trocas entre cédulas e moedas;
- 052 – Estabelecem trocas de moedas em função do valor;

- 054 – Calculam o resultado de multiplicações por 10 e por 100;
- 055 – Calculam o resultado de uma subtração de frações com o mesmo denominador;
- 056 – Representam um número natural na reta numérica;
- 058 – Resolvem um problema aplicando a invariância da diferença;
- 060 – Representam relações espaciais utilizando os elementos posicionais para frente / para baixo;
- 086 – Aplicam a propriedade comutativa da multiplicação para resolver problema;
- 088 – Resolvem problema simples envolvendo operações de multiplicação e adição;
- 093 – Interpretam dados representados em gráfico de setores;
- 095 – Identificam a representação de um número decimal na reta numérica;
- 101 – Resolvem problema envolvendo cálculo de uma adição e de uma subtração;
- 103 – Armam uma expressão numérica para resolver um problema;
- 105 – Identificam entre diversas medidas de mesma grandeza a mais apropriada para medir a altura de um prédio;
- 106 – Decompõem número escrito por extenso na forma polinomial;
- 111 – Identificam a representação de um número decimal na reta numérica;
- 117 – Identificam um objeto entre vários pela posição indicada no enunciado (esquerda e abaixo de um outro objeto);
- 122 – Identificam, pelo sistema de numeração, o valor posicional de um algarismo na construção da escrita de um número;
- 123 – Comparam leituras realizadas em termômetros situados em ambientes diferentes, reconhecendo as diferenças indicadas;
- 127 – Identificam a posição de um número dados dois referenciais;
- 133 – Calculam o número de minutos de um dia;
- 134 – Resolvem situações-problema envolvendo operações que utilizam cédulas e moedas;
- 136 – Identificam elemento em uma ilustração gráfica seguindo referência da posição (segunda criança da esquerda para a direita);
- 138 – Inferem que devem realizar uma operação de adição para solucionar um problema;

Nível 325

- 004 – Identificam número de vértices de uma figura;
- 006 – Comparam e ordenam números racionais na forma decimal;
- 014 – Comparam grandezas de mesma natureza utilizando unidades não convencionais;
- 015 – Identificam a escrita de uma fração na forma de um número decimal;
- 030 – Comparam grandezas de mesma natureza, utilizando cálculos com unidades convencionais;
- 044 – Reconhecem o significado de frações como quocientes;
- 053 – Armam e efetuam uma multiplicação de número da ordem das centenas, sendo que um deles possui um zero intermediário;
- 059 – Identificam a forma de um cilindro;
- 061 – Comparam grandezas utilizando unidades não convencionais;
- 062 – Aplicam a propriedade distributiva da multiplicação para resolver problema;
- 073 – Comparam unidades de tempo (dias, semanas e meses);
- 074 – Identificam características próprias do retângulo e do quadrado;
- 078 – Interpretam um croqui utilizando elementos posicionais (frente e esquerda);
- 085 – Comparam e convertem horas e minutos em minutos;
- 089 – Interpretam a representação da movimentação do girando para frente de quem está olhando e na direção esquerda da figura;
- 094 – Transformam medida de quilômetros para metros;
- 098 – Interpretam dados representados em um gráfico de colunas;
- 099 – Resolvem problema envolvendo situação comparativa com existência de um elemento neutro;
- 107 – Identificam semelhanças entre duas figuras tridimensionais pelo número de faces;
- 119 – Identificam o vértice de uma figura plana;
- 126 – Identificam a posição representada de uma imagem espelhada;
- 141 – Reconhecem dois segmentos como paralelos.

A4 – Consulta feita às Secretarias de Educação de Estado, a respeito de políticas públicas implementadas com base nos resultados do SAEB

*Prezado(a) Gestor(a) de Avaliação da Secretaria de Estado de Educação do Estado
XXXX*

Sou aluna do Curso de Doutorado do Instituto de Psicologia da Universidade de Brasília e realizo pesquisa sobre “Testes de desempenho em avaliações em larga escala”. Minha pesquisa é realizada com o apoio do CNPq. Utilizo os dados das avaliações do Sistema Nacional de Avaliação da Educação Básica (SAEB), cedidos pelo Instituto Nacional de Estudos e Pesquisas em Educação (INEP), relativos às edições 1995 a 2003.

Neste momento, necessito de sua colaboração no sentido de me informar se o seu Estado implementou alguma política pública tendo por base os resultados do SAEB no período 1995 a 2003. Se houve a implementação de alguma política, solicitaria que me informasse o conteúdo da política adotada.

As questões que procuro responder por meio de minha pesquisa são:

- O que pode ser feito para melhorar o processo de avaliação da educação básica brasileira?*
- Como deve ser elaborado um teste que avalia o desempenho de um sistema?*
- Como as análises pedagógica e psicométrica dos itens podem se tornar úteis no relato dos resultados?*
- Como devem ser relatados os resultados das avaliações e a quais públicos?*
- Que contribuições o SAEB pode fornecer para a adoção de políticas públicas que melhorem a qualidade da educação básica brasileira?*

Considerando que estou próxima do término de meu trabalho, apelo para sua boa vontade em me enviar essas informações com a máxima urgência.

Certa de poder contar com sua contribuição, agradeço sua atenção,

*Margarida Maria Mariano Rodrigues
Aluna do Curso de Doutorado do Instituto de Psicologia da Universidade de Brasília
Matrícula 03/51326*

ANEXO



Anexo 1 – Matrizes Curriculares do SAEB

A – Língua Portuguesa

Quadro 1 - Tópico I – Procedimentos de Leitura

DESCRITORES	Séries em que se aplicam		
	4ª EF	8ª EF	3ª EM
Localizar informações explícitas em um texto.	D 1	D 1	D 1
Inferir o sentido de uma palavra ou expressão.	D 3	D 3	D 3
Inferir uma informação implícita em um texto.	D 4	D 4	D 4
Identificar o tema de um texto.	D 6	D 6	D 6
Distinguir um fato da opinião relativa a esse fato.	D 11	D 14	D 14

Quadro 2 - Tópico II – Implicações do suporte, do gênero e/ou enunciador na compreensão do texto

DESCRITORES	Séries em que se aplicam		
	4ª EF	8ª EF	3ª EM
Interpretar texto com auxílio gráfico diverso (propagandas, quadrinhos, fotos etc.)	D 5	D 5	D 5
Identificar a finalidade de textos de diferentes gêneros.	D 9	D 12	D 12

Quadro 3 - Tópico III – Relação entre textos

DESCRITORES	Séries em que se aplicam		
	4ª EF	8ª EF	3ª EM
Reconhecer diferentes formas de tratar uma informação na comparação de textos que tratam do mesmo tema, em função das condições em que eles foram produzidos e daquelas em que serão recebidos.	D 15	D 20	D 20
Reconhecer posições distintas entre duas ou mais opiniões relativas ao mesmo fato ou ao mesmo tema.	-	D 21	D 21

Quadro 4 - Tópico IV – Coerência e coesão no processamento do texto

DESCRITORES	Séries em que se aplicam		
	4ª EF	8ª EF	3ª EM
Estabelecer relações entre as partes de um texto, identificando repetições ou substituições que contribuem para a continuidade do texto.	D 2	D 2	D 2
Identificar o conflito gerador do enredo e os elementos que constroem a narrativa.	D 7	D 10	D 10
Estabelecer relação causa/consequência entre partes e elementos do texto.	D 8	D 12	D 12
Estabelecer relações lógico discursivas presentes no texto, marcadas por conjunções, advérbios etc.	D12	D 15	D 15
Identificar a tese de um texto.	-	D 7	D 7
Estabelecer relação entre a tese e os argumentos oferecidos para sustentá-la.	-	D 8	D 8
Diferenciar as partes principais das secundárias em um texto.	-	D 9	D 9

Quadro 5– Tópico V – Relação entre recursos expressivos e efeitos de sentido

DESCRITORES	Séries em que se aplicam		
	4ª EF	8ª EF	3ª EM
Identificar efeitos de ironia ou humor em textos variados.	D 13	D 16	D 16
Reconhecer o efeito de sentido decorrente do uso da pontuação e de outras notações.	D 14	D17	D17
Reconhecer o efeito de sentido decorrente do uso da pontuação e de outras notações.	-	D 18	D 18
Reconhecer o efeito de sentido decorrente da exploração de recursos ortográficos e/ou morfosintáticos.	-	D 19	D 19

Quadro 6 – Tópico VI – Variação lingüística

Identificar as marcas lingüísticas que evidenciam o locutor e o interlocutor de um texto.	D 10	D 13	D 13

B – Matemática - SAEB

Quadro 7 – Matrizes de Matemática referentes aos três ciclos, considerando-se os temas

