

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
CAEd – CENTRO DE POLÍTICAS PÚBLICAS E AVALIAÇÃO DA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO PROFISSIONAL EM GESTÃO E AVALIAÇÃO
DA EDUCAÇÃO PÚBLICA

FERNANDA GOMES DA SILVA

**A INCLUSÃO DE CIÊNCIAS NATURAIS NOS SISTEMAS DE AVALIAÇÃO DA
EDUCAÇÃO: O CASO DO AVALIA BH**

JUIZ DE FORA

2014

FERNANDA GOMES DA SILVA

**A INCLUSÃO DE CIÊNCIAS NATURAIS NOS SISTEMAS DE AVALIAÇÃO DA
EDUCAÇÃO: O CASO DO AVALIA BH**

Dissertação apresentada como requisito parcial à conclusão do Mestrado Profissional em Gestão e Avaliação da Educação Pública, pela Faculdade de Educação da Universidade Federal de Juiz de Fora.

Orientador(a): Lina Kátia Mesquita de Oliveira

JUIZ DE FORA

2014

TERMO DE APROVAÇÃO

FERNANDA GOMES DA SILVA

**A INCLUSÃO DE CIÊNCIAS NATURAIS NOS SISTEMAS DE AVALIAÇÃO DA
EDUCAÇÃO: O CASO DO AVALIA BH**

*Dissertação apresentada à Banca Examinadora designada pela equipe de
Dissertação do Mestrado Profissional CAEd/FACED/UFJF, aprovada em
29/09/2014.*

Membro da banca-orientador(a)

Membro da banca Externa

Membro da Banca Interna

Juiz de Fora, 2014

Àquela que me inculcou, desde muito antes da entrada na escola, a importância dos estudos para o progresso profissional e financeiro, mas, sobretudo, para a formação cidadã, reflexiva e humana. Entre tantos célebres educadores que tive contato no meio acadêmico, você, Mãe, sem títulos de mestrado ou doutorado, é a maior educadora que eu conheço.

AGRADECIMENTOS

Ao Geraldo Luiz, à Maria do Socorro, à Juliana e à Tábata por serem meus alicerces sempre.

À Cecília, por esperar, pacientemente, ou não, a tia Nanda acabar de estudar para poder brincar.

Ao Heverton, principalmente, pelo equilíbrio emocional que me trouxe em momento mais oportuno, mas também por sua dedicação em me acompanhar nas pesquisas de campo e pelo auxílio nas traduções dos textos em inglês.

Ao Kelmer que contribuiu de forma desmedida para esse trabalho com a presteza que ultrapassa a esperada da função de tutor para a solidariedade de um amigo.

À GAPED por ter me recebido, com toda a atenção, mesmo em dia de expediente atribulado e cedido a entrevista que fundamentou este trabalho.

Ao CAEd pelo incentivo à formação continuada em serviço.

RESUMO

Esta dissertação centra-se no estudo de caso do sistema de avaliação da educação pública pertencente à Secretaria Municipal de Educação da cidade de Belo Horizonte – MG, o Avalia BH, no que concerne à implementação de testes de proficiência em Ciências da Natureza para o Ensino Fundamental. São investigadas as motivações para tal implementação e os seus efeitos na gestão do próprio sistema, na gestão escolar, no desempenho estudantil e na prática docente. Conclui-se, a partir de entrevista junto ao órgão responsável pelo programa (GAPED - Gerência de Avaliação de Políticas Educacionais) e das análises de questionários aplicados a docentes e gestores, que a inclusão da aferição de competências científicas no referido sistema repercute em ações tímidas diante dos seus resultados; apresenta, ainda, incerteza a respeito de seus procedimentos e instrumentos de medição e necessita, portanto, de reformulações. Propõe-se, assim, como Plano de Ação Educacional, uma ampliação da investigação aqui realizada a partir da meta-avaliação, a fim de sistematizar os principais pontos a serem modificados e possibilitar a criação de estratégias para tornar a política pública compensatória ao ensino de Ciências.

Palavras-chave: Avaliação. Ciências Naturais. Avalia BH. Meta-avaliação.

ABSTRACT

This dissertation focuses on the case study of the evaluation system of public education pertaining to the Municipal Education Department of the city of Belo Horizonte-MG – called Avalia BH – concerning the implementation of proficiency tests in Natural Sciences. The motivations for such an implementation and its effects on the management of the system itself, the school management, student achievement and teacher practice are investigated. It is concluded, through interviews with the agency responsible for the program (Gerência de Avaliação de Políticas Educacionais – GAPED, in Portuguese) and by the analysis of questionnaires given to teachers and educational managers, that the inclusion of measuring scientific competencies on this system resonates timid actions before their results; it also presents uncertainty about its procedures and measurement instruments, therefore in need of reformulation. It is proposed as an Educational Action Plan, an extension of the research conducted here starting from the meta-evaluation in order to systematize the main points to be modified to make compensatory public policy to Science Teaching.

Keywords: Educational evaluation. Natural Sciences. Avalia BH. Meta-evaluation.

LISTA DE SIGLAS

Avaliando IDEPB – Sistema de Avaliação da Educação da Paraíba
CETPP – Centro de Estudos e Testes e Pesquisas Psicológicas
ECIEL - Estudos Conjuntos de Integração Econômica Latino-americana
FGV – Fundação Getúlio Vargas
IAEP – International Assessment of Educational Progress
NAEP – National Assessment of Educational Progress
PAEBES – Programa de Avaliação da Educação Básica do Espírito Santo
PISA - Programme for International Student Assessment
SABE – Sistema de Avaliação Baiano da Educação
SADEAM – Sistema de Avaliação do Desempenho Educacional do Amazonas
SAEB – Sistema Nacional de Avaliação da Educação Básica
SAEGO – Sistema de Avaliação da Educação de Goiás
SAEMS – Avaliação da Educação Básica do Mato Grosso do Sul
SAEP – Sistema de Avaliação da Educação Básica do Paraná
SAEPE – Sistema de Avaliação da Educação Básica de Pernambuco
SAEPI – Sistema de Avaliação Educacional do Piauí
SAERJ – Sistema de Avaliação da Educação do Estado do Rio de Janeiro
SAERO – Sistema de Avaliação Educacional de Rondônia
SAERS – Sistema de Avaliação do Rendimento Escolar do Rio Grande do Sul
SALTO – Sistema de Avaliação da Educação Tocantins
SARESP – Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo
SAVEAL – Sistema de Avaliação do Alagoas
SEAPE – Sistema Estadual de Avaliação da Aprendizagem Escolar
SIMADE – Sistema Maranhense de Avaliação do Desempenho
SIMAVE – Sistema Mineiro de Avaliação da Educação Pública
SPAECE – Sistema Permanente de Avaliação da Educação Básica do Ceará

LISTA DE GRÁFICOS

Gráfico 1 - Respostas dos docentes à pergunta: A oficina de divulgação dos resultados mudou sua relação com os dados de Ciências produzidos pelo Avalia BH?	72
Gráfico 2 - Respostas dos gestores em relação à relevância do conteúdo da oficina de divulgação de resultados de Língua Portuguesa, Matemática e Ciências do AVALIA BH 2012.....	74
Gráfico 3 - Respostas dos gestores quando perguntados sobre o que acharam do preparo oferecido pela oficina para a apropriação dos resultados de Língua Portuguesa, Matemática e Ciências dos testes do Avalia BH de 2012.	75
Gráfico 4 - Respostas dos gestores participantes de oficina de divulgação dos resultados de Língua Portuguesa, Matemática e Ciências de 2012 em relação ao objetivo dessa oficina.	76
Gráfico 5 - Respostas dos gestores em relação à qualidade do material oferecido na oficina de divulgação de resultados do Avalia BH 2012 para a compreensão dos dados de avaliação em larga escala em Língua Portuguesa, Matemática e Ciências	77
Gráfico 6 - Percentual de respostas dos gestores participantes de oficina de divulgação dos resultados de Língua Portuguesa, Matemática e Ciências do Avalia BH 2012 em relação aos esclarecimentos e explicações sobre dados de avaliação externa.	78
Gráfico 7 - Percentual de respostas dos gestores em relação às atividades e dinâmicas oferecidas na oficina de divulgação de resultados do Avalia BH 2012.....	80
Gráfico 8 - Percentual de respostas dos professores em relação a sua percepção das melhorias no processo de ensino aprendizagem de Ciências a partir da inclusão dessa disciplina nos testes do Avalia BH	82

LISTA DE QUADROS

Quadro 1 - Iniciativas pontuais para a verificação de competências científicas no Brasil anteriores ao SAEB	28
Quadro 2 - Escala de Proficiência de Ciências - PISA 2006	32
Quadro 3 - Escala de Proficiência de Ciências da Natureza - SAEB 1997	37
Quadro 4 - Padrão de Desempenho em Biologia no SADEAM 2012	42
Quadro 5 - Padrão de Desempenho em biologia dos estudantes baianos avaliados pelo SAE em 2011	44
Quadro 6 - Níveis de Proficiência em Ciências da Natureza - SARESP 2008	45
Quadro 7 - Padrão de Desempenho em Biologia dos estudantes capixabas - PAEBES 2011	47
Quadro 8 - Escala de Proficiência de Ciências da Natureza do Avalia BH.....	56
Quadro 9 - Padrões de Desempenho em Ciências dos avaliandos do 3º ano do Ensino Fundamental no AVALIA BH 2012	66
Quadro 10 - Padrões de Desempenho em Ciências dos avaliandos do 2º ciclo do Ensino Fundamental no AVALIA BH 2012	68
Quadro 11 - Padrões de Desempenho em Ciências dos avaliandos do 3º ciclo do Ensino Fundamental do AVALIA BH 2012	70
Quadro 12 - Chave de Avaliação para realização de meta-avaliações	94
Quadro 13 - Proposta de Chave de Meta-avaliação para as avaliações de Ciências do Avalia BH.....	95
Quadro 14 - Julgamento de aspectos da avaliação de Ciências do Avalia BH de acordo com padrões de qualidade	98

LISTA DE TABELAS

Tabela 1 - Sistemas de Avaliação Estaduais da Educação	39
Tabela 2 - Médias de Proficiência dos estudantes da rede pública estadual de São Paulo em Ciências da Natureza no SARESP de 2008, 2010 e 2012	45
Tabela 3 - Cortes na Escala de Proficiência que definem os padrões de desempenho por série	64
Tabela 4 - Médias de Proficiência, por série, em Ciências, dos estudantes da rede municipal de Belo Horizonte - MG nos ciclos do Avalia BH de 2011 e 2012	64