TEMA	4ª SÉRIE	8ª SÉRIE	3ª SÉRIE
1. ESPAÇO E FORMA	<p>D1 Identificar a localização/movimentação de objetos em mapas, croquis e outras representações gráficas.</p> <p>D2 Identificar propriedades comuns e diferenças entre poliedros e corpos redondos, relacionando figuras tridimensionais com suas planificações.</p> <p>D3 Identificar propriedades comuns e diferenças entre figuras bidimensionais pelo número de lados, pelos tipos de ângulos.</p> <p>D4 Identificar quadriláteros observando as posições relativas entre seus lados (paralelos, concorrentes, perpendiculares).</p> <p>D5 Reconhecer a conservação ou modificação de medidas dos lados, do perímetro, da área em ampliação e/ou redução de figuras poligonais usando malhas quadriculadas.</p>	<p>D1 Identificar a localização/movimentação de objetos em mapas, croquis e outras representações gráficas.</p> <p>D2 Identificar propriedades comuns e diferenças entre poliedros e corpos redondos, relacionando figuras tridimensionais com suas planificações.</p> <p>D3 Identificar propriedade de triângulos pela comparação de medidas de lados e ângulos.</p> <p>D4 Identificar relação entre quadriláteros por meio de suas propriedades.</p> <p>D5 Reconhecer a conservação ou modificação de medidas dos lados, do perímetro, da área em aplicação e/ou redução de figuras poligonais usando malhas quadriculadas.</p> <p>D6 Reconhecer ângulos como mudanças de direção ou giros, identificando ângulos retos e não retos.</p> <p>D7 Reconhecer que as imagens de uma figura construída por uma transformação homotética são semelhantes, identificando propriedades e/ou medidas que se modificam ou não se alteram.</p> <p>D8 Resolver problema utilizando propriedades dos polígonos (soma de seus ângulos internos, números de diagonais, cálculo da medida de cada ângulo interno nos polígonos regulares).</p> <p>D9 Interpretar informações apresentadas por meio de coordenadas cartesianas.</p> <p>D10 Utilizar relações métricas do triângulo retângulo para resolver problemas significativos.</p> <p>D11 Reconhecer círculo/circunferência, seus elementos e algumas de suas relações.</p>	<p>D1 Identificar figuras semelhantes mediante o reconhecimento de relações de proporcionalidade.</p> <p>D2 Reconhecer aplicações das relações métricas do triângulo retângulo em um problema que envolva figuras planas ou espaciais.</p> <p>D3 Relacionar diferentes poliedros ou corpos redondos com suas planificações ou vistas.</p> <p>D4 Identificar a relação entre o número de vértices, faces e/ou arestas de poliedros expressa em um problema.</p> <p>D5 Resolver problema que envolve razões trigonométricas no triângulo retângulo (seno, cosseno, tangente).</p> <p>D6 Identificar a localização de pontos no plano cartesiano.</p> <p>D7 Interpretar geometricamente os coeficientes da equação de uma reta.</p> <p>D8 Identificar a equação de uma reta apresentada a partir de dois pontos dados ou de um ponto e sua inclinação.</p> <p>D9 Relacionar a determinação do ponto de interseção de duas ou mais retas com a resolução de um sistema de equação com duas incógnitas.</p> <p>D10 Reconhecer, dentre as equações do 2º grau com duas incógnitas, as que representam circunferência.</p>
2. GRANDEZAS E MEDIDAS	<p>D6 Estimar a medida de grandezas utilizando unidades de medida convencionais ou não.</p> <p>D7 Resolver problemas significativos utilizando unidades de medida padronizadas como Km/m/cm/mm, Kg/g/mg, l/ml.</p> <p>D8 Estabelecer relações entre unidades de medida de tempo.</p> <p>D9 Estabelecer relações entre o horário de início e término e/ou o intervalo da duração de um evento ou acontecimento.</p> <p>D10 Num problema, estabelecer trocas entre cédulas e moedas do sistema monetário brasileiro, em função de seus valores.</p>	<p>D12 Resolver problema envolvendo o cálculo de perímetro de figuras planas.</p> <p>D13 Resolver problema envolvendo o cálculo de área de figuras planas.</p> <p>D14 Resolver problema envolvendo noções de volume.</p> <p>D15 Resolver problema envolvendo relações entre diferentes unidades de medida.</p>	<p>D11 Resolver problema envolvendo o cálculo de perímetro de figuras planas.</p> <p>D12 Resolver problema envolvendo o cálculo de área de figuras planas.</p> <p>D13 Resolver problema envolvendo a área total e/ou volume de um sólido (prisma, pirâmide, cilindro, cone, esfera).</p>

TEMA	4ª SÉRIE	8ª SÉRIE	3ª SÉRIE
	<p>D11 Resolver problemas envolvendo o cálculo do perímetro de figuras planas, desenhadas em malhas quadrangulares.</p> <p>D12 Resolver problemas envolvendo o cálculo ou estimativa de áreas de figuras planas, desenhadas em malhas quadrangulares.</p>		
<p>3. NÚMERO E OPERAÇÕES / ÁLGEBRA E FUNÇÕES</p>	<p>D13 Reconhecer e utilizar características do sistema de numeração decimal, tais como agrupamento e trocas na base 10 e princípio do valor posicional.</p> <p>D14 Identificar a localização de números naturais na reta numérica.</p> <p>D15 Reconhecer a decomposição de números naturais nas suas diversas ordens.</p> <p>D16 Reconhecer a composição e a decomposição de números naturais em sua forma polinomial.</p> <p>D17 Calcular o resultado de uma adição ou subtração de números naturais.</p> <p>D18 Calcular o resultado de uma multiplicação ou divisão de números naturais.</p> <p>D19 Resolver problemas com números naturais, envolvendo diferentes significados da adição ou subtração: juntar, alteração de um estado inicial (positiva ou negativa), comparação e mais de uma transformação (positiva ou negativa).</p> <p>D20 Resolver problemas com números naturais, envolvendo diferentes significados da multiplicação ou divisão: multiplicação comparativa, idéia de proporcionalidade, configuração retangular e combinatória.</p> <p>D21 Identificar diferentes representações de um mesmo número racional.</p> <p>D22 Identificar a localização de números racionais representados na forma decimal na reta numérica.</p> <p>D23 Resolver problema utilizando a escrita decimal de cédulas e moedas do sistema monetário brasileiro.</p> <p>D24 Identificar fração como representação que pode estar associada a diferentes significados.</p> <p>D25 Resolver problemas com números racionais expressos na forma decimal envolvendo diferentes significados da adição ou subtração.</p> <p>D26 Resolver problema envolvendo noções de porcentagem (25%,50%, 100%).</p>	<p>D16 Identificar a localização de números inteiros na reta numérica.</p> <p>D17 Identificar a localização de números racionais na reta numérica.</p> <p>D18 Efetuar cálculos com números inteiros, envolvendo as operações (adição, subtração, multiplicação, divisão, potenciação).</p> <p>D19 Resolver problema com números naturais, envolvendo diferentes significados das operações (adição, subtração, multiplicação, divisão, potenciação).</p> <p>D20 Resolver problema com números inteiros envolvendo as operações (adição, subtração, multiplicação, divisão, potenciação).</p> <p>D21 Reconhecer as diferentes representações de um número racional.</p> <p>D22 Identificar fração como representação que pode estar associada a diferentes significados.</p> <p>D23 Identificar frações equivalentes.</p> <p>D24 Reconhecer as representações decimais dos números racionais como uma extensão do sistema de numeração decimal, identificando a existência de “ordens” como décimos, centésimos e milésimos.</p> <p>D25 Efetuar cálculos que envolvam operações com números racionais (adição, subtração, multiplicação, divisão, potenciação).</p> <p>D26 Resolver problemas com números racionais envolvendo as operações (adições, subtração, multiplicação, divisão, potenciação).</p> <p>D27 Efetuar cálculos simples com valores aproximados de radicais.</p> <p>D28 Resolver problema que envolva porcentagem.</p> <p>D29 Resolver problema que envolva variações proporcionais, diretas ou inversas, entre grandezas.</p> <p>D30 Calcular o valor numérico de uma expressão algébrica.</p> <p>D31 Resolver problema que envolva equação do 2º grau.</p> <p>D32 Identificar a expressão algébrica que expressa uma regularidade observada em seqüência de números ou figuras (padrões).</p> <p>D33 Identificar uma equação ou inequação do 1º grau</p>	<p>D14 Identificar a localização de número reais na reta numérica.</p> <p>D15 Resolver problema que envolvam variações proporcionais, diretas ou inversas, entre grandezas.</p> <p>D16 Resolver problema que envolva porcentagem.</p> <p>D17 Resolver problema envolvendo equação do 2º grau.</p> <p>D18 Reconhecer expressão algébrica que representa uma função a partir de uma tabela.</p> <p>D19 Resolver problema envolvendo uma função do 1º grau.</p> <p>D20 Analisar crescimento/decrescimento, zeros de funções reais apresentadas em gráficos.</p> <p>D21 Identificar o gráfico que representa uma situação descrita em um texto.</p> <p>D22 Resolver problema envolvendo P.A./P.G. dada a fórmula do termo geral.</p> <p>D23 Reconhecer o gráfico de uma função polinomial de 1º grau por meio de seus coeficientes.</p> <p>D24 Reconhecer a representação algébrica de uma função do 1º grau dado o seu gráfico.</p> <p>D25 Resolver problemas que envolvam os pontos de máximo ou de mínimo no gráfico de uma função polinomial do 2º grau.</p> <p>D26 Relacionar as raízes de um polinômio com sua decomposição em fatores do 1º grau .</p> <p>D27 Identificar a representação algébrica e/ou gráfica de uma função exponencial.</p> <p>D28 Identificar a representação algébrica e/ou gráfica de uma função logarítmica, reconhecendo-a como inversa da função exponencial.</p> <p>D29 Resolver problema que envolva função exponencial.</p> <p>D30 Identificar gráficos de funções trigonométricas (seno, cosseno, tangente) reconhecendo suas propriedades.</p> <p>D31 Determinar a solução de um sistema linear associando-o à uma matriz.</p> <p>D32 Resolver problema de contagem utilizando o princípio multiplicativo ou noções de permutação simples, arranjo simples e/ou combinação simples.</p>

TEMA	4ª SÉRIE	8ª SÉRIE	3ª SÉRIE
		que expressa um problema. D34 Identificar um sistema de equação do 1º grau que expressa um problema. D35 Identificar a relação entre as representações algébricas e geométricas de um sistema de equação do 1º grau.	D33 Calcular a probabilidade de um evento.
4. TRATAMENTO DA INFORMAÇÃO	D27 Ler informações e dados apresentados em tabelas. D28 Ler informações e dados apresentados em gráficos (particularmente em gráficos de colunas.)	D36 Resolver problema envolvendo informações apresentadas em tabelas e/ou gráficos. D37 Associar informações apresentadas em listas e/ou tabelas simples aos gráficos que as representam e vice-versa.	D34 Resolver problemas envolvendo informações apresentadas em tabelas e/ou gráficos. D35 Associar informações apresentadas em listas e/ou tabelas simples aos gráficos que as representam e vice-versa.