SUMÁRIO

INTRODUÇÃO	13
1. O AVALIA BH NO CONTEXTO DAS AVALIAÇÕES EM LARGA ESCALA DE CIÊNCIAS NATURAIS	16
1.1 Um breve histórico do ensino de Ciências no Brasil	17
1.2 A importância do ensino e da avaliação de Ciências da Natureza	21
1.3 Os primeiros movimentos rumo às avaliações em larga escala de competências científicas	25
1.4 A inclusão de exames de Ciências da Natureza nos programas atuais de avaliação em larga escala	29
1.4.1 O PISA como principal influência	29
1.4.2 O histórico da avaliação de competências científicas no SAEB	33
1.4.3 A inclusão de ciências nos programas estaduais de avaliação da educação	39
1.5 O AVALIA BH e o processo de inclusão dos testes de Ciências	48
1.5.1 Antecedentes	49
1.5.2 O Desenho do Sistema	51
1.5.3 A inclusão dos testes de proficiência em Ciências	54
1.6 Retomada do problema de pesquisa	57
2. O AVALIA BH SOB ANÁLISE: RESULTADOS E IMPLICAÇÕES DOS TESTES DE CIÊNCIAS	59
2.1 Percurso metodológico	59
2.2 A percepção dos docentes sobre a inclusão de ciências no AVALIA BH	60
2.3 A Repercussão da Inclusão e dos Resultados dos Testes de Ciências no Avalia BH	62
2.3.1 O desafio da definição dos pontos de cortes na Escala de Proficiência pela Gerência de Avaliação de Políticas Educacionais (GAPED)	62
2.3.2. Os padrões de desempenho e as médias de proficiência em Ciências	64
2.3.3 A divulgação dos resultados do Avalia BH em Ciências	71
2.3.4 Os usos dos resultados do Avalia BH em Ciências	81
2.4 Outros efeitos da aferição de Ciências no Avalia BH	85
2.4.1 A Matriz de Referência de Ciências e a sua utilização como ferramenta pedagógica: um efeito perverso do Avalia BH?	85

2.4.2 O efeito sobre a gestão do Avalia BH com a inclusão de Ciências	87
2.4.3 Conclusões da pesquisa.....	88
3. A META-AVALIAÇÃO COMO PLANO DE AÇÃO EDUCACIONAL PARA O AVALIA BH/CIÊNCIAS DA NATUREZA	90
3.1 O que é meta-avaliação?	90
3.2 Modelo de meta-avaliação para o Avalia BH	92
3.2.1 A escolha do meta-avaliador e a formalização da meta-avaliação	92
3.2.2 A definição dos padrões de qualidade	96
3.2.3 O cronograma da meta-avaliação.....	99
3.2.4 Os resultados da meta-avaliação.....	99
3.3 Considerações finais	99
REFERÊNCIAS.....	102

INTRODUÇÃO

Cada vez mais estados e municípios aderem às políticas de avaliação educacional em larga escala como resposta à demanda por transparência sobre a situação do ensino, criando seus próprios programas avaliativos, nos moldes do Sistema Nacional de Avaliação da Educação Básica – SAEB, implementado a partir de 1990. E, cada vez mais, estados e municípios que antes aferiam conhecimentos apenas nos campos da Linguagem e da Matemática, passam a aplicar testes de Ciências da Natureza para os seus estudantes. Em 2008, os estados do Amazonas, da Bahia e de São Paulo começam a aplicar testes padronizados de Biologia, Física e Química; em 2010, Belo Horizonte desponta como o primeiro município do país a avaliar Ciências desde o 3º até o 9º ano do Ensino Fundamental; em 2011, foi a vez do Espírito Santo e do Rio de Janeiro recriarem seus sistemas de avaliação incluindo a verificação das competências científicas¹; em 2012, segue essa tendência o estado do Ceará.

Como licenciada em biologia e analista de instrumentos de avaliação em Ciências Naturais, esse fato chamou-me a atenção: o que levou a essa corrida dos sistemas estaduais e municipais de avaliação em larga escala por abranger a avaliação de competências e habilidades científicas? Como tem se dado esse processo? Os programas têm atingido o seu objetivo? O que mudou no ensino e na aprendizagem de Ciências Naturais a partir da implementação de testes padronizados dessa disciplina para estudantes da rede pública no país?

O presente Plano de Ação Educacional surge dessas inquietações. Na impossibilidade de investigar todos os programas de avaliação sistêmica nacionais, o Avalia BH, pertencente à prefeitura de Belo Horizonte - MG, desponta como o mais adequado para a realização desta pesquisa pela sua proximidade em termos geográficos, o que facilita a obtenção de dados. Além disso, esse programa tornou-se atrativo pelo seu pioneirismo, em âmbito municipal, na inclusão de testes de Ciências. Outra justificativa se relaciona à constatação de que, decorridos 3 anos dessa inclusão, em levantamentos de literatura até o presente momento, não foram encontrados estudos comparativos ou qualquer tipo de avaliação formal que

¹ Competências científicas: no presente trabalho, serão entendidas como as relacionadas à compreensão de teorias e fenômenos científicos, ao reconhecimento de questões que podem e devem ser tratadas cientificamente e a aplicação do conhecimento científico no cotidiano.

questionem os métodos do Avalia BH ou que analisem seus efeitos sobre a gestão, a docência e o ensino.

Nesse sentido, este caso de gestão objetivou, de forma geral, analisar o Avalia BH no que diz respeito à implementação dos testes de proficiência em Ciências Naturais, e propor, ao final da investigação, uma meta-avaliação. De forma específica, estabeleceu-se como objetivos conhecer as motivações para a inclusão desses testes, bem como reconhecer os seus efeitos na gestão do sistema de avaliação, na gestão escolar, no desempenho estudantil e para os docentes.

Os dados para esta pesquisa foram obtidos por meio de entrevista semiestruturada com membros da Gerência de Avaliação de Políticas Educacionais – GAPED, órgão pertencente à prefeitura de Belo Horizonte, responsável pelo desenho, implementação e monitoramento do Avalia BH; questionários respondidos por um grupo de 23 docentes da rede pública municipal; questionários respondidos por gestores, coordenadores pedagógicos e técnicos educacionais; além da análise documental baseada, principalmente, em boletins de divulgação de resultados.

O primeiro capítulo situa as avaliações educacionais em larga escala no percurso histórico do ensino de Ciências no Brasil, destacando a sua crescente importância relacionada à capacidade de ampliar a visão desse ensino, identificar seus problemas primordiais e apontar caminhos para a resolução. O PISA – Programa Internacional de Avaliação de Alunos – é apresentado como a principal influência para a inclusão da aferição de competências científicas nos programas nacionais, os quais também são descritos, com o fim de trazer um panorama dos métodos avaliativos que determinaram o desenho do Avalia BH e conhecer, ainda que de forma pouco aprofundada, os resultados de desempenho em alguns estados brasileiros.

Antes de chegar ao foco do trabalho, esse capítulo inicial não poderia deixar de abordar o SAEB que, embora tenha descortinado a avaliação de Ciências a partir de 1999, deixou como herança ao Avalia BH, seus fundamentos e procedimentos metodológicos. Enfim, expõe-se o sistema de avaliação educacional em larga escala da prefeitura de Belo Horizonte no que concerne à sua iniciativa de voltar a atenção também para o ensino das Ciências Naturais, apresentando os antecedentes dessa iniciativa, o desenho do programa e os bastidores da implementação dos testes de

proficiência nessa disciplina e retomam-se os questionamentos relacionados à repercussão dessa implementação.

O segundo capítulo explicita os métodos utilizados para a obtenção dos dados nessa pesquisa e traz esses dados que visam responder às questões colocadas no capítulo anterior. São apresentados, analisados e discutidos a percepção docente sobre o processo de inclusão dos testes de Ciências, os desafios da gestão do Avalia BH, os resultados de desempenho nos três ciclos de aplicação dos referidos testes, a forma de divulgação desses resultados, seus usos e seus efeitos sobre a gestão escolar, o trabalho docente e o ensino. Chega-se à conclusão da pesquisa que dará suporte à proposição do plano de ação educacional.

O último capítulo propõe uma ampliação da análise realizada neste trabalho, um estudo sistematizado que avalie o programa de avaliação da educação de Belo Horizonte, isto é, uma meta-avaliação. Com base em artigo publicado por Daniel Stufflebeam (2001), em que descreve uma metodologia geral para a realização de meta-avaliações, são delineadas propostas para a meta-avaliação do Avalia BH.

1. O AVALIA BH NO CONTEXTO DAS AVALIAÇÕES EM LARGA ESCALA DE CIÊNCIAS NATURAIS

O presente capítulo se inicia com um conciso relato da história do ensino de Ciências no Brasil, no intento de mostrar que esse ensino tem crescido em importância, o que é ratificado pelas iniciativas de estados e municípios brasileiros de incluí-lo em suas políticas de avaliação educacional em larga escala, sobretudo a partir do final da década de 2000.

A seção seguinte, respaldada em diferentes autores, vem salientar que a despeito da relevância das competências científicas para a inserção na sociedade moderna, existe uma crise na formação dos estudantes em relação a esse campo do conhecimento. Dentre as alternativas para a superação dessa crise, a avaliação sistêmica é defendida como a habilitada a trazer dados empíricos, produzir indicadores e apontar os problemas de forma precisa para que a solução possa ser planejada e alcançada com algum sucesso.

A terceira seção descreve os primeiros movimentos rumo às avaliações de Ciências, tímidos, porém relevantes pela sua posição de inauguradores. A quarta seção apresenta o processo de inclusão dos testes de desempenho nessa disciplina nos principais programas de avaliação em larga escala nacionais, destacando a influência que receberam do Programa Internacional de Avaliação de Alunos (PISA) e do Sistema Nacional de Avaliação da Educação Básica (SAEB) e a influência que exerceram sobre o programa de avaliação da educação do município de Belo Horizonte, o Avalia BH.

Traçado o panorama da inclusão de Ciências nos sistemas avaliativos, a quinta seção advém para inserir o Avalia BH nesse contexto. Nesse momento, é tomado o foco do estudo com a apresentação do programa, seus antecedentes, seu desenho e o processo que culminou com o pioneirismo na implementação dos testes de Ciências para alunos do 3º ao 9º anos do Ensino Fundamental. A última seção é dedicada à retomada do problema da pesquisa, com o fim de preparar o leitor ao segundo capítulo.

1.1 Um breve histórico do ensino de Ciências no Brasil

As Ciências Naturais foram instituídas como disciplina no Brasil por meio do Decreto n. 19.890, que dispõe sobre a organização do ensino secundário, em 1931. No entanto, somente a partir da década de 50, essa disciplina foi crescendo em importância, sendo objeto de diversas transformações em função das influências internacionais e dos contextos político e econômico vigentes (KRASILCHIK, 2000).

Segundo Schwartzman (1989), no período que antecede à Segunda Guerra Mundial, a Ciência era predominantemente acadêmica, mantendo pouca relação com a tecnologia. Já no período em que se deflagra o conflito, de 1939 a 1945, ela é colocada a serviço das atividades militares, sendo responsável pelo desenvolvimento de radares e pela construção da bomba atômica, por exemplo. No pós-guerra, a ciência passa a ser associada à atividade produtiva e ao desenvolvimento econômico e social.

É nesse contexto que a União Soviética, em 1957, lança o satélite Sputnik. Esse episódio abalou a supremacia tecnológica dos Estados Unidos, que atribuem a perda da superioridade científica ao seu sistema educacional e passam, então, para se manterem em competitividade, a aplicar grande volume de recursos financeiros na formação de professores e na expansão de programas de ensino superior em Ciências (BROOKE, 2012). Porém, o maior investimento foi na reforma curricular com a criação de projetos de materiais didáticos, hoje denominados de 1ª Geração do Ensino de Física, Química, Biologia e Matemática para o Ensino Médio (KRASILCHIK, 2000), que foram distribuídos em países de sua influência cultural, inclusive o Brasil.

Segundo Brooke (2012), esses materiais didáticos consistiam em livros-textos que apresentavam a ciência como um processo contínuo, enfatizando as ideias principais, mas permitindo aos estudantes o amadurecimento dessas ideias e até mesmo a sua invalidação. Possibilitava a redescoberta do conhecimento científico e incitava a iniciativa própria para pesquisas. Krasilchik (2000) afirma que a elaboração desses projetos contou com a participação intensa de sociedades científicas, universidades e acadêmicos renomados e, até hoje, são reconhecidos mundialmente pelas suas siglas: PSSC (*Physical Science Study Committee*), CBA (*Chemical Bond Approach*) e BSCS (*Biological Science Curriculum Study*).

No Brasil, os livros-textos advindos das fundações americanas pouco foram aproveitados pelos professores. Calhoun & Joyce (1996) *apud* Brooke (2012) atribuem esse fato à violação de um dos princípios do profissional da educação: o seu envolvimento no processo de produção dos materiais que irá utilizar. Porém, a confiança na reforma do currículo de Ciências como propulsora da modernização e do avanço econômico repercutiu nas políticas nacionais.

Em 1959, educadores brasileiros publicam o Manifesto dos Educadores Democratas em Defesa do Ensino Público. Nesse documento, eles relatam suas preocupações em reestruturar o ensino de acordo com o progresso que a sociedade conclamava: fazia-se necessário abandonar o ensino da Ciência pura, baseado no conhecimento e memorização dos experimentos realizados por cientistas, e torná-lo mais aplicável para a formação de mão-de-obra compatível com as tecnologias do mundo industrializado. Era o nacional-desenvolvimentismo, governo de Juscelino Kubitschek, onde se apregoava a necessidade do desenvolvimento de cinquenta anos em cinco (SANFELICE, 2007).

É nesse cenário, também, que é promulgada a Lei 4.024 – Diretrizes e Bases da Educação de 1961, que amplia sobremaneira a participação da Física, Química e Biologia no, então, curso colegial, correspondente ao atual Ensino Médio. Para Krasilchik (2000), as ciências passavam a ter a função de desenvolver cidadãos críticos e capazes de tomar decisões com base em informações científicas.

Na década de 1970, porém, o currículo de Ciências se modifica, no Brasil, junto com as transformações ocasionadas na política e na economia pelo golpe militar de 1964. O sistema ditatorial fomentará os investimentos em educação técnica e em habilidades voltadas para o mercado de trabalho, na intenção de criar atitudes condizentes com o modelo capitalista em ascensão no período. Além disso, encaminhar os estudantes para um destino laboral no ensino secundário desviaria a sua atenção do ensino superior, resolvendo o problema da falta de vagas nas universidades (BROOKE, 2012).

A Lei nº 5692 de 1971, consonantemente, determina que a parte de formação especial do currículo escolar

- a) terá o objetivo de sondagem de aptidões e iniciação para o trabalho, no ensino de 1º grau, e de habilitação profissional, no ensino de 2º grau;
- b) será fixada, quando se destina a iniciação e habilitação profissional, em consonância com as necessidades do mercado de

trabalho local ou regional, à vista de levantamentos periodicamente renovados. (§ 2º, Art. 5º)

O currículo de Ciências, então, passa a ter, obrigatoriamente, caráter tecnicista e profissionalizante, deixando de enfatizar a cidadania e a participação. Zotti (2004) afirma que a educação intentaria tornar-se um braço do capitalismo, no sentido de servir às suas necessidades e cita Gramsci (*apud* GERMANO, 1994), concordando com esse autor que a escola que se formava, do tipo profissional, no lugar da formativa, objetivava satisfazer interesses práticos imediatos, destinando-se, não somente, a perpetuar como cristalizar as desigualdades sociais.

Já na década de 1980, a redemocratização do país trará o movimento da educação para a ciência, conhecido como CTS – Ciência, Tecnologia e Sociedade. Influenciado pelas tendências progressistas da Educação, esse movimento passou a questionar a abordagem e a organização dos currículos, sugerindo uma transformação que os tornassem mais significativos para os estudantes, próximos do seu cotidiano e integrados a outros conteúdos. Também nesse período, as pesquisas acadêmicas iniciaram uma aproximação da educação científica às ciências humanas e sociais. Essa nova concepção de educação para a ciência considera, como relevantes, as ideias e os conhecimentos prévios dos estudantes independentes do ensino formal (DRIVER *et al*, *apud* MORTIMER, 1996).

Na década de 1990, as discussões sobre o ensino de Ciências incorporam o discurso da formação de um cidadão consciente, participativo e crítico. Fazia-se necessário desenvolver nos estudantes o pensamento reflexivo que levasse ao questionamento das relações existentes entre ciência, tecnologia, meio ambiente e sociedade (NASCIMENTO JÚNIOR *et al*, 2010). Valorizava-se o estudo de conteúdos científicos com aplicabilidade na vida extraescolar. Nesse contexto, em 1996, foi aprovada a nova LDB (9394/96), a qual estabelece que a educação escolar deve vincular-se ao mundo do trabalho e à prática social.

Em 1997, são publicados as Diretrizes Curriculares Nacionais para a Educação Básica e os Parâmetros Curriculares Nacionais – PCN, documentos produzidos pelo Ministério da Educação com o fim de servirem de base para a formulação de currículos estaduais e municipais. Nos PCN de Ciências da Natureza e suas Tecnologias, preconiza-se um currículo capaz de formar estudantes que se percebam integrantes, dependentes e transformadores do ambiente; contribuam

para a melhoria do meio ambiente; conheçam e cuidem do próprio corpo, valorizando e adotando hábitos saudáveis; responsabilizem-se pela saúde individual e coletiva e utilizem diferentes fontes de informação e recursos tecnológicos para a construção de conhecimentos (BRASIL, 1997).

Surgem, então, as temáticas, a serem estudadas de forma interdisciplinar, da poluição, do consumo consciente, da produção de lixo, da economia dos recursos naturais, da prevenção de doenças, entre outras (WALDHELM, 2007). É nesse período que emergem o conceito de competência e a crítica ao ensino considerado conteudista, aquele que apresenta excesso de informações, muitas vezes desnecessárias, descontextualizadas e pouco significativas para os estudantes.

O Parecer CEB/CNE nº 15/98, que dispõe sobre as diretrizes curriculares de 1997, define o currículo voltado para competências: é aquele que valoriza o aprender a aprender, a pensar e a relacionar o conhecimento com a experiência cotidiana, isto é, fazer a ponte entre a teoria e prática. Esse currículo destaca a importância da fundamentação da crítica, da argumentação baseada em fatos e do abandono do conhecimento enciclopédico. Procura-se a construção de habilidades em detrimento do acúmulo de informações.

É, também, na década de 1990, que tem início o movimento de avaliação educacional em larga escala no Brasil com o Sistema Nacional de Avaliação da Educação Básica – SAEB, sob coordenação do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP. O SAEB avalia o desempenho, bianualmente, dos estudantes das 5ª e 9ª séries do Ensino Fundamental e do 3º ano do Ensino Médio. Até o ciclo avaliativo de 1999, o programa incluía testes de Ciências da Natureza. No ciclo de 2001, porém, ocorreu uma reformulação do SAEB “visando à definição, revisão e construção dos instrumentos que integraram o sexto ciclo de avaliação” (BRASÍLIA, 2002). Dentre as modificações ocorridas, documento publicado pelo INEP, em 2002, intitulado SAEB 2001: Novas Perspectivas, explica que

Construiu-se [...] um modelo de prova que passou a priorizar determinados Tópicos e Temas dos conteúdos a serem avaliados em Língua Portuguesa e Matemática, respectivamente, em função das competências e habilidades definidas para as diferentes séries e disciplinas. Foram atribuídas, a seguir, diferentes prioridades aos Tópicos e Temas, o que, na prática, significa alocar um número

maior ou menor de itens a cada descritor² diante de sua importância na série e disciplina. (pag.7)

No entanto, esse documento não explica o porquê de ter removido os exames de Ciências. Waiselfisz (2009) atribui esse fato “à política de focalização no fundamental da educação fundamental: o ensino de Língua Portuguesa e de Matemática” (p. 10) surgida na segunda metade da década de 1990.

A década de 2000 vê o fortalecimento das políticas de avaliação, na medida em que estados e municípios começam a criar seus próprios sistemas de monitoramento da situação educacional. A princípio, a exemplo do SAEB, a maioria dos programas se concentrava em medir a proficiência em Língua Portuguesa e Matemática. Atualmente, porém, cada vez mais estados e municípios incluem ciências nos seus testes de rendimento escolar: em 2008, Amazonas, Bahia e São Paulo; em 2010, o município de Belo Horizonte; em 2011, o Espírito Santo; em 2012, o Ceará. O próprio SAEB, em 2013, no mês de novembro, retorna, em caráter experimental, a apurar habilidades em ciências para alunos do 9º ano do Ensino Fundamental e 3º ano do Ensino Médio (BRASIL, 2013).

1.2 A importância do ensino e da avaliação de Ciências da Natureza

Essa seção objetiva trazer um respaldo teórico para a discussão que se seguirá a respeito do advento das avaliações em larga escala de competências científicas. Argumenta-se, aqui, a favor desse tipo de política pública como uma possibilidade de ampliar a visão do ensino de ciências e apontar caminhos para a sua melhoria, dada a sua importância, sobretudo, nos dias atuais.

Keti Tenenblat *et al* (2007), em trabalho produzido no âmbito da Academia Brasileira de Ciências – ABC, defendem que a ciência e a tecnologia constituem recursos essenciais para o desenvolvimento econômico, o manejo sustentável dos recursos naturais, a utilização de políticas públicas adequadas na área de saúde e a gestão da urbanização. Esses pesquisadores argumentam, ainda, que além da sua aplicabilidade no cotidiano, a alfabetização científica é imprescindível na formação da cidadania em sociedades modernas, capacitando indivíduos a tomarem decisões próprias a respeito do mundo natural.

² Descritor: constitui a descrição de uma habilidade esperada ao final de cada período de escolarização. O conjunto de descritores forma a Matriz de Referência.

O documento que orienta o currículo da educação básica no Brasil, Parâmetros Curriculares Nacionais – PCN, complementa a justificativa da importância das ciências biológicas, físicas e químicas para a civilidade, alegando que o conhecimento nesses âmbitos permite a compreensão dinâmica da vivência, o convívio harmônico com o mundo da informação, o entendimento histórico da vida social e produtiva e a percepção evolutiva da vida, do planeta e do cosmos (BRASIL, 1996).