Anexo 2 – Matrizes referenciais e tabelas de especificação do NAEP

A - Leitura

Quadro 8 – Contextos especificados para LEITURA

Contexto	Descrição
Leitura para experiência literária	Os leitores exploram eventos, personagens, temas, ambientes, enredos, ações e a linguagem de palavras literárias comuns na leitura de romances, histórias curtas, poemas, peças de teatro, biografias, lendas, mitos e folclores.
Leitura para informação	Os leitores tiram proveito de informações que permitam o entendimento do mundo por meio da leitura de revistas, jornais, livros didáticos, ensaios e discursos.
Leitura para o desempenho de uma tarefa	Os leitores aplicam o que lêem, como por exemplo, reconhecer horários e direções de ônibus e trens, seguir orientações para reparos, jogos, montagens, formulários de taxas, mapas e outras mais.

Quadro 9 – Aspectos da Leitura descritos na Matriz do NAEP

Formar um entendimento geral do texto	Desenvolver uma interpretação do texto	Estabelecer conexões entre o leitor e o texto	Examinar o conteúdo e a estrutura do texto
Considerar o texto integralmente ↓ Entendimento de forma ampla.	Focar sobre partes específicas ↓ Ligar informações através de partes do texto.	Pensar além do texto ↓ Aplicar o texto a situações do mundo real.	Considerar o porquê e como o texto foi desenvolvido ↓ Considerar o conteúdo, organização e forma do texto.

Quadro 10 – Questões e estratégias de leitura por contexto e aspecto (Tabela de Especificação)

CONTEXTO PARA LEITURA	ASPECTOS DA LEITURA			
	FORMAR UM ENTENDIMENTO GERAL	INTERPRETAR	ESTABELECEER CONEXÕES ENTRE O LEITOR E O TEXTO	EXAMINAR O CONTEÚDO E A ESTRUTURA
LEITURA PARA EXPERIÊNCIA LITERÁRIA	Do que se trata a história/enredo? ➢ Síntese ➢ Análise ➢ Inferência	Como esse personagem mudou do início para o fim da história? ➢ Síntese ➢ Análise ➢ Inferência ➢ Usar detalhes	Que outro personagem você conhece que tem um problema similar? ➢ Analogia ➢ Síntese ➢ Usar detalhes ➢ Relacionar informações e idéias	Qual é o humor dessa história? Que linguagem o autor usa para demonstrá-lo? ➢ Usar detalhes ➢ Inferência ➢ Síntese ➢ Pesquisa
LEITURA PARA INFORMAÇÃO	O que diz o autor a respeito de determinado tópico? ➢ Generalização ➢ Usar detalhes ➢ Síntese	O que causou determinada mudança? ➢ Causa e efeito ➢ Inferência ➢ Pesquisa	Que outro evento na história ou em notícias recentes é similar a este? ➢ Analogia ➢ Síntese ➢ Análise ➢ Inferência ➢ Usar detalhes ➢ Relacionar	O autor apresenta algum viés em relação ao artigo que escreveu? Dê a sua resposta com base na leitura desse artigo. ➢ Síntese ➢ Análise ➢ Julgamento ➢ Inferência

CONTEXTO PARA LEITURA	ASPECTOS DA LEITURA			
	FORMAR UM ENTENDIMENTO GERAL	INTERPRETAR	ESTABELECEER CONEXÕES ENTRE O LEITOR E O TEXTO	EXAMINAR O CONTEÚDO E A ESTRUTURA
			informações e idéias	<ul style="list-style-type: none"> ➢ Usar detalhes ➢ Distinguir fato e opinião
LEITURA PARA DESEMPENHAR UMA TAREFA	<p>Que tempo você gasta para poder voar para X sem parar?</p> <ul style="list-style-type: none"> ➢ Pesquisa 	<p>O que você deve fazer antes do passo x?</p> <ul style="list-style-type: none"> ➢ Pesquisa ➢ Inferência ➢ Seqüência 	<p>Descreva uma situação na qual você omitiria o passo x.</p> <ul style="list-style-type: none"> ➢ Inferência ➢ Análise ➢ Usar detalhes ➢ Relatar informações e idéias 	<p>A informação indicada neste texto é fácil de usar?</p> <ul style="list-style-type: none"> ➢ Avaliação ➢ Usar detalhes ➢ Síntese ➢ Pesquisa

Quadro 11 – Amostra de itens do NAEP por elementos de Textos Literários e Aspectos da Leitura

ELEMENTOS DE UM TEXTO LITERÁRIO	ASPECTOS DA LEITURA			
	FORMAR UM ENTENDIMENTO GERAL	INTERPRETAR	ESTABELECEER CONEXÕES ENTRE O LEITOR E O TEXTO	EXAMINAR O CONTEÚDO E A ESTRUTURA
TEMA	Qual é a moral da história? Use uma evidência da história na sua resposta.	Como o ambiente ajuda a ilustrar o tema da história?	Você pensa que a lição dessa história é verdadeira nos dias de hoje? Por que?	Explica o que torna essa história uma fábula.
PERSONAGENS PRINCIPAIS	Qual era a opinião do principal personagem de _____?	Qual a causa o personagem principal ter feito _____? Use evidências da história em sua resposta.	Que ações diferentes do personagem você pensaria para os dias de hoje? Busque apoio para sua resposta em evidências do texto.	Como a descrição do autor de _____ ajudaria a explicar as ações do personagem?
EVENTOS PRINCIPAIS	Escreva um resumo dos principais eventos da história.	O que se espera depois de _____?	Como você pensa que a história terminaria se _____ não acontecesse?	Como o primeiro evento ajuda você a prever o final?
PROBLEMA	O que pode tornar o problema pior?	Como pode ajudar a resolver o problema?	O problema da história pode ser comparado com o de outra que você já tenha lido? Inclua evidência do texto e a outra história.	Por que o autor explica o problema na primeira parte da história? Explique com evidências da história.
VOCABULÁRIO	Quais palavras descrevem o que principalmente a história diz respeito?	Quais palavras deixam você saber que tempo se passa a história?	Explique o duplo significado de _____. Descubra qual significado melhor explica as idéias principais do texto.	Por que o autor usa as palavras _____ para descrever como _____ sente?

Quadro 12 – Amostra de itens do NAEP por elementos de Textos Informativos e Aspectos da Leitura

ELEMENTOS DE UM TEXTO INFORMATIVO	ASPECTOS DA LEITURA			
	FORMAR UM ENTENDIMENTO GERAL	INTERPRETAR	ESTABELECEER CONEXÕES ENTRE O LEITOR E O TEXTO	EXAMINAR O CONTEÚDO E A ESTRUTURA
PROPÓSITO CENTRAL	Qual poderia ser a mensagem do autor nesse artigo?	Como o autor transmite a mensagem?	Você concorda com a mensagem do autor? Dê evidências do texto.	Baseado sobre o que você leu, qual poderia ser a razão do autor ter escrito isso.

ELEMENTOS DE UM TEXTO INFORMATIVO	ASPECTOS DA LEITURA			
	FORMAR UM ENTENDIMENTO GERAL	INTERPRETAR	ESTABELECEER CONEXÕES ENTRE O LEITOR E O TEXTO	EXAMINAR O CONTEÚDO E A ESTRUTURA
IDÉIAS PRINCIPAIS	Faça um resumo das idéias principais.	Como a grande idéia do primeira seção se relaciona com a grande idéia da última seção?	Quem poderia precisar ou querer essa informação? Use detalhes do texto em sua resposta.	O que faria o autor para apresentar a informação claramente?
IDÉIAS AUXILIARES	Identifique as idéias que mais proximamente se relacionam ao tópico. Dê evidências do texto para apoiar sua escolha.	Como o autor mostra a você que a idéia principal é importante?	Quais detalhes acerca do _____ ajudam você a ter uma clara imagem do tópico: explique por que você os escolheu.	Que informação o autor tem que saber antes de escrever o artigo?
APOIOS AUXILIARES	O gráfico neste artigo é mais usado para _____.	Como a informação no gráfico apóia a informação no artigo?	Por que o autor inclui a imagem junto ao gráfico? Explique usando o que você sabe e a informação do texto.	O que é significativo na ilustração para o artigo?
VOCABULÁRIO	Quais palavras descrevem o que o texto mais diz respeito? Use evidência do texto para apoiar a sua escolha.	Quais palavras você pensa que significam a mesma coisa do título? Diga por que você pensa assim.	Explique o duplo significado de _____. Diga qual o significado melhor explica as idéias principais no texto.	Por que o autor deu uma definição de _____ no parágrafo ____?