Mas é Waiselfisz (2009) quem melhor sustenta essa tese, destacando que o século XXI inaugura a Revolução do Conhecimento, em que a educação em ciências ganha posição de relevo, pois dela depende

a compreensão e o enfrentamento de questões sumamente relevantes da atualidade, entre elas novas doenças (Gripe A, aids), desafios (mudanças climáticas) e até ameaças (sofisticadas armas de destruição em massa). (p.7)

No entanto, a despeito da sua relevância na oportunização do acesso ao saber científico, das intensas discussões, do surgimento de diversas correntes e da produção de documentos oficiais com vistas à melhoria das suas práticas, o ensino de ciências nas escolas brasileiras, como afirmam diversos autores (FOUREZ, 2003; CACHAPUZ *et al*, 2005; BIZZO, 2004; SANTOS, 2007), tem fracassado em seu propósito. Gérard Fourez (2003) explica que o problema é universal e está no fato de o ensino de ciências querer obrigar os estudantes a enxergarem o mundo pelos olhos dos cientistas, enquanto o que teria sentido para eles seria um ensino que proporcionasse a compreensão do seu próprio mundo. Esse mesmo autor também atribui o fracasso à formação dos professores que “esteve mais centrada sobre o projeto de fazer deles técnicos de ciências do que de fazê-los educadores” (p. 111).

Cachapuz *et al* (2005) corrobora esse pensamento, considerando que a abordagem da disciplina de Ciências ainda constitui mera reprodução de saberes de forma neutra, distanciados do cotidiano do estudante e que não têm contribuído para o desenvolvimento de uma visão crítica, gerando, antes, uma visão distorcida da Ciência, seus métodos e produtos, o que leva ao desinteresse e passividade frente aos fenômenos naturais e tecnológicos. Santos (2007) afirma que as estratégias didáticas que predominam, hoje, nas escolas, limitam-se a um processo de memorização de vocábulos, de classificações e fórmulas, fazendo com que os estudantes aprendam termos científicos, mas sejam incapazes de extrair o significado deles. Por fim, Tenenblat *et al* (2007) afirmam que a crise da formação

científica dos alunos brasileiros participa de um problema ainda mais amplo, que é a precariedade da educação básica no Brasil como um todo.

A consequência dessa crise, conforme Nelio Bizzo (2004), repercute na economia e legítima desigualdades. Segundo esse autor, diante da globalização crescente, o quadro brasileiro se agrava pelo fato de existir um desencontro entre a preparação de mão de obra nacional e as demandas tecnológicas atuais, o que reflete nas decisões concernentes ao tipo de investimentos que o Brasil atrai. Ademais, continua Bizzo, um país que não consegue formar cidadãos inventivos, termina por consumir a inventividade externa e “essa inventividade de outros países reproduz, de forma avassaladora, a concentração de riqueza, traduzida em concentração internacional do saber” (2004, p. 9).

No plano social, Waiselfisz (2009, p.8) se vale dos resultados de desempenho no Programa Internacional de Avaliação de Alunos – PISA, de 2006, ano em que o foco dos testes foi a aferição de habilidades científicas, para evidenciar que “mais de 60% dos alunos brasileiros não apresentam competência suficiente na área de ciências para lidar com as exigências e os desafios mais simples da vida cotidiana atual”.

Nesse sentido, torna-se imperiosa a necessidade de concentrar esforços para a melhoria da qualidade do processo de ensino e aprendizagem de Ciências da Natureza no Brasil. Na literatura especializada, são encontradas diversas proposições e sugestões que visam tal melhoria: revisão curricular, valorização e capacitação dos professores, readequação de metodologias e práticas de ensino, investimentos em laboratórios nas escolas, parcerias interinstitucionais, ampliação da carga horária das disciplinas de Ciências, Biologia, Física e Química, entre outras (BIZZO, 2004; ABREU, 2008; TENENBLAT *et al*, 2007).

Aliada a essas proposições, surge a avaliação educacional em larga escala definida pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP – como um instrumento capaz de

oferecer subsídios à formulação, à reformulação e ao monitoramento de políticas públicas e programas de intervenção ajustados às necessidades diagnosticadas; identificar problemas e diferenças regionais na educação básica; produzir informações sobre os fatores do contexto socioeconômico, cultural e escolar que influenciam o desempenho dos estudantes; proporcionar aos agentes educacionais e à sociedade uma visão dos resultados dos processos de ensino e aprendizagem e das condições em que são desenvolvidos; desenvolver competência técnica e científica na área de avaliação

educacional, ativando o intercâmbio entre instituições educacionais de ensino e pesquisa; produzir informações sobre o desempenho dos estudantes, assim como sobre as condições intra e extraescolares que incidem sobre o processo de ensino e aprendizagem, no âmbito das redes de ensino e unidades escolares; avaliar a qualidade, a equidade e a eficiência dos sistemas e redes de ensino brasileiras (BRASILIA, 2013, p. 12).

Dessa forma, as políticas de avaliação externa da educação que, desde 1990, com a implementação do Sistema Nacional de Avaliação da Educação Básica – SAEB – estiveram voltadas para o desempenho estudantil em Língua Portuguesa e Matemática, passam, cada vez mais, a serem vistas como uma alternativa para assistir o ensino de ciências também. Bizzo (2004) critica o fato de o programa nacional deixar de divulgar dados de desempenho estudantil em Ciências a partir de 2000 e comemora o retorno dos testes dessa disciplina, anunciado pelo INEP, diante da independência que ocasionarão em relação aos indicadores internacionais.

Colenci Jr. *et al* também reclamou a falta de um indicador interno que mostrasse o nível de instrução dos alunos brasileiros em ciências: “sem medir, não temos como verificar como estamos e, portanto, corrigir possíveis erros em qualquer planejamento realizado” (2011, p. 9). Esse educador vai mais além em seus posicionamentos, sugerindo, de forma incisiva, que o Índice de Avaliação da Educação Básica – IDEB – o qual, atualmente, cruza dados de desempenho em Linguagem e Matemática com os das taxas de aprovação escolar, inclua a aferição de competências relacionadas às ciências:

O indicador estabelecido internamente pelo governo, o IDEB, só mede o desempenho de nossos estudantes na leitura/escrita e matemática, sem qualquer avaliação em ciências. Será que esta não é importante no planejamento da educação brasileira? Por que não realizar essa avaliação também pelo próprio IDEB? Ou devemos abdicar de qualquer avaliação que nos mostre o que infelizmente não queremos ver e enfrentar: que a educação das ciências no ensino fundamental está em segundo plano por vários motivos, desde a falta de instalações educacionais condizentes, professores capacitados e motivados e um plano didático bem elaborado e prático para ser aplicado? (p. 9).

A necessidade de investigar o ensino de ciências por meio de avaliações sistêmicas, pondera Waiselfisz (2009), não está fundamentada na ausência de trabalhos sobre o tema no Brasil, mas no fato de que esses estão focados em aspectos epistemológicos, metodológicos ou conceituais e, assim, precisam ser

acrescentados de dados empíricos referentes às variáveis de ensino-aprendizagem, como sala de aula, professor, aluno e aproveitamento, a fim de se obter um panorama mais abrangente da realidade.

Essa necessidade também tem sido reconhecida, ainda que tardiamente, por alguns estados e municípios, os quais tomaram a iniciativa de incluir em seus programas e projetos de avaliação educacional, a partir do final da década de 2000, testes de proficiência na área de Ciências da Natureza, como é o caso do Amazonas, Bahia, Espírito Santo, São Paulo, Ceará e Rio de Janeiro, bem como do município de Belo Horizonte - MG. Nesse contexto, é mister afirmar que embora ainda exista a priorização da verificação de competências em Língua Portuguesa e Matemática, justificada, muitas vezes, por tratarem de disciplinas suportes para a aprendizagem de outras áreas e, ainda, por existir uma experiência acumulada nesse sentido (CUETO, 2009³), existe um movimento atual de valorização do ensino de ciências.

Tenenblat *et al* (2007) endossam que esse movimento é um avanço importante e chamam a atenção para o aperfeiçoamento da qualidade técnica dos indicadores, afirmando que as avaliações em larga escala devem ser claramente inteligíveis para gestores educacionais, gestores escolares, docentes e pais dos alunos; devem estar vinculadas aos programas de ensino e os seus resultados devem ser capazes de conduzir a ações específicas que corrijam deficiências encontradas e melhorem, sistematicamente, o padrão de qualidade das escolas e do ensino.

1.3 Os primeiros movimentos rumo às avaliações em larga escala de competências científicas

Como sinalizado na seção anterior, os programas de avaliações de Ciências em larga escala emergiram, no Brasil, a partir da segunda metade da década de 2000. Como políticas de Estado e não de governo, esses programas tendem a se consolidar, aferindo, anualmente, o desempenho estudantil, produzindo dados comparáveis e estabelecendo metas. Antes dessa década, porém, surgiram algumas iniciativas pontuais para a verificação dos conhecimentos científicos. Esta seção

³ Santiago Cueto, em entrevista concedida ao Portal Aprendiz, sobre priorização do ensino de Língua Portuguesa e Matemática, disponível em: http://www.educacionista.org.br/jornal/index.php?option=com_content&task=view&id=4525&Itemid=49. Acesso em: 25 maio 2013.

versa sobre essas iniciativas, algumas das quais, serviram de molde ou influenciaram de alguma forma o desenho dos programas de avaliação atuais.

Bernadete Gatti (2009) aponta que a primeira pesquisa relativamente ampla para verificação de competências e sua relação com diferentes variáveis (sexo, nível socioeconômico, entre outras) ocorreu em 1966 com a criação do Centro de Estudos e Testes e Pesquisas Psicológicas (CETPP) da Fundação Getúlio Vargas (FGV). Esse Centro desenvolveu exames educacionais para as últimas séries do Ensino Médio, que constituíam um questionário sobre aspectos socioeconômicos e provas objetivas de Linguagem, Matemática, Estudos Sociais e Ciências Físicas e Naturais. No entanto, os testes não foram aplicados na rede pública de ensino, servindo apenas para processos seletivos de universidades e concursos para cargos públicos.

Ainda segundo Gatti (2009), em meados da década de 1970, o Programa de Estudos Conjuntos de Integração Econômica Latino-americana (ECIEL) desenvolveu um estudo avaliativo no Brasil e em outros países da América Latina sobre os fatores determinantes da escolaridade e do rendimento de estudantes com diferentes características pessoais e socioeconômicas. Os alunos responderam um questionário contextual e testes de compreensão em Leitura e Ciências.

Em 1987, o Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), pertencente ao Ministério da Educação (MEC), pretendeu criar um programa de avaliação externa do, então, 1º grau, hoje, Ensino Fundamental. Esse programa, denominado Avaliação do Rendimento em Escolas de 1º Grau da Rede Pública, avaliou amostras de alunos em 69 cidades, localizadas em todos os estados brasileiros, quanto aos conhecimentos desenvolvidos em Língua Portuguesa e Matemática para as 1ª e 3ª séries e além dessas disciplinas, Ciências para os alunos de 5ª e 7ª séries (GATTI, 1994).

É interessante destacar que, como Gatti (1994) relata, os programas utilizados para balizar a avaliação referida foram definidos consensualmente por professores do Ensino Fundamental e que os instrumentos avaliativos foram estudados em relação aos coeficientes de fidedignidade e discriminação média. Além disso, as provas apresentaram dificuldade mediana, o que concorreu para resultados consistentes e permitiu a definição de diferentes níveis de desempenho, facilitando a identificação de pontos críticos no quadro aprendizagem dos alunos de diversas séries (GATTI, 1994). Esse fato demonstra, portanto, o início de uma

preocupação para além do avaliar, mas de como avaliar produzindo dados confiáveis.

Em 1988, a Secretaria do Estado da Educação do Paraná dá continuidade à iniciativa do MEC, realizando uma avaliação dos alunos de 2ª e 4ª séries em relação aos conteúdos de Língua Portuguesa, Matemática, Estudos Sociais e Ciências. Os testes foram aplicados em amostras de estudantes de 41 escolas em sete cidades do estado (GATTI, 1994).

Sobre essas iniciativas do MEC e do Paraná, Gatti (2009) afirma que

trouxeram o impacto dos baixos resultados médios, muito aquém do esperado, que repercutiram tanto no Ministério, como nas Secretarias de Educação e na mídia, criando nas administrações públicas interesse pelos processos avaliativos. Com os resultados obtidos e dados das escolas, dos professores e dos alunos, dispunha-se de muitas informações, sobre as quais se poderiam debruçar, refletir e tirar inferências tanto para políticas dirigidas às redes de ensino como um todo, como para questões da aprendizagem de alunos nas salas de aula. Estes primeiros estudos mais abrangentes de alunos e escolas serviram de base para a implantação do SAEB – Sistema de Avaliação da Educação Brasileira. (p.11)

Além dessas avaliações de âmbito nacional e estadual, merecem destaque uma iniciativa internacional e uma municipal incluindo Ciências. De acordo com Vianna (1992), em 1990, o Brasil participou do *International Assessment of Educational Progress (IAEP)*, projeto relacionado ao *National Assessment of Educational Progress (NAEP)*, programa americano que, desde 1969, avalia, anualmente, amostras de estudantes das séries correspondentes aos 5º e 9º anos do Ensino Fundamental e 3º ano do Ensino Médio brasileiros em relação ao rendimento em leitura, escrita, matemática, ciências naturais, artes, cidadania, economia, geografia e história.

O *IAEP* foi uma pesquisa sobre o desempenho em Ciências e Matemática de uma amostra de estudantes de 13 anos selecionados em escolas das cidades de São Paulo e Fortaleza. As escolas também foram analisadas quanto à estrutura física, enquanto os gestores e professores foram questionados a respeito de sua formação, a relação com família dos alunos e as práticas de ensino. O objetivo do projeto, aplicado, também, em mais 19 países, foi o de identificar ambientes culturais e práticas educacionais associadas a um alto desempenho (VIANNA, 1992). Gatti (2009), sobre essa avaliação, conclui que apesar de ter sido realizada em apenas

duas capitais, já pode mostrar o precário desempenho do Brasil em relação a outros países: penúltima colocação.

Por fim, de acordo com Gatti (2009), apenas uma municipalidade, a de São Paulo, preocupou-se com um estudo avaliativo de sua rede de ensino no início dos anos 1980. O Departamento de Planejamento da Secretaria Municipal de Educação da Cidade de São Paulo avaliou todos os alunos do terceiro estágio de Educação Infantil, os de 1º, 3º, 5º, 7º anos do Ensino Fundamental e os do 1º ano do Ensino Médio através de provas que abrangeram Língua Portuguesa, Matemática e Ciências. O objetivo era o de produzir dados e a partir deles, planejar as atividades escolares e extraescolares. No entanto, Gatti (2009) aponta que

tendo havido mudança de administração em 1982, o estudo ficou com seus resultados sem utilização, uma vez que os novos dirigentes não se interessaram em utilizar os resultados obtidos e levar adiante o processo iniciado dois anos antes. (p.9)

Dessa forma, considera-se que essas avaliações foram tímidas, mas nem por isso irrelevantes. Muito pelo contrário, elas abriram precedentes e inauguraram a ideia de que não só é possível como necessária a avaliação de competências científicas, ideia essa que será seguida por estados e municípios em processo que será descrito na próxima seção.

O quadro a seguir resume essas primeiras iniciativas.

Quadro 1 - Iniciativas pontuais para a verificação de competências científicas no Brasil anteriores ao SAEB

Ano	Órgão responsável	Disciplinas avaliadas	Público-alvo
1966	FGV	Linguagens, Estudos Sociais, Matemática, Ciências Físicas e Naturais	Candidatos ao ingresso em universidades e cargos públicos.
1970	ECIEL	Leitura e Ciências	Alunos no Brasil e na América Latina
1980	SME/SP	Língua Portuguesa, Matemática e Ciências	Alunos do terceiro estágio de Educação Infantil, os de 1º, 3º, 5º, 7º anos do Ensino Fundamental e os do 1º ano do Ensino Médio
1987	INEP	Língua Portuguesa, Matemática e Ciências	Amostras de alunos de 69 cidades de todos os estados brasileiros (1º, 3º, 5º e 7º séries)
1988	SEE/PA	Língua Portuguesa, Matemática, Estudos Sociais e Ciências	Amostras de alunos de 41 escolas em sete cidades do Paraná (2ª e 4ª séries)
1990	NAEP	Ciências e Matemática	Amostras de estudantes de 13 anos de São Paulo e Fortaleza

1.4 A inclusão de exames de Ciências da Natureza nos programas atuais de avaliação em larga escala

Esta seção busca inserir o Avalia BH no contexto da corrida pela avaliação em larga escala de competências científicas, iniciada no final da década de 2000. Inicia-se com a caracterização do *Programme for International Student Assessment* - PISA, sobre o qual não faria sentido discorrer, se não tivesse exercido e ainda exercesse tanta influência sobre as políticas de avaliação educacional brasileiras.

O SAEB, embora tenha retirado os exames de Ciências da Natureza dos seus ciclos de avaliação, também será descrito, justificado pelo fato de ser o parâmetro nacional para a elaboração de metodologias e instrumentos de avaliação do desempenho. Seguindo uma lógica hierárquica, parte-se para o tratamento dos programas estaduais e municipais, confluindo para o objeto de estudo, o Avalia BH.

1.4.1 O PISA como principal influência

O *Programme for International Student Assessment* foi lançado, oficialmente, em 1997 pela *Organisation for Economic Co-operation and Development* – OECD com o objetivo de avaliar os sistemas de ensino em todo o mundo, testando habilidades e o conhecimento de estudantes de 15 anos de idade⁴, faixa etária em que se pressupõe o término da escolarização básica obrigatória na maioria dos países (BRASÍLIA, 2001).

Aplicados de três em três anos, desde 2000, os exames do PISA avaliam conhecimentos nas áreas de Leitura, Matemática e Ciências, abordando com maior enfoque uma dessas disciplinas em cada ciclo de aplicação. Assim, a cada triênio, dois terços da prova são dedicados à avaliação de um domínio principal, enquanto o outro terço visa apenas a propiciar um perfil básico das habilidades dos avaliandos (BRASÍLIA, 2001).

O Brasil participa do PISA desde a sua primeira edição e, conforme relatório produzido em 2001 pelo INEP, órgão administrador do programa no país, a sua implementação teve os seguintes propósitos:

- 1) Obter informações para situar o desempenho dos alunos brasileiros no contexto da realidade educacional, nacional e internacional.

⁴ Informação obtida na página da OECD na internet, disponível em: <<http://www.oecd.org/pisa/aboutpisa/>>. Acesso em: 31 out. 2013.

- 2) Fomentar a discussão sobre indicadores de resultados educacionais adequados à realidade brasileira.
- 3) Participar das discussões sobre as áreas de conhecimento avaliadas pelo Pisa em fóruns internacionais de especialistas.
- 4) Promover a apropriação de conhecimentos e metodologias na área de avaliação educacional.
- 5) Disseminar as informações geradas pelo Pisa, tanto em termos de resultados quanto em termos de conceitos e metodologias, entre diversos atores do sistema educacional, governamentais e não-governamentais. (p.22)

É importante ressaltar, entretanto, que o PISA não funciona somente como um gerador de dados sobre a educação brasileira, mas como um molde para as políticas de avaliação da educação. Existe uma grande aproximação entre as metodologias, proposições e tendências desse programa e as encontradas nos programas nacionais (TROJAN *et al*, 2012). O conceito de *scientific literacy*, por exemplo, muito presente nos discursos brasileiros sobre o que se espera de um estudante em relação às habilidades no campo das ciências foi difundido a partir do PISA.

Scientific literacy, traduzido, comumente, como letramento científico, designa

a capacidade de empregar o conhecimento científico para identificar questões, adquirir novos conhecimentos, explicar fenômenos científicos e tirar conclusões baseadas em evidências sobre questões científicas. Também faz parte do conceito de letramento científico a compreensão das características que diferenciam a ciência como uma forma de conhecimento e investigação; a consciência de como a ciência e a tecnologia moldam nosso meio material, cultural e intelectual; e o interesse em engajar-se em questões científicas, como cidadão crítico capaz de compreender e tomar decisões sobre o mundo natural e as mudanças nele ocorridas. (OECD, 2006, p.12)

De forma mais clara, trata-se do uso social do conhecimento científico, isto é, da capacidade de relacionar as ciências e a tecnologia ao cotidiano, como nas decisões sobre questões envolvendo saúde, energia, alimentação, recursos naturais, ou em situações práticas como na diluição de domissanitários, compreensão de uma bula de remédio, adoção de profilaxia de doenças, entre outras (SANTOS, 2007).

Veremos a admissão desse conceito em diversos programas de avaliação subnacionais e, inclusive, no nacional, o SAEB, que retorna à avaliação de ciências em 2013. Documento intitulado Inclusão de Ciências no SAEB, divulgado pelo INEP em 2013, especifica o que deve ser objetivo do ensino de Ciências no Brasil e, portanto, o que deve ser avaliado, claramente a concepção de letramento científico:

os objetivos da educação em Ciências da Natureza devem ultrapassar a ideia de uma transmissão de conhecimentos, avançando no sentido da apropriação de conhecimentos socialmente valorizados e construídos a partir de métodos próprios da ciência. Devem avançar para a apropriação de ferramentas que possibilitem a construção e o uso de conhecimentos científicos escolares e a compreensão dos métodos a partir dos quais o conhecimento científico é produzido. Dar-se-á desde um nível de reconhecimento de fenômenos e processos científicos que ocorrem em situações vividas pelo indivíduo até um nível de tomada de decisões, no âmbito da ciência e da tecnologia, que envolvam interesses de diferentes grupos sociais e reflexões sobre suas possíveis consequências. (p.20)

Também as metodologias utilizadas pelo PISA são acolhidas pelos sistemas nacionais de avaliação. O PISA utiliza como modelo estatístico a Teoria de Resposta ao Item – TRI. Esse modelo surgiu da necessidade de vencer a dificuldade de comparar resultados de diferentes testes em diversas situações e as limitações da apresentação de resultados apenas por percentuais de acertos (KLEIN *apud* SOUZA, 2005).