Quadro 13 – Amostra de itens do NAEP por elementos de Textos Práticos (desempenho de uma tarefa) e Aspectos da Leitura

ELEMENTOS DE UM TEXTO PRÁTICO	ASPECTOS DA LEITURA			
	FORMAR UM ENTENDIMENTO GERAL	INTERPRETAR	ESTABELECEER CONEXÕES ENTRE O LEITOR E O TEXTO	EXAMINAR O CONTEÚDO E A ESTRUTURA
PROPÓSITO CENTRAL	Qual é o propósito do documento? Use evidência do documento para sua resposta.	Essas orientações listam todo o material que você precisa? É uma razão para sua resposta.	Você pensa que as orientações tenham que ser as mesmas para _____? Use evidências do texto.	Isso é principalmente voltado para leitores familiarizados com a atividade? Explique por que.
INFORMAÇÃO PRINCIPAL	Diga como essa informação seria útil. Use evidências do documento.	O que é importante fazer antes _____? Use evidências do documento em sua resposta.	Compare as orientações a um conjunto de orientações que você tenha usado no passado. Qual o conjunto é mais fácil seguir? Conte por que.	Nomeie um passo que seja importante para seguir nas orientações. Explique por que esse passo é importante.
CARACTERÍSTICAS PRINCIPAIS DA ORGANIZAÇÃO	Diga o que você precisa para completar a tarefa. Quais características indicam isso?	O que acontece depois de _____? Explique por que isso é um passo importante.	Quando você poderia usar esses passos? Apóie sua resposta com evidências do texto.	Explique como o autor organizou esse documento. Use apoio do texto.

ELEMENTOS DE UM TEXTO PRÁTICO	ASPECTOS DA LEITURA			
	FORMAR UM ENTENDIMENTO GERAL	INTERPRETAR	ESTABELECEER CONEXÕES ENTRE O LEITOR E O TEXTO	EXAMINAR O CONTEÚDO E A ESTRUTURA
ILUSTRAÇÕES	Qual o significado das ilustrações para o conjunto total de orientações? Use apoio do texto em sua resposta.	Por que a segunda imagem mostra _____?	Quais ilustrações adicionais você acrescentaria para tornar as orientações mais claras?	Por que as palavras estão em negrito em frente a cada seção?
VOCABULÁRIO	Quais palavras descrevem ao que o documento mais diz respeito? Use evidência do texto para apoiar a sua escolha.	Use o contexto para dizer o significado de _____.	Explique por que a frase _____ é útil nas orientações.	Por que o autor usa as palavras _____ para apresentar o último passo das orientações?

B – Matemática – Matriz de Matemática

Quadro 14 – Sentido dos números

SENTIDO DOS NÚMEROS		
GRAU 4	GRAU 8	GRAU 12
a) Identificar o valor posicional e o valor real de dígitos em números inteiros.	a) Usar valor posicional para modelar e descrever números inteiros e decimais	
b) Representar números usando modelos tais como representações na base 10, reta numérica e modelos de duas dimensões.	b) Modelar ou descrever números racionais ou relações numéricas usando reta numérica e diagramas.	
c) Compor ou decompor quantidades inteiras pelo valor posicional (por exemplo, escrever números inteiros em notação expandida, usando o valor posicional).		
d) Escrever ou renomear números inteiros (e.g., $10 : 5 + 5$, $12 - 2$, 2×5)	d) Escrever ou renomear números racionais.	d) Escrever, renomear, representar, ou comparar números reais (e.g., π , $\sqrt{2}$, relações numéricas usando retas numéricas, modelos ou diagramas).
e) Associar, número por extenso, ou número usando vários modelos e representações por números inteiros, frações e decimais.	e) Reconhecer, traduzir, ou aplicar múltiplas representações de números racionais (frações, decimais e percentuais) em contextos significativos.	
	f) Expressar ou interpretar números usando notação científica de contextos da vida real.	f) Representar números muito grandes ou muito pequenos usando notação científica em contextos significativos.
	g) Encontrar ou modelar valor absoluto ou aplicar para situações-problema.	g) Encontrar ou modelar valor absoluto ou aplicar para situações-problema.
		h) Interpretar números dispostos em calculadora ou computador dados em notação científica.
	i) Ordenar ou comparar números racionais (frações, decimais, percentuais, ou inteiros) usando vários modelos e representações (e.g., reta numérica)	
j) Ordenar ou comparar números inteiros, decimais, ou frações	j) Ordenar ou comparar números racionais, incluindo inteiros muito grandes e muito pequenos, e decimais, e frações próximas de zero.	j) Ordenar ou comparar números reais, incluindo os muito pequenos ou os muito grandes.

Quadro 15 – Estimação

ESTIMAÇÃO		
GRAU 4	GRAU 8	GRAU 12
a) Usar representações (números bem conhecidos usados como pontos significativos para comparação) para números inteiros, decimais, ou frações em contextos (e.g., $\frac{1}{2}$ e 0,5 podem ser usados como representações para frações e decimais entre 0 e 1,00).	a) Estabelecer ou aplicar representações para números racionais e irracionais comuns (e.g., π) em contextos.	a) Estabelecer ou aplicar representações para números reais em contextos.
b) Fazer estimativas apropriadas para uma dada situação com números inteiros, frações, ou decimais: <ul style="list-style-type: none"> • sabendo quando estimar; • selecionando o tipo apropriado de estimativa, incluindo super estimativa, sub estimativa e grau da estimativa, ou • selecionando o método apropriado de estimação (e.g., arredondando) 	b) Fazer estimativas apropriadas para uma dada situação: <ul style="list-style-type: none"> • identificando quando a estimativa é apropriada, • determinando o nível de acurácia necessária, • selecionando o método apropriado de estimação, ou • analisando o efeito de um método de estimação sobre a acurácia dos resultados. 	b) Fazer estimativas de números muito grandes ou muito pequenos apropriadas para uma dada situação: <ul style="list-style-type: none"> • identificando quando a estimativa é apropriada ou não, • determinando o nível de acurácia necessária, • selecionando o método apropriado de estimação, ou • analisando o efeito de um método de estimação sobre a acurácia dos resultados.
c) Verificar soluções ou determinar a racionalidade de resultados em contextos significativos.	c) Verificar soluções ou determinar a racionalidade de resultados em uma variedade de situações, incluindo calculadora e resultados de computador.	c) Verificar soluções ou determinar a racionalidade de resultados em uma variedade de situações, incluindo notações científica, calculadora e resultados de computador.
	d) estimar raiz quadrada ou cúbica de números menores que 1 000 entre dois números inteiros.	d) estimar raiz quadrada ou cúbica de números menores que 1 000 entre dois números inteiros.

Quadro 16 – Operações com números

OPERAÇÕES COM NÚMEROS		
GRAU 4	GRAU 8	GRAU 12
a) Adicionar e subtrair: <ul style="list-style-type: none"> • Números inteiros, ou • Frações com denominadores comuns, ou • Décimos até centésimos. 	a) Executar cálculos com números racionais.	a) Executar cálculos com números reais, incluindo números irracionais ou números de valores absolutos.
b) Multiplicar números inteiros: <ul style="list-style-type: none"> • Não maiores que dois dígitos por dois dígitos, com cálculo 		

OPERAÇÕES COM NÚMEROS		
GRAU 4	GRAU 8	GRAU 12
<p>com lápis e papel, ou</p> <ul style="list-style-type: none"> Números maiores com uso de calculadora. 		
<p>c) Dividir números inteiros:</p> <ul style="list-style-type: none"> Até três dígitos por um dígito, com cálculo com lápis e papel, ou Até cinco dígitos por dois dígitos, com uso de calculadora. 		
<p>d) Descrever o efeito de operações sobre tamanho. (números inteiros)</p>	<p>d) Descrever o efeito de multiplicar e dividir por números, incluindo o efeito de multiplicar ou dividir um número racional por:</p> <ul style="list-style-type: none"> Zero, ou Um número menor que zero, ou Um número entre zero e um, ou, Um, ou Um número maior que um. 	<p>d) Descrever o efeito de multiplicar e dividir por números, incluindo o efeito de multiplicar ou dividir um número real por:</p> <ul style="list-style-type: none"> Zero, ou Um número menor que zero, ou Um número entre zero e um, ou, Um, ou Um número maior que um.
	<p>e) Proporcionar um argumento matemático para explicar operações com duas ou mais frações.</p>	
<p>f) Interpretar operações com números inteiros e as relações entre eles.</p>	<p>f) Interpretar operações com números racionais e as relações entre eles.</p>	
<p>g) Resolver problemas de aplicação envolvendo números e operações.</p>	<p>g) Resolver problemas de aplicação envolvendo números racionais e operações usando respostas exatas ou estimativas apropriadas.</p>	<p>g) Resolver problemas de aplicação envolvendo números racionais e irracionais comuns, operações usando respostas exatas ou estimativas apropriadas.</p>