Klein *apud* Souza (2005) explica que a

TRI é um conjunto de modelos matemáticos onde a probabilidade de resposta a um item é modelada como função da proficiência (habilidade) do aluno (variável latente, não observável) e de parâmetros que expressam certas propriedades dos itens. (p. 121)

Oliveira (2008) observa que a TRI⁵ apresenta fortes pressupostos quanto ao comportamento de um indivíduo que responde aos itens de um teste, o que confere a esse modelo, as vantagens de permitir a comparação longitudinal de resultados de diferentes testes; de possibilitar a avaliação com alto grau de precisão e abrangência de determinada área do conhecimento, sem que cada estudante precise responder a longos testes; e de propiciar a comparação entre diferentes séries, por meio da construção de uma escala única de resultados, a escala de proficiência.

O PISA divulga as médias de proficiência em uma escala interpretativa, que indica as tarefas que os estudantes são capazes de realizar em relação às competências científicas. Essa escala foi construída por especialistas, com base nos dados obtidos em todas as edições do teste e apresenta seis níveis. O quadro 2, a

⁵ Explicações técnicas sobre a modelagem TRI fogem do escopo deste trabalho, mas podem ser encontradas em Andrade *et al* (2000). Interessante é notar que o SAEB e a maioria, se não todos, dos programas estaduais e municipais de avaliação do desempenho a utilizam.

seguir, representa essa escala de proficiência, de forma resumida, com a descrição das competências apresentadas pelos estudantes que se encontram em cada um dos seus níveis e o *score* que os classificou nesses níveis.

Quadro 2 - Escala de Proficiência de Ciências - PISA 2006

Nível	Score	Competências científicas apresentadas pelos estudantes
1	Até 409,5	Aplicam o conhecimento científico a umas poucas situações familiares, apresentam explicações científicas óbvias sobre uma evidência apresentada.
2	Até 484,1	Fornecem explicações referentes a situações familiares e tiram conclusões através de investigação simples. Interpretam resultados de forma simples ou interpretam de forma literal os resultados de pesquisas científicas.
3	Até 558,7	Identificam questões científicas em contextos diferentes. Selecionam conceitos científicos para explicar fenômenos e aplicam estratégias simples de pesquisa. Selecionam informações relevantes a partir de dados variados e avaliam situações.
4	Até 633,3	Inferem sobre o papel da ciência ou da tecnologia. Integram explicações e argumentos científicos, relacionando-os com aspectos cotidianos. Refletem sobre ações e tomam decisões a partir de evidências científicas.
5	Até 707,9	Identificam componentes científicos de situações complexas; comparam e avaliam evidências científicas para dar resposta a essas situações. Criticam situações reais, formulando explicações baseadas em evidência científica.
6	Acima de 707,9	Identificam, explicam e aplicam conhecimentos científicos em situações complexas. Desenvolvem argumentos para subsidiar decisões e recomendações de nível pessoal, social ou global.

Fonte: WAISELFISZ (2009, p. 26)

O Brasil, em seu primeiro ano de participação no PISA, 2000, apresentou média de proficiência em ciências de 375 pontos; em 2003, de 390; em 2006, ano em que o foco foi o letramento científico, a média permaneceu estagnada nos 390 pontos e em 2009, subiu para 405 pontos⁶. De acordo com esses dados, os estudantes brasileiros nunca conseguiram sair do nível 1 da escala, isto é, ainda encontram-se no nível mais baixo de letramento, sendo capazes de utilizar o conhecimento científico apenas para aplicá-lo em situações simples e óbvias. Nas próximas seções, veremos que, a exemplo do PISA, o SAEB, os programas estaduais de avaliação e o Avalia BH também divulgam uma escala de proficiência em ciências.

⁶ Disponível em: <<http://portal.inep.gov.br/internacional-novo-pisa-resultados>>. Acesso em: 2 nov. 2013.

A influência do PISA é constatada, ainda, sobre a construção do Índice de Desenvolvimento da Educação Básica – IDEB (SILVA, 2010). Esse indicador de qualidade cruza informações do desempenho estudantil no SAEB com os índices de aprovação (fluxo escolar) obtidos através do censo escolar. Por sua vez, o IDEB irá influenciar a criação de índices com o mesmo objetivo no sistema de avaliação do estado de São Paulo, o SARESP, e no do Amazonas, o SADEAM. Silva (2010) também afirma que o estabelecimento da meta pelo INEP de que o Brasil alcance a nota 6 no IDEB de 2022, corresponde aos resultados obtidos pelo PISA nos países membros da OECD, “para o alcance dessa meta, o Inep/MEC fez uma projeção para o Brasil e seus entes federados de 2007 a 2021, vislumbrando um aumento médio de 0,3 pontos a cada edição bianual do IDEB”.

1.4.2 O histórico da avaliação de competências científicas no SAEB

O SAEB surgiu de uma exigência do Banco Mundial por realizar uma avaliação do impacto do Projeto Nordeste – Segmento Educação, financiado por essa instituição, no âmbito do VI Acordo MEC/Banco Internacional de Reconstrução e Desenvolvimento – BIRD, alinhada ao interesse nacional em implementar um sistema amplo de avaliação da educação básica. Inicialmente denominado Sistema Nacional de Avaliação do Ensino Público de 1º Grau – SAEP, foi aplicado em 1988 apenas nos estados do Paraná e Rio Grande do Norte com o fim de testar a pertinência e adequação dos procedimentos e realizar ajustes para que pudessem ser expandidos (BONAMINO & FRANCO, 1999).

O primeiro ciclo de aplicação de provas para todo o Brasil ocorreu em 1990 e, até hoje, contabilizam-se doze ciclos. Ressalta-se que, desde a sua implementação, o SAEB tem sofrido diversas modificações de ordem metodológica e operacional, “a fim de aperfeiçoar normas e procedimentos específicos e assegurar cientificidade, confiabilidade e comparabilidade a seus resultados” (FERRÃO *et al*, 2001). A observação do histórico desse programa de avaliação educacional em larga escala, permite concluir que ocorreram alterações em relação aos seus objetivos principais, às séries avaliadas, aos instrumentos contextuais, aos modelos estatísticos para análise de resultados e às disciplinas cobradas.

Em 1990 e 1993⁷, foram avaliadas amostras de alunos das 1ª; 3ª, 5ª e 7ª séries da rede pública de ensino em relação às competências e habilidades desenvolvidas em Língua Portuguesa, Matemática e Ciências. Não dispondo ainda de uma Matriz de Referência⁸, os testes foram definidos com base nos currículos utilizados pelos sistemas estaduais de ensino (BRASÍLIA, 2013). Bonamino & Franco (1999) evidenciam que houve, nesse período, participação efetiva de professores, especialistas, técnicos das secretarias de educação e universitários tanto na determinação do conteúdo mínimo a ser cobrado como na validação dos instrumentos, na aplicação e correção dos testes.

Nesses primeiros anos do SAEB, não há, também, divulgação tampouco discussão dos dados produzidos. A razão disso, Bonamino & Franco (1999) imputam ao maior enfoque dado para os aspectos processuais do sistema de avaliação e os objetivos de

desenvolver e aprofundar a capacidade avaliativa das unidades gestoras do sistema educacional; regionalizar a operacionalização do processo avaliativo, criando nexos e estímulos para o desenvolvimento de infra-estrutura de pesquisa e avaliação educacional; propor uma estratégia de articulação dos resultados das pesquisas e avaliações já realizadas ou em vias de implementação. (p.111)

A partir de 1995, no entanto, a ênfase do SAEB passa a ser a produção de resultados que contribuam para o monitoramento da situação educacional brasileira e subsidie políticas públicas para a sua melhoria (BONAMINO & FRANCO, 1999). É nesse ciclo que se inicia a utilização da TRI para a análise dos dados, a construção de escalas de proficiência e a divulgação de relatórios⁹ contendo os resultados.

Além disso, nesse terceiro ciclo de aplicação, o SAEB inovou em aspectos como: inclusão do ensino médio e rede privada de ensino na amostra avaliada; incorporação de questionários para o levantamento de dados socioeconômicos, culturais e sobre os hábitos de estudo dos alunos; redefinição das séries avaliadas para o 4º e 8º anos do ensino fundamental (hoje, 5º e 9º anos) e 3º ano do ensino

⁷ Nota: o segundo ciclo de aplicação do SAEB deveria ter ocorrido em 1992, já que o programa é bianual. Porém, por problemas orçamentários, ele foi reprogramado para o ano seguinte (NETO, 2007).

⁸ Matriz de Referência: documento que explicita o objeto de uma avaliação. Constitui um conjunto de descritores que mostram as habilidades que são esperadas dos alunos em diferentes etapas de escolarização e passíveis de serem aferidas em testes padronizados de desempenho.

⁹ Relatórios sobre o SAEB 1995 são encontrados em:
<<http://www.publicacoes.inep.gov.br/resultados.asp?cat=6&subcat=4>>. Acesso em: 7 nov. 2013.

médio e construção de matrizes de referência para as provas de Leitura e Matemática (BRASÍLIA, 2013).

Ciências da Natureza não foram cobradas em 1995, voltando a figurar nos testes em 1997, onde foram acrescentadas as disciplinas de Física, Química e Biologia para o ensino médio e Ciências para o ensino fundamental (BONAMINO & FRANCO, 1999). A inclusão dessas disciplinas, de acordo com Lauglo (1997), deve-se às prescrições do Banco Mundial, que as considera especialmente importantes para o desenvolvimento econômico de um país.

Em 1997 urge a necessidade de desenvolver referências para uma avaliação, já que a metodologia de trabalho utilizada até então, preconizava a aferição das proficiências dos alunos por meio do seu desempenho real nos testes construídos sem parâmetros consensuais, advindos da consulta nacional sobre os conteúdos praticados e indicados pelas escolas brasileiras (BRASÍLIA, 1997). Assim, para orientar a produção de itens, foram estabelecidas matrizes de referência.

O processo de elaboração dessas matrizes envolveu a construção de uma Tabela de Convergência de Conteúdos que reunia conteúdos de propostas curriculares de estados e municípios brasileiros e a constituição de um grupo de trabalho formado por especialistas, professores da rede pública, entidades e técnicos do INEP para a

Análise de adequação e pertinência dos conteúdos validados constantes da Tabela de Convergência; Hierarquização destes conteúdos por ciclos, com terminalidade na 4ª e 8ª séries do ensino fundamental e na 3ª série do ensino médio; Associação dos conteúdos às competências cognitivas e habilidades instrumentais que lhes são próprias, expressas em forma de descritores do desempenho do aluno; Elaboração de indicadores concretos para a aferição de cada área em cada nível de complexidade cognitiva, sinalizados com a proporção de itens a serem demandados em cada uma delas; Classificação dos descritores, de acordo com o nível das operações mentais envolvidas; Distribuição dos descritores, de acordo com Temas, Tópicos e/ou Assuntos dos Conteúdos, observando os níveis das competências e habilidades envolvidas e a distribuição nos diferentes ciclos de avaliação, indicando a proporcionalidade de itens a serem construídos em cada bloco (BRASÍLIA, 1997, p.8).

O INEP (1997) ressalta a grande dificuldade no trabalho de elaboração das matrizes de referência de ciências da natureza devido ao seu “vasto e complexo campo de conteúdos”. Por esse motivo, explica o instituto, selecionaram conteúdos

considerados essenciais à formação básica do estudante e os agruparam nas categorias Tempo e Universo: Noções de Astronomia, Terra e Ambiente, Ser Humano: Desenvolvimento e Saúde. As competências, dentro desses domínios, foram subdivididas nas categorias: presentativa, onde se inserem as habilidades de identificar conceitos e nomear fenômenos do mundo natural ou transformado; operacional, que inclui a habilidade de relacionar esses conceitos e fenômenos; e a global, que reúne as habilidades de interpretar contextos científicos, resolver situações-problema e inferir sobre fenômenos (BRASIL, 1998).

Outro aspecto a destacar do SAEB 1997 é a forma de apresentação dos resultados que dá continuidade à perspectiva de interpretação de escalas de proficiência (BONAMINO & FRANCO, *et al*, 1999). Fontanive & Klein (2000) explicam, de forma mais detida, como se deu a construção dessas escalas:

As escalas de proficiência são obtidas a partir da calibração dos itens de teste através do modelo da Teoria de Resposta ao Item (TRI). Calibrar um item de teste significa identificar seus parâmetros, sendo uma das propriedades importantes da TRI o fato de que os parâmetros dos itens e as proficiências dos indivíduos são invariantes [...]. (p.416)

Os itens aplicados no teste foram analisados de acordo com os parâmetros de discriminação, dificuldade e acerto ao acaso e estimados quanto à proficiência exigida para que fossem gabaritados. Dessa forma, foi possível construir uma escala onde se alocam os itens e a proficiência dos estudantes em um *continuum* do nível mais alto para o mais baixo. Fontanive & Klein (2000) elucidam os procedimentos adotados para obter escalas comuns por disciplina abrangendo as três séries avaliadas:

- 1) aplicação de blocos de itens comuns entre séries [...]
- 2) transformação (equalização) das escalas de cada disciplina entre as séries para obtenção de uma escala comum a todas as séries, fixando-se uma média de 250 e um desvio padrão de 50 para a oitava série.

No caso das disciplinas Física, Química e Biologia, foi possível a construção de uma escala comum com Ciências, graças à confirmação da hipótese de unidimensionalidade (ou seja, a garantia de que cada teste abordou essencialmente uma única dimensão da área de conhecimento) e a presença de blocos de itens comuns entre Ciências de 8ª série e aquelas disciplinas. (p. 416)

O quadro 3, abaixo, mostra a escala de Ciências da Natureza construída a partir dos dados gerados em 1997. Para interpretar os desempenhos dos

avaliandos, foram fixados cinco pontos (níveis âncoras) nessa escala: 100, 175, 250, 325 e 400. A proficiência é explicada pela descrição das habilidades que eles demonstraram possuir quando situados em torno desses pontos.

Quadro 3 - Escala de Proficiência de Ciências da Natureza - SAEB 1997

100	Interpretam pequenos textos e observam figuras para identificar elementos do cotidiano relacionados ao ambiente ou à vida humana. Identificam, por exemplo, árvores entre outras representações de plantas. Reconhecem quais veículos utilizam derivados de petróleo como combustível.
175	Interpretam problemas elementares ligados ao meio ambiente e à saúde que aparecem em textos ou tabelas muito simples. Conhecem conceitos básicos das ciências naturais. Calculam velocidade média. Utilizam a linguagem do cotidiano e do senso comum, pois têm pouca familiaridade com a nomenclatura científica.
250	Identificam, nomeiam, classificam, comparam e sequenciam eventos a partir de textos ou representações figurativas. Reconhecem relações entre parte e todo e estabelecem relações temporais entre fenômenos. Demonstram conhecimentos elementares de astronomia, meio ambiente e saúde. Aplicam vocábulos científicos elementares cujo significado lhes é fornecido.
325	Compreendem a descrição de fenômenos e processos naturais e técnicos em linguagem cotidiana. Fazem uso rudimentar da notação científica e de símbolos específicos. Interpretam esquemas, diagramas, tabelas e gráficos com dados ambientais e de saúde. Estabelecem relações causais envolvendo fatores não diretamente observáveis. Dominam conceitos científicos de temas vivenciais como temperatura, poluição, cadeia alimentar.
400	Dominam conceitos científicos básicos e aplicam alguns deles. Utilizam a terminologia científica relacionada à estrutura, reprodução e hereditariedade. Explicam certos desequilíbrios ambientais e indicam fatores que contribuem para a estabilidade do meio ambiente e manutenção da biodiversidade. Identificam seres vivos a partir de suas características. Relacionam estruturas de seres vivos às suas funções. Reconhecem a seleção natural como fundamental para a evolução. Solucionam problemas relativos ao corpo humano, à evolução e à hereditariedade, em diferentes linguagens. Identificam fenômenos e processos físicos. Utilizam linguagens simbólicas, notações vetoriais, diagramas e gráficos relativos a processos físicos. Interpretam situações físicas relacionando causa e efeito. Calculam grandezas físicas. Convertem unidades. Utilizam princípios gerais da física em situações simples. Relacionam fenômenos, processos e modelos abstratos. Identificam alguns processos químicos em linguagem cotidiana e os transcrevem utilizando notação química. Interpretam gráficos de fenômenos, obtendo dados, efetuando cálculos simples e relacionando variáveis. Analisam uma reação química em solução aquosa e no estado gasoso, relacionando reagentes e produtos por meio de cálculos simples. Identificam a influência de fatores externos sobre as reações químicas. Utilizam modelos microscópios simples para explicar fenômenos químicos.

Fonte: FONTANIVE & KLEIN (2000). Adaptado.

É importante deixar claro que as interpretações dos níveis âncoras nessa escala não significaram expectativas do INEP/MEC quanto ao que os estudantes deveriam saber, mas o que eles de fato mostraram saber nos testes que fizeram em 1997 (FONTANIVE & KLEIN, 2000). A média de proficiência em ciências apresentada pelos estudantes do 4º ano do ensino fundamental foi de 181 pontos; dos estudantes do 8º ano, 250; no ensino médio, as proficiências médias em física, química e biologia foram, respectivamente, de 290, 297 e 298. Como se vê, o SAEB 1997 atesta a fragilidade do ensino dessas disciplinas.

O quinto ciclo do SAEB, em 1999, traz de inédito a inclusão das disciplinas Geografia e História nos exames. Em relação às Ciências da Natureza, ocorre uma revisão das matrizes de referência de 1997 quanto a sua adequação e pertinência (Brasília, 2013), porém os resultados dos testes não foram praticamente divulgados nem trabalhados (WASELFIZ, 2009). Esse ciclo também encerra a aferição de competências científicas e, como já adiantado neste trabalho, o INEP não explica, oficialmente, o que motivou a exclusão. Em 2013, porém, o Diário Oficial da União, nº 119, em 24 de junho, anuncia o retorno da avaliação de ciências para o 9º ano do ensino fundamental, com a proposta de

[...] aplicar, em caráter experimental para validação das matrizes e escalas, os testes de Ciências definidos nas Matrizes de Referência do Sistema de Avaliação da Educação Básica - SAEB para 9º ano do ensino fundamental e para o 3º ano do ensino médio [...]. (p.34)

O INEP aponta na legislação educacional e nas diretrizes curriculares nacionais a justificativa para tal inclusão, mostrando que a Constituição Federal de 1988, a Lei de Diretrizes e Bases da Educação Nacional (1996), as Diretrizes Curriculares Nacionais para a Educação Básica e os Parâmetros Curriculares Nacionais confluem para a determinação de conteúdos, no ensino fundamental, que permitam uma abordagem de temas abrangentes e contemporâneos que afetam a vida humana em escala global, regional e individual, enquanto, para o ensino médio apregoam a importância do “domínio dos princípios científicos e tecnológicos que presidem a produção moderna” (BRASÍLIA, 2013).

1.4.3 A inclusão de ciências nos programas estaduais de avaliação da educação

A tabela 1, a seguir, apresenta uma relação dos estados brasileiros que criaram sistemas próprios de avaliação em larga escala do desempenho estudantil, as respectivas datas de criação, as disciplinas que esses programas avaliam atualmente e o momento em que incluíram testes de Ciências da Natureza.

Tabela 1 - Sistemas de Avaliação Estaduais da Educação¹⁰

Estado	Programa	Início	Disciplinas			
			LP	MAT	CN	CH
AC	SEAPE	2009	X	X		
AL	SAVEAL	2001	X	X		
AM	SADEAM	2008	X	X	X	X
BA	SABE	2008	X	X	X	X
CE	SPAECE	1992	X	X		
		2012	X	X	X	X
ES	PAEBES	2004	X	X		
		2011	X	X	X	X
GO	SAEGO	2001	X	X		
MA	SIMADE	2010	X	X		
MS	SAEMS	2003	X	X		
MG	SIMAVE	2000	X	X		
PB	Avaliando IDEPB	2012	X	X		
PR	SAEP	2012	X	X		
PE	SAEPE	2008	X	X		
PI	SAEPI	2011	X	X		
RJ	SAERJ	2008	X	X		
		2011	X	X	X	X
RS	SAERS	2007	X	X		
RO	SAERO	2012	X	X		
SP	SARESP	1995	X	X		
		2008	X	X	X	X
TO	SALTO	2011	X	X		
		2013	X	X	X	

Como se observa, o número de estados que aferem a aprendizagem dos seus alunos a nível sistêmico é bastante expressivo, revelando a cultura de

¹⁰ Informações obtidas através do contato com a coordenação do setor de contratos do Centro de Políticas Públicas e Avaliação da Educação – CAEd/UFJF e de documento publicado pelo Grupo de Avaliação e Medidas Educacionais – GAME/UFMG, em 2011, disponível em: <<http://www.fvc.org.br/pdf/relatorio-avaliacoes-externas.pdf>>. Acesso em: 27 out. 2013.

avaliação que se instaurou no Brasil a partir de 1990 e ganhou ênfase no final da década de 2000 e nos últimos anos. Observa-se, também, a importância dada aos conhecimentos nas áreas de Linguagem e Matemática, competências examinadas em todos os programas desde a sua implementação, bem como é possível denotar a nova tendência de incluir exames de Ciências da Natureza e Ciências Humanas.