Quadro 17 – Razão e Proporção

RAZÃO E PROPORÇÃO		
GRAU 4	GRAU 8	GRAU 12
<p>a) Usar razões simples para descrever situações-problema.</p>	<p>a) Usar razão para descrever situações-problema.</p>	
	<p>b) Usar frações para representar e expressar razões e proporções.</p>	<p>b) Usar proporções para modelar problemas.</p>

RAZÃO E PROPORÇÃO		
GRAU 4	GRAU 8	GRAU 12
	c) Usar razão proporcional para modelar e resolver problemas (incluindo taxas e acúmulo)	c) Usar razão proporcional para resolver problemas (incluindo taxas)
	d) Resolver problemas envolvendo porcentagens (acréscimo e decréscimo percentual, lucro, desconto, taxa, gorjeta, ou relação parte / todo)	d) Resolver problemas envolvendo porcentagens (acréscimo e decréscimo percentual, lucro, desconto, taxa, gorjeta, ou relação parte / todo)

Quadro 18 – Propriedades dos Números e Operações

PROPRIEDADES DOS NÚMEROS E OPERAÇÕES		
GRAU 4	GRAU 8	GRAU 12
a) Identificar números pares e ímpares.	a) Descrever inteiros pares e ímpares e como eles se comportam em diferentes operações.	
b) Identificar fatores de números inteiros.	b) Reconhecer, encontrar, ou usar fatores, múltiplos, ou fatoração de número primo.	b) Resolver problemas envolvendo fatores, múltiplos, ou fatoração de número primo.
	c) Reconhecer ou usar número primo e compor números para resolver problemas.	c) Usar número primo ou compor números para resolver problemas.
	d) Usar divisibilidade ou resíduos em problemas ambientais.	d) Usar divisibilidade ou resíduos em problemas ambientais.
e) Aplicar propriedades básicas das operações.	e) Aplicar propriedades básicas das operações.	e) Aplicar propriedades básicas das operações.
f) Explicar ou justificar um conceito matemático ou relacionamento (e.g., explicar por que 15 é um número ímpar ou por que $7-3$ não é o mesmo que $3-7$).	f) Explicar ou justificar um conceito matemático ou relacionamento (e.g., explicar por que 17 é um número primo).	f) Proporcionar um argumento matemático acerca de uma propriedade ou uma relação numérica.

Quadro 19 – Atributos físicos da medida

ATRIBUTOS FÍSICOS DA MEDIDA		
GRAU 4	GRAU 8	GRAU 12
a) Identificar o atributo que seja apropriado para medir uma dada situação.		
b) Comparar objetos de um dado atributo, tais como comprimento, área, volume, tempo, ou temperatura.	b) Comparar objetos a respeito de comprimento, área, volume, medida de ângulo, peso ou massa.	
c) Estimar o tamanho de um objeto a respeito de um dado atributo de medida (e.g., comprimento, perímetro, ou área usando uma rede quadriculada).	c) Estimar o tamanho de um objeto a respeito de um dado atributo de medida (e.g., área).	c) Estimar ou comparar perímetros ou áreas de figuras geométricas bidimensionais.

ATRIBUTOS FÍSICOS DA MEDIDA		
GRAU 4	GRAU 8	GRAU 12
		d) Estimar ou comparar volume ou superfície de figuras tridimensionais.
		e) Resolver problemas envolvendo o plano coordenado tais como a distância entre dois pontos, o ponto médio de um segmento, ou inclinações de linhas paralelas ou perpendiculares.
		f) Resolver problemas de medidas de ângulo, incluindo aquela envolvendo triângulos ou outros polígonos ou linhas paralelas cortadas por uma transversal.
g) Selecionar ou usar instrumentos de medida apropriados tais como régua, metro, relógio, termômetro, ou outros instrumentos escalares.	g) Selecionar ou usar instrumentos de medida apropriados para determinar ou criar um dado comprimento, área, volume, ângulo, peso, ou massa.	
h) Solucionar problemas envolvendo perímetro de figuras planas.	h) Resolver problemas matemáticos ou do mundo real envolvendo perímetro ou área de figuras planas tais como triângulos, retângulos, círculos ou figuras compostas.	h) Resolver problemas matemáticos ou do mundo real envolvendo perímetro ou área de figuras planas tais como polígonos, círculos ou figuras compostas.
i) Solucionar problemas envolvendo área de quadrados e retângulos..		
	j) Solucionar problemas envolvendo volume ou superfície de sólidos retangulares, cilindros, prismas, ou formas compostas.	j) Solucionar problemas envolvendo volume ou superfície de sólidos retangulares, cilindros, cones, pirâmides, prismas, esferas, ou formas compostas.
	k) Solucionar problemas envolvendo medidas indiretas tais como descobrir a altura de uma construção comparando sua sombra com a altura e a sombra de um objeto conhecido.	k) Solucionar problemas envolvendo medidas indiretas tais como descobrir a altura de uma construção pela descoberta da distância entre a base da construção e o ângulo de elevação para o topo.
	l) Solucionar problemas envolvendo taxas tais como velocidade ou densidade populacional.	l) Solucionar problemas envolvendo taxas tais como velocidade, densidade, densidade populacional, ou taxas correntes.
		m) Usar relações trigonométricas em triângulos retângulos para resolver problemas.

Quadro 20 – Sistemas de medida

ATRIBUTOS FÍSICOS DA MEDIDA		
GRAU 4	GRAU 8	GRAU 12
a) Selecionar ou usar tipo apropriado de unidade para o atributo que está sendo medido tais como comprimento, tempo, ou temperatura.	a) Selecionar ou usar tipo apropriado de unidade para o atributo que está sendo medido tais como comprimento, área, ângulo, tempo, ou volume.	a) Selecionar ou usar tipo apropriado de unidade para o atributo que está sendo medido tais como volume ou superfície.
b) Resolver problemas envolvendo	b) Resolver problemas envolvendo	b) Resolver problemas envolvendo

ATRIBUTOS FÍSICOS DA MEDIDA		
GRAU 4	GRAU 8	GRAU 12
conversões dentro do mesmo sistema de medidas tais como conversões envolvendo polegadas e pés, ou horas e minutos.	conversões dentro do mesmo sistema de medidas tais como conversões envolvendo polegadas quadradas e pés quadrados.	conversões dentro ou entre sistemas de medidas dadas as relações entre as unidades.
	c) Estimar a medida de um objeto em um sistema dada a medida do objeto em outro sistema e o fator de conversão aproximado. Por exemplo: <ul style="list-style-type: none"> • Conversão da distância: 1 quilômetro é aproximadamente $\frac{5}{8}$ de uma milha; • Conversão de dinheiro: US dollar é aproximadamente 1,5 dólares canadenses; 	
d) Determinar tamanho apropriado de unidade de medida em situação-problema envolvendo atributos como extensão, tempo, capacidade, ou peso.	d) Determinar tamanho apropriado de unidade de medida em situação-problema envolvendo atributos como extensão, área, ou volume.	
e) Determinar situações nas quais uma medida altamente acurada seja importante.	e) Determinar a precisão de uma medida apropriada a situações-problema (por exemplo, a acurácia de cada uma das várias medidas de comprimento necessárias para obter uma precisão especificada de um comprimento total) e encontrar o grau de precisão da medida.	e) Determinar a precisão de uma medida apropriada a situações-problema (por exemplo, a acurácia de medidas das dimensões para obter precisão especificada da área) e encontrar o grau de precisão da medida.
	f) Construir ou resolver problemas (por exemplo, a área do piso de uma sala) envolvendo escalas desenhadas.	f) Construir ou resolver problemas (por exemplo, especificar a quantidade de rolos de papéis necessários para envolver as paredes internas de uma casa) envolvendo escalas desenhadas.
		g) Comparar comprimentos áreas ou volumes de figuras similares usando proporções.

Quadro 21 – Dimensão e forma

DIMENSÃO E FORMA		
GRAU 4	GRAU 8	GRAU 12
a) Explorar propriedades de trajetos entre pontos	a) Desenhar ou descrever um trajeto de extensão mais curta entre pontos para solucionar problemas do contexto.	
b) Identificar ou descrever (informalmente) objetos do mundo real usando figuras planas simples (p. ex., triângulos, quadrados e círculos) e figuras sólidas simples (p. ex., cubos, esferas e cilindros).	b) Identificar um objeto geométrico, dada uma descrição escrita de suas propriedades.	b) Usar representações bidimensionais de objetos tridimensionais para visualizar e resolver problemas envolvendo a área de uma superfície e o volume.
c) Identificar ou desenhar ângulos	c) Identificar, definir, ou descrever	c) Dar descrições matemáticas

DIMENSÃO E FORMA		
GRAU 4	GRAU 8	GRAU 12
e outras figuras geométricas no plano.	formas geométricas no plano e no espaço tridimensional, dada uma representação visual.	precisas ou definições de formas geométricas no plano e no espaço tridimensional.
	d) Desenhar ou esboçar de uma descrição escrita de polígonos, círculos ou semi-círculos.	d) Desenhar ou esboçar de uma descrição escrita de figuras planas (p. ex., triângulos isósceles, polígonos regulares, e figuras curvas) e imagens planas de figuras tridimensionais (p. ex., poliedros, esferas e hemisférios).
	e) Representar ou descrever uma situação tridimensional em um desenho bidimensional de diferentes visões.	e) Descrever ou analisar propriedades das esferas e hemisférios.
f) Descrever atributos de formas bidimensionais e tridimensionais.	f) Demonstrar compreensão acerca das formas bidimensionais e tridimensionais em nosso mundo através da identificação, desenho, modelagem, construção, ou planificações.	

Quadro 22 – Transformação de formas e preservação das propriedades

TRANSFORMAÇÃO DE FORMAS E PRESERVAÇÃO DAS PROPRIEDADES		
GRAU 4	GRAU 8	GRAU 12
a) Identificar se uma figura é simétrica, ou desenhar linhas de simetria	a) Identificar linhas de simetria em figuras planas ou reconhecer e classificar tipos de simetrias de figuras planas.	a) Reconhecer ou identificar tipos de simetrias (p.ex., ponto, linha, rotações, auto-congruências) de figuras bi e tridimensionais.
		b) Dar ou reconhecer o relacionamento matemático preciso (p. ex., congruência, similaridade, orientação) entre uma figura e sua imagem sob uma transformação.
c) Identificar as imagens resultantes de movimentações (flexões), planos inclinados (translações), ou giros (rotações).	c) Reconhecer, ou informalmente descrever o efeito de uma transformação em formas geométricas bidimensionais (movimentações através de linhas de simetria, rotações, translações, aumentos e contrações).	c) Executar ou descrever o efeito de uma única transformação sobre formas geométricas bi e tridimensionais (movimentações através de linhas de simetria, rotações, translações, e dilatações).
d) Reconhecer quais atributos (tais como forma e área) mudam ou não, quando figuras planas são cortadas ou rearranjadas.	d) Predizer resultados de combinações, subdivisões, e mudanças de formas de figuras planas e sólidos (p. ex., dobraduras de papel, ladrilhar, e cortar e rearranjar peças).	d) Descrever o resultado final de transformações sucessivas.
e) Comparar ou desenhar figuras congruentes em uma dada coleção.	e) Justificar relacionamentos de congruência e similaridade, e aplicar esses relacionamentos usando escala e raciocínio proporcional.	e) Justificar relacionamentos de congruência e similaridade, e aplicar esses relacionamentos usando escala e raciocínio proporcional.

TRANSFORMAÇÃO DE FORMAS E PRESERVAÇÃO DAS PROPRIEDADES		
GRAU 4	GRAU 8	GRAU 12
f)	f) Para figuras similares, identificar e usar as relações de conservação de ângulos e de proporcionalidade da extensão lado e perímetro.	

Quadro 23 – Relações entre figuras geométricas

RELAÇÕES ENTRE FIGURAS GEOMÉTRICAS		
GRAU 4	GRAU 8	GRAU 12
a) Analisar ou descrever padrões de figuras geométricas pelo aumento do número de lados, mudando o tamanho ou orientação (p.ex., polígonos com mais lados).		
b) Reunir formas planas simples para construir uma forma dada.	b) Aplicar propriedades geométricas e relações na resolução de problemas simples em duas e três dimensões	b) Aplicar propriedades geométricas e relações na resolução de problemas simples em duas e três dimensões (incluindo figuras rígidas e não-rígidas).
c) Reconhecer faces bidimensionais de figuras tridimensionais	c) Representar situações-problema com modelos geométricos simples para solucionar problemas matemáticos ou do mundo real.	c) Representar situações-problema com modelos geométricos para solucionar problemas matemáticos ou do mundo real.
	d) Usar o Teorema de Pitágoras para solucionar problemas.	d) Usar o Teorema de Pitágoras para solucionar problemas em situações bi ou tridimensionais.
		e) Descrever e analisar propriedades dos círculos (p. ex., perpendicularidade da tangente e raio, ângulo inscrito em um semicírculo).
f) Descrever e comparar propriedades de figuras combinadas e simples compostas de triângulos, quadrados e retângulos.	f) Descrever ou analisar propriedades de, ou relações entre, triângulos, quadriláteros, e outras figuras planas poligonais.	f) Analisar propriedades ou relacionamentos de triângulos, quadriláteros, e outras figuras planas poligonais.
	g) Descrever ou analisar propriedades e relacionamentos de linhas paralelas ou cruzadas.	g) Descrever ou analisar propriedades e relacionamentos de linhas paralelas, perpendiculares, ou cruzadas, incluindo as relações entre ângulos que aumentam nesses casos.

Quadro 24 – Posição e Direção

POSIÇÃO E DIREÇÃO		
GRAU 4	GRAU 8	GRAU 12
a) Descrever posições relativas de pontos e linhas usando idéias geométricas de paralelismo ou perpendicularidade.	a) Descrever posições relativas de pontos e linhas usando idéias geométricas de ponto médio, pontos sobre linha comum através de um ponto comum,	

POSIÇÃO E DIREÇÃO		
GRAU 4	GRAU 8	GRAU 12
	paralelismo ou perpendicularidade.	
	b) Descrever a intersecção de duas ou mais figuras geométricas no plano (p. ex., intersecção de um círculo e uma linha).	b) Descrever as intersecções de linhas no plano e no espaço, intersecções de uma linha e um plano, ou dois planos no espaço.
	c) Visualizar ou descrever a seção cruzada de um sólido.	c) Descrever ou identificar cônicas e outras seções cruzadas de sólidos.
d) Construir figuras geométricas com vértices em pontos sobre uma rede coordenada.	d) Representar figuras geométricas usando coordenadas retangulares sobre um plano.	d) Representar figuras bidimensionais algebricamente, usando coordenadas e/ou equações.
		e) Usar vetores para representar velocidade e direção.

Quadro 25 – Raciocínio Matemático

RACIOCÍNIO MATEMÁTICO		
GRAU 4	GRAU 8	GRAU 12
a) Distinguir quais objetos em uma coleção satisfazem uma dada definição geométrica e explicar escolhas.	a) Fazer e testar uma conjectura geométrica acerca de polígonos regulares.	a) Fazer, testar e validar conjecturas geométricas usando uma variedade de métodos, incluindo raciocínio dedutivo e contra exemplos.

Quadro 26 – Representação dos dados

REPRESENTAÇÃO DOS DADOS		
GRAU 4	GRAU 8	GRAU 12
As seguintes representações de dados são indicadas para cada nível. Os objetivos os quais somente um subconjunto dessas representações são aplicáveis estão entre parênteses associados com os objetivos		
Gráficos de imagens, de barras, circular, de linha, tabelas e contagens.	Histogramas, gráficos de linha, de dispersão, de barras, de caixa, circulares, de caule e folhas, distribuições de frequência, tabelas	Histogramas, gráficos de linha, de dispersão, de caixa, circulares, de caule e folhas, distribuições de frequência e tabelas
a) Ler ou interpretar um único conjunto de dados.	a) Ler ou interpretar dados, incluindo interpolação e extrapolação de dados.	a) Ler ou interpretar dados, incluindo interpolação e extrapolação de dados.
b) Para um dado conjunto de dados, completar um gráfico (limites de tempo torna difícil construir gráficos completamente).	b) Para um dado conjunto de dados, completar um gráfico e então solucionar um problema usando os dados no gráfico (histogramas, gráficos de linha, de dispersão, de círculo e de barras).	b) Para um dado conjunto de dados, completar um gráfico e então solucionar um problema usando os dados no gráfico (histogramas, gráficos de linha, de dispersão).
c) Solucionar problemas pela estimativa e cálculo dentro de um único conjunto de dados.	c) Solucionar problemas pela estimativa e cálculo com dados de um único conjunto ou através de conjuntos de dados.	c) Solucionar problemas pela estimativa e cálculo com dados univariados ou bivariados (incluindo gráficos de dispersão e tabelas de duas entradas).
	d) Dado um gráfico ou um conjunto de dados, determinar se a informação está representada efetivamente e apropriadamente (histogramas,	d) Dado um gráfico ou um conjunto de dados, determinar se a informação está representada efetivamente e apropriadamente (histogramas,

REPRESENTAÇÃO DOS DADOS		
GRAU 4	GRAU 8	GRAU 12
As seguintes representações de dados são indicadas para cada nível. Os objetivos os quais somente um subconjunto dessas representações são aplicáveis estão entre parênteses associados com os objetivos		
Gráficos de imagens, de barras, circular, de linha, tabelas e contagens.	Histogramas, gráficos de linha, de dispersão, de barras, de caixa, circulares, de caule e folhas, distribuições de frequência, tabelas	Histogramas, gráficos de linha, de dispersão, de caixa, circulares, de caule e folhas, distribuições de frequência e tabelas
	gráficos de dispersão, de círculo, de linha e de barras).	gráficos de dispersão, de caixa, de linha e de barras).
	e) Comparar e contrastar a efetividade de diferentes representações dos mesmos dados.	e) Comparar e contrastar a efetividade de diferentes representações dos mesmos dados.

Quadro 27 – Características dos conjuntos de dados

CARACTERÍSTICAS DOS CONJUNTOS DE DADOS		
GRAU 4	GRAU 8	GRAU 12
	a) Calcular, usar, ou interpretar média, mediana, moda, ou intervalo.	a) Calcular, usar, ou interpretar média, mediana, moda, intervalo interquartilico, ou desvio padrão.
b) Dado um conjunto de dados ou um gráfico, descrever a distribuição dos dados usando mediana, intervalo, ou moda.	b) Descrever como média, mediana, moda, ou intervalo interquartilico relaciona com a forma da distribuição.	b) Reconhecer como transformações lineares de dados de uma variável afetam a média, mediana, moda e intervalo (p.ex., o efeito da média adicionando uma constante a cada ponto)
	c) Identificar casos extremos (<i>outliers</i>) e determinar seu efeito na média, mediana, ou intervalo.	c) Determinar o efeito de casos extremos (<i>outliers</i>) na média, mediana, moda, intervalo interquartilico, ou desvio padrão.
d) Comparar dois conjuntos de dados relacionados.	d) Usando medidas estatísticas apropriadas, comparar dois ou mais conjuntos de dados descrevendo a mesma característica para duas populações diferentes ou subconjuntos da mesma população.	d) Comparar dois ou mais conjuntos de dados usando média, mediana, moda, intervalo interquartilico, ou desvio padrão descrevendo a mesma característica para duas populações diferentes ou subconjuntos da mesma população.
	e) Visualmente, escolher a linha de melhor ajuste a um gráfico de dispersão e informalmente explicar o significado da linha. Usar a linha para fazer previsões.	e) Dado um conjunto de dados ou um gráfico de dispersão, visualmente, escolher a linha de ajuste melhor e explicar o significado da linha. Usar a linha para fazer previsões.
		f) Usar ou interpretar uma distribuição normal como um modelo matemático apropriado para sumarizar certos conjuntos de dados.
		g) Dado um gráfico de dispersão, tomar decisões ou fazer previsões envolvendo uma linha ou ma curva de melhor ajuste.
		h) Dado um gráfico de dispersão,

CARACTERÍSTICAS DOS CONJUNTOS DE DADOS		
GRAU 4	GRAU 8	GRAU 12
		estimar o coeficiente de correlação (p. ex., dado um gráfico de dispersão, a correlação está mais próxima de 0, 0,5 ou 1,0? É uma correlação positiva ou negativa?).

Quadro 28– Experimentos e amostras

EXPERIMENTOS E AMOSTRAS		
GRAU 4	GRAU 8	GRAU 12
	a) Dada uma amostra, identificar possíveis fontes de viés.	a) Identificar possíveis fontes de viés em métodos de coleta de dados e descrever como tais vieses podem ser controlados e reduzidos.
	b) Distinguir amostras randômicas e não-randômicas.	b) Reconhecer e descrever um método para selecionar uma amostra randômica simples.
		c) Fazer inferências de resultados amostrais.
	d) Avaliar o delineamento de um experimento.	d) Identificar ou avaliar as características de uma boa pesquisa ou de um experimento bem delineado.

Quadro 29 – Probabilidade

PROBABILIDADE		
GRAU 4	GRAU 8	GRAU 12
a) Usar pensamento probabilístico informal para descrever eventos de chance (Isto é, provável e improvável, certo e incerto)	a) Analisar uma situação que envolva probabilidade de um evento independente.	a) Analisar uma situação que envolva probabilidade de eventos dependentes ou independentes.
b) Determinar uma probabilidade simples de um contexto que inclua uma imagem.	b) Determinar a probabilidade teórica de eventos simples e compostos em contextos familiares.	b) Determinar a probabilidade teórica de eventos simples e compostos em contextos familiares ou não-familiares.
	c) Estimar a probabilidade de eventos simples e compostos através da experimentação ou simulação.	c) Dados os resultados de um experimento ou simulação, estimar a probabilidade de eventos simples e compostos em contextos familiares ou não-familiares.
	d) Usar probabilidade teórica para avaliar ou prever resultados experimentais.	d) Usar probabilidade teórica para avaliar ou prever resultados experimentais.
e) Listar todos os possíveis resultados de uma dada situação ou evento.	e) Determinar o espaço amostral para uma dada situação.	e) Determinar o número de vezes que um evento possa ocorrer usando diagramas de árvore, fórmulas para combinações ou permutações, ou outras técnicas de contagem.
	f) Usar espaço amostral para determinar a probabilidade de resultados possíveis de um evento.	f) Determinar a probabilidade de resultados possíveis de um evento.

PROBABILIDADE		
GRAU 4	GRAU 8	GRAU 12
g) Representar a probabilidade de um dado resultado usando uma imagem ou outro gráfico.	g) Representar a probabilidade de um dado resultado usando frações, decimais, e porcentagens.	
	h) Determinar a probabilidade de eventos dependentes e independentes. (Eventos dependentes devem ser limitados a funções lineares com tamanho amostral pequeno.)	h) Determinar a probabilidade de eventos dependentes e independentes.
		i) Determinar probabilidade condicional usando tabelas de duas entradas.
	j) Interpretar probabilidades dentro de um dado contexto.	j) Interpretar probabilidades dentro de um dado contexto.

Quadro 30 – Padrões, relações e funções

PADRÕES, RELAÇÕES E FUNÇÕES		
GRAU 4	GRAU 8	GRAU 12
a) Reconhecer, descrever, ou estender padrões numéricos.	a) Reconhecer, descrever, ou estender padrões numéricos e geométricos, usando tabelas, gráficos, palavras, ou símbolos.	a) Reconhecer, descrever, ou estender padrões numéricos e geométricos, usando tabelas, gráficos, palavras, ou símbolos.
b) Dar um padrão ou seqüência, construir ou explicar uma regra que possa gerar os termos do padrão ou seqüência.	b) Generalizar um padrão que apareça em uma seqüência numérica ou tabela ou gráfico usando palavras ou símbolos.	b) Expressar a função em termos gerais (ou recursivamente ou explicitamente), dada uma tabela, ou uma descrição verbal, ou alguns termos de uma seqüência.
c) Dar uma descrição, estender ou encontrar um termo ausente em um padrão ou seqüência.	c) Analisar ou criar padrões, seqüências, ou funções lineares usando uma regra.	
d) Criar uma representação diferente de um padrão ou seqüência dada uma descrição verbal.		
e) Reconhecer ou descrever um relacionamento no qual quantidades mudam a proporcionalidade.	e) Identificar funções como lineares ou não-lineares ou contrastar distinguindo propriedades de funções de tabelas, gráficos, ou equações.	e) Identificar ou analisar distinguindo propriedades de funções lineares, quadráticas, inversas ($y = k/x$) ou exponenciais de tabelas, gráficos, ou equações.
	f) Interpretar o significado de inclinações ou interceptos em funções lineares.	
		g) Determinar o domínio e o intervalo de funções dados vários contextos.
		h) Reconhecer e analisar as formas gerais de funções lineares, quadráticas, inversas, ou exponenciais (p. ex., em $y = ax + b$, reconhecer os papéis de a e b).
		i) Expressar funções lineares e

PADRÕES, RELAÇÕES E FUNÇÕES		
GRAU 4	GRAU 8	GRAU 12
		exponenciais em forma recursiva e explícita dada uma tabela ou uma descrição verbal.

Quadro 31 – Representações algébricas

REPRESENTAÇÕES ALGÉBRICAS		
GRAU 4	GRAU 8	GRAU 12
a) Interpretar entre as diferentes formas de representação (simbólicas, numéricas, verbais, ou ilustradas) o relacionamento de números inteiros (tais como de uma descrição escrita para uma equação ou de uma função tabelada para uma descrição escrita).	a) Interpretar entre diferentes representações de expressões lineares usando símbolos, gráficos, tabelas, diagramas, ou descrições escritas.	a) Interpretar entre diferentes representações de expressões algébricas usando símbolos, gráficos, tabelas, diagramas, ou descrições escritas.
	b) Analisar ou interpretar relacionamentos lineares expressos em símbolos, gráficos, tabelas, diagramas, ou descrições escritas.	b) Analisar ou interpretar relacionamentos expressos em símbolos, gráficos, tabelas, diagramas, ou descrições escritas.
c) Representar por um gráfico ou interpretar pontos com números inteiros ou letras coordenadas em grades ou no primeiro quadrante do plano coordenado.	c) Representar por um gráfico ou interpretar pontos que são representados por pares ordenadas de números sobre um sistema de coordenadas retangular.	c) Representar por um gráfico ou interpretar pontos que são representados por um ou mais pares ordenadas de números sobre um sistema de coordenadas retangular.
	d) Resolver problemas envolvendo pares coordenados sobre um sistema de coordenadas retangular.	d) Executar ou interpretar transformações nos gráficos de funções lineares e quadráticas.
e) Verificar uma conclusão usando propriedades algébricas.	e) Fazer, validar e justificar conclusões e generalizações acerca de relacionamentos lineares.	e) Usar propriedades algébricas para desenvolver um argumento matemático válido.
		f) Usar um modelo algébrico de uma situação para fazer inferências ou previsões.
	g) Identificar ou representar relacionamentos funcionais em contextos significativos incluindo relação proporcional, linear, e não-linear comum (p.ex., juro composto, crescimento bacteriológico) em tabelas, gráficos, palavras, ou símbolos.	g) Dada uma situação do mundo real, determinar se uma função linear, quadrática, inversa, ou exponencial ajusta à situação (p. ex., meia vida do crescimento bacteriológico).
		h) Solucionar problemas envolvendo crescimento e decréscimo exponencial.

Quadro 32 – Variáveis, expressões e operações

VARIÁVEIS, EXPRESSÕES E OPERAÇÕES		
GRAU 4	GRAU 8	GRAU 12
a) Usar letras e símbolos para representar uma quantidade desconhecida em uma expressão matemática simples.		
b) Expressar relacionamentos matemáticos simples usando sentenças numéricas.	b) Escrever expressões algébricas, ou inequações para representar uma situação.	b) Escrever expressões algébricas, ou inequações para representar uma situação.
	c) Executar operações básicas, usando instrumentos apropriados, sobre expressões algébricas lineares (incluindo agrupamento e ordenação de operações múltiplas envolvendo operações básicas, expoentes, raízes, simplificações e expansões).	c) Executar operações básicas, usando instrumentos apropriados, sobre expressões algébricas (incluindo agrupamento e ordenação de operações múltiplas envolvendo operações básicas, expoentes, raízes, simplificações).
		d) Escrever formas equivalentes de expressões algébricas, equações ou inequações para representar e explicar relacionamentos matemáticos.

Quadro 33 – Equações e inequações

REPRESENTAÇÕES ALGÉBRICAS		
GRAU 4	GRAU 8	GRAU 12
a) Encontrar o valor desconhecido em uma sentença de números inteiros.	a) Solucionar equações lineares ou inequações (p. ex., $ax + b = c$ ou $ax + b = cx + d$ ou $ax + b > c$	a) Solucionar equações lineares, racionais, ou quadráticas e inequações.
	b) Interpretar “=” como uma equivalência entre duas expressões e usar essa interpretação para solucionar problemas.	
	c) Analisar situações ou solucionar problemas usando equações lineares e inequações com coeficientes racionais simbolicamente ou graficamente (p. ex., $ax + b = c$ ou $ax + b = cx + d$).	c) Analisar situações ou solucionar problemas usando equações lineares ou quadráticas ou inequações simbolicamente ou graficamente.
	d) Interpretar relacionamentos entre expressões lineares simbólicas e gráficos de linha pela identificação e cálculo de inclinações e interceptos (p. ex., conhecido em $y = ax + b$, que a é o valor da mudança e b é o intercepto vertical do gráfico).	d) Reconhecer o relacionamento entre a solução de um sistema de equações lineares e seus gráficos.
	e) Usar e avaliar fórmulas comuns (p. ex., relacionamento entre uma circunferência do círculo e o diâmetro ($C = \pi d$), distância e tempo sob velocidade constante.	e) Resolver problemas envolvendo fórmulas mais avançadas [p. ex., volumes e áreas de superfícies de sólidos tridimensionais, ou fórmulas tais como: $A = P(1 + r)^t$, $A = Pe^{rt}$]

REPRESENTAÇÕES ALGÉBRICAS		
GRAU 4	GRAU 8	GRAU 12
		f) Dada uma fórmula familiar, solucionar para uma das variáveis.
		g) Solucionar ou interpretar sistemas de equações e inequações.

Anexo 3 – Exemplo de itens que tratam da mesma habilidade, mas encontram-se em níveis diferentes da escala

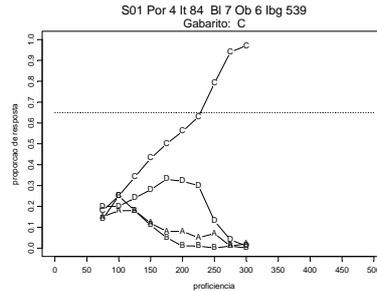
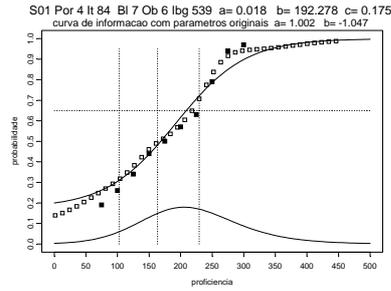
Leia o texto para responder às questões 6 a 9.

<p>MICRÓBIOS - amigos ou inimigos? Matam ou engordam?</p>  <p>Tem uma coisa que os adultos dizem que eu tenho certeza de que aborrece as crianças: “Vá lavar as mãos antes de comer! Ela está cheia de micróbios. Não coma esse troço que caiu no chão! Lave logo o machucado, senão os micróbios tomam conta!”</p> <p>Daí a criança vai logo pensando: “Coisa chata essa de micróbio!” E os micróbios vão ficando com essa fama de monstrinhos, sempre prontos a atacar em caso de desleixo.</p>	<p>Mas sem micróbios e bactérias também não dá para viver, porque há um montão deles que são essenciais para manter vida em nosso planeta.</p> <p>Quando a gente vai lavar as mãos antes de comer fica até meio desapontado, pois não vê micróbio nenhum. E acha aquilo um exagero. É que os micróbios são microscópicos. Atualmente são considerados micróbios ou microorganismos os fungos, as bactérias, os vírus, algumas algas e os protozoários.</p> <p>Os micróbios - não há como negar - são responsáveis por uma série de aborrecimentos: gripe, sarampo, tifo, malária, febre amarela, paralisia infantil e um bocado de coisas mais.</p> <p>Mas também há inúmeros micróbios benéficos, que decompõem o corpo morto das plantas e animais, transformando suas moléculas complexas em moléculas pequenas, aproveitáveis na nutrição das plantas.</p> <p>O vilão de nossa história, portanto, não é totalmente malvado. Se ele desaparecesse, nós também acabaríamos junto com ele.</p> <p>Adaptado: <i>CIÊNCIA HOJE DAS CRIANÇAS</i>. Rio de Janeiro: SBPC, ano 6. n.30, p.20-23.</p>
---	---

6) A leitura do texto oferece ao leitor

- (A) uma advertência.
- (B) um divertimento.
- (C) uma informação.
- (D) uma recomendação.

ÍNDICES									PROPORÇÕES DE RESPOSTA						COEFICIENTES BISSERIAIS					
ITEM	BL	OB	GAB	DIFI	DISCR	ABAI	ACIM	BISE	A	B	C	D	""	""	A	B	C	D	""	""
84	7	6	C	.47	.45	.24	.69	.47	.11	.09	.47	.26	.05	.02	-.24	-.39	.47	.00	-.47	-.68



Item: 84 Nível: 250

Tema: (2) Implicações do Suporte, do Gênero e/ou do Enunciador na Compreensão do Texto

Descritor: (9) Identificar a finalidade de textos de diferentes gêneros.

Total de respostas: 12943

Proporção do total sem peso: .06

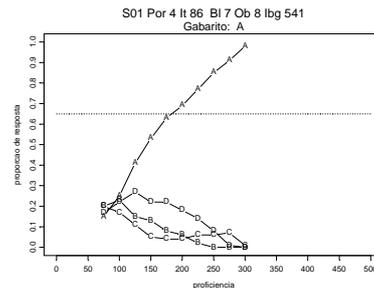
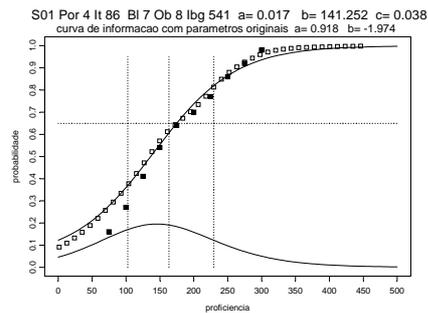
Proporção do total com peso: .05

Nível	75	100	125	150	175	200	225	250	275	300	325	350	375	400	425
Proporção nos intervalos	.19	.26	.34	.44	.50	.57	.63	.79	.94	.97	.94	.00	.00	.00	.00
Indivíduos que erraram	310	787	1222	1390	1255	911	522	165	37	10	2	0	0	0	0
Indivíduos que acertaram	76	275	671	967	1182	1088	947	612	353	121	40	0	0	0	0

8) A finalidade do texto é

- (A) informar sobre as funções dos micróbios.
- (B) divertir mostrando que os micróbios engordam.
- (C) recomendar mostrando que os micróbios são só amigos.
- (D) explicar como os micróbios são “coisa chata”.

ÍNDICES									PROPORÇÕES DE RESPOSTA						COEFICIENTES BISSERIAIS					
ITEM	BL	OB	GAB	DIFI	DISCR	ABAI	ACIM	BISE	A	B	C	D	""	""	A	B	C	D	""	""
86	7	8	A	.55	.51	.28	.79	.52	.55	.11	.08	.20	.04	.02	.52	-.33	-.23	-.14	-.44	-.70



Item: 86 Nível: 200

Tema: (2) Implicações do Suporte, do Gênero e/ou do Enunciador na Compreensão do Texto

Descritor: (9) Identificar a finalidade de textos de diferentes gêneros.

Total de respostas: 12878

Proporção do total sem peso: .15

Proporção do total com peso: .15

Nível	75	100	125	150	175	200	225	250	275	300	325	350	375	400	425
Proporção nos intervalos	.16	.27	.41	.54	.64	.70	.77	.86	.92	.98	1.00	.00	.00	.00	.00
Indivíduos que erraram	301	771	1113	1108	904	609	325	109	27	6	0	0	0	0	0
Indivíduos que acertaram	71	284	767	1234	1527	1385	1141	666	363	125	42	0	0	0	0