O Ceará foi o primeiro estado a criar um sistema próprio de avaliação educacional em larga escala, o Sistema Permanente de Avaliação da Educação Básica do Ceará – SPAECE, em 1992. Desde essa data, avalia anualmente e censitariamente as competências e habilidades em Língua Portuguesa e Matemática. Em 2012, porém, os estudantes que finalizaram o ensino médio fizeram, também, provas de Ciências da Natureza e Ciências Humanas. Essa inclusão pode estar relacionada a um movimento surgido no estado com o objetivo de motivar e preparar os estudantes da rede pública de ensino para o ENEM que inclui essas disciplinas em sua avaliação.

As provas de Biologia, Física e Química do SPAECE se balizaram nas matrizes de referência do ENEM. Os testes foram montados com itens aplicados em anos anteriores cedidos pelo INEP e itens inéditos elaborados por especialistas do CAEd, órgão responsável pelo desenvolvimento do programa. Em caráter experimental, essas provas funcionaram mais como um simulado do ENEM, do que como uma avaliação em larga escala e seus resultados não receberam a atenção, ainda, que recebem os de proficiência em Linguagem e Matemática, que são publicados em boletins, amplamente divulgados e discutidos em oficinas e reuniões com gestores, docentes e demais envolvidos com a escola.

Minas Gerais, com o Sistema Mineiro de Avaliação da Educação – SIMAVE, chegou a aferir competências científicas dos seus estudantes no seu segundo ciclo de aplicação, em 2001. Porém, em 2004, quando essa avaliação deveria ser repetida, a Secretaria de Estado da Educação resolveu por concentrar a avaliação nas áreas de Língua Portuguesa e Matemática, apenas (SILVA, 2007).

Os estados do Amazonas, Bahia e São Paulo atentaram-se para a importância de se avaliar o desempenho em Ciências da Natureza em 2008. Os dois primeiros realizaram testes nesse campo do conhecimento logo no primeiro ano de aplicação, já São Paulo, com programa de avaliação consolidado desde 1996,

voltou-se para outras competências que não somente Linguagens e Matemática um pouco mais tardiamente.

Sendo conduzidos pela mesma instituição, o Centro de Políticas Públicas e Avaliação da Educação - CAEd, o Sistema Estadual de Avaliação do Amazonas (SADEAM) e o Sistema de Avaliação Baiano de Educação (SABE) enquadram-se em moldes bastante semelhantes: são anuais; censitários; adotam a TRI para a construção dos instrumentos de avaliação e análises dos resultados; aplicam, também, questionários contextuais sobre a situação socioeconômica dos estudantes e a infraestrutura escolar; divulgam e analisam os dados produzidos em oficinas de apropriação com gestores educacionais e docentes e por meio da internet.

Atualmente, o SADEAM avalia competências desenvolvidas por seus alunos da rede estadual matriculados nos 3º, 5º, 7º e 9º anos do Ensino Fundamental, anos iniciais e finais da Educação de Jovens e Adultos (EJA), 1º e 3º anos do Ensino Médio e o 3º da EJA. Fazem os exames de Ciências da Natureza esses dois últimos grupos. Documento que apresenta o programa de avaliação amazonense às escolas, intitulado Revista do Sistema, evidencia a importância da avaliação externa para a discussão da qualidade do ensino das Ciências da Natureza e no estabelecimento de um norte na redefinição curricular (AMAZONAS, 2011).

A Matriz de Referência para as avaliações do SADEAM foi construída com base nas Orientações Curriculares para o Ensino Fundamental e Médio do estado, nos Parâmetros Curriculares Nacionais – PCN, nas Orientações Educacionais Complementares aos PCN – PCN+, na consulta aos professores regentes da rede e na matriz de referência para o ENEM 2011. As habilidades descritas nessa matriz são selecionadas de acordo com os critérios de relevância para a etapa de escolarização avaliada e possibilidade de serem aferidas em testes de múltipla escolha. Elas são distribuídas dentro de cinco competências consideradas essenciais: Matéria e Energia, Terra e Universo, Ser Humano e Saúde, Vida e Ambiente e Tecnologia e Sociedade, organização que será encontrada em todos os programas de avaliação coordenados pelo CAEd.

O SADEAM classifica os estudantes de acordo com sua média de proficiência em padrões de desempenho:

os padrões de desempenho são categorias definidas a partir de cortes numéricos que agrupam os níveis de proficiência, com base nas metas educacionais estabelecidas pelo Sadeam. Esses cortes dão origem a quatro Padrões de desempenho – Abaixo do básico,

Básico, Proficiente e Avançado – os quais apresentam o perfil de desempenho dos alunos (AMAZONAS, 2012, p.36).

O quadro 4, a seguir, apresenta a classificação dos estudantes amazonenses em padrões de desempenho, de acordo com as habilidades científicas que demonstraram ter desenvolvido nos testes de 2012.

Quadro 4 - Padrão de Desempenho em Biologia no SADEAM 2012

Padrão	Proficiência	Habilidades apresentadas pelos avaliandos
Abaixo do Básico	Até 550	Desenvolvimento de habilidades consideradas elementares, como identificar ações antrópicas e seus impactos no ambiente e reconhecer hábitos alimentares saudáveis.
Básico	De 550 a 650	Desenvolvimento de habilidades com pouco nível de complexidade, como identificar níveis tróficos em cadeias alimentares; classificar vertebrados de acordo com sua morfofisiologia; identificar mecanismos de transmissão de algumas doenças que afetam a espécie humana e reconhecer teorias evolucionistas.
Proficiente	De 650 a 750	Desenvolvimento de maior número de habilidades consideradas essenciais, como associar órgãos e funções, compreender a importância do ciclo do nitrogênio, reconhecer a aplicabilidade dos testes de DNA e resolver problemas envolvendo a Primeira Lei de Mendel.
Avançado	Acima de 750	Desenvolvimento de habilidades necessárias à realização de tarefas que exigem maior nível cognitivo, como compreender relações ecológicas, identificar etapas do desenvolvimento embrionário, compreender conceitos básicos de genética, associar organoides citoplasmáticos às suas funções, relacionar a reprodução com a proliferação dos seres vivos e a variabilidade genética e comparar o desenvolvimento embrionário de diferentes grupos de vertebrados.

Fonte: AMAZONAS, 2012. Adaptado

Em 2012, a média de proficiência em Biologia para o 3º ano do Ensino Médio regular foi de 502,3, indicando que os estudantes amazonenses encontram-se bem aquém do esperado, abaixo do básico, em relação ao desenvolvimento de competências científicas nessa etapa de conclusão da jornada escolar.

Um aspecto importante a se mencionar sobre o SADEAM é o fato de as proficiências apresentadas pelos estudantes nas provas do SADEAM comporem, juntamente com a taxa de aprovação, o Índice de Educação do Amazonas – IDEAM. As escolas que obtêm os melhores desempenhos no IDEAM são premiadas, bem

como seus funcionários (AMAZONAS, 2012), sendo uma política de incentivo ao alcance de metas e melhoria da qualidade do ensino aprendizagem a cada ano de avaliação.

O SABE se destaca por ser o primeiro programa estadual de aferição de desempenho de caráter longitudinal, isto é, que acompanha o desenvolvimento de um mesmo grupo de estudantes durante um determinado período (BAHIA, 2011). Em 2008, avaliou o 1º ano do ensino médio em relação à aquisição de habilidades interdisciplinares, utilizando a matriz do ENEM. Em 2009, avaliou o 2º ano do ensino médio, cobrando habilidades dos campos da Linguagem, Ciências Humanas, Ciências da Natureza e Matemática ainda com base na Matriz do Enem. Em 2010, o 3º ano do ensino médio é avaliado e a matriz de referência para os testes passa a ser do próprio sistema. Em 2011, inicia-se outro triênio. É importante salientar que o SABE avalia não só o ensino regular, mas a Educação Profissional Integrada ao Ensino Médio.

As Matrizes de Referência do SABE foram construídas levando em conta a matriz do ENEM, a do SAEB, os Parâmetros Curriculares Nacionais e contribuições de professores regentes da rede. Na Matriz de Referência para avaliação em Ciências da Natureza e suas tecnologias, os domínios representam os eixos temáticos das áreas de conhecimento. Os domínios selecionados, de acordo com a Revista da Avaliação, publicada em 2011 pelo CAEd, permitem a interação entre os componentes curriculares de Biologia, Física e Química.

Assim como o SADEAM, o programa de avaliação da Bahia subdivide seus avaliandos em categorias que explicitam a situação em que se encontram em relação ao seu desempenho, são eles: Muito Crítico, Crítico, Básico e Avançado. Segundo documento intitulado Revista do Sistema Avalie Ensino Médio, publicada em 2012,

esses padrões proporcionam uma interpretação pedagógica das habilidades desenvolvidas pelos estudantes e oferecem o entendimento a respeito do nível em que eles se encontram. Por meio dos padrões, é possível analisar os aspectos cognitivos que demarcam os percentuais de estudantes situados nos diferentes níveis de desempenho, bem como a diferença de aprendizagem entre eles, refletindo a distância existente entre aqueles que têm grandes chances de atingir o sucesso escolar; e aqueles que estão suscetíveis ao abandono escolar caso não sejam implementadas ações e políticas educacionais com vistas à promoção da equidade. (p.17)

O quadro 5, a seguir, apresenta o padrão de desempenho dos estudantes avaliados pelo SABE em Biologia em 2011.

Quadro 5 - Padrão de Desempenho em biologia dos estudantes baianos avaliados pelo SABE em 2011

Padrão/Proficiência	Habilidades apresentadas pelos avaliandos do 1ºEM
Muito Crítico Até 500	Não desenvolveram habilidades consideradas mínimas para a etapa de escolarização avaliada.
Crítico De 500 a 600	Iniciaram o desenvolvimento de habilidades consideradas essenciais, como identificar as consequências do aumento do número de veículos movidos por combustíveis fósseis, identificar os fatores limitantes da distribuição de água no planeta Terra, interpretar uma pirâmide nutricional.
Básico De 600 a 700	Desenvolveram habilidades que exigem processos cognitivos mais elaborados, como identificar reagentes e produtos da fotossíntese e respiração celular, reconhecer os papéis biológicos da água, reconhecer aplicabilidades da Engenharia Genética.
Avançado Acima de 700	Desenvolveram habilidades que permitem relacionar conhecimentos de Biologia às situações do cotidiano, como associar mitoses descontroladas ao surgimento do câncer, reconhecer princípios do teste de DNA e associar a prática de atividade esportiva ao metabolismo energético do organismo.

Fonte: BAHIA, 2011. Adaptado.

Nesse ciclo de avaliação, os avaliandos apresentaram proficiência média de 499,4 em Biologia, o que os enquadra no padrão Muito Crítico e reflete a precariedade do ensino dessa disciplina integrante das Ciências Naturais no estado.

O Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo – SARESP – foi criado em 1996 e ocorre, anualmente, a partir dessa data. É aplicado, obrigatoriamente, na rede estadual e, por meio de adesão, nas redes municipais e escolas privadas. Avalia, desde a sua fundação, os alunos do 3º, 5º, 7º e 9º anos do Ensino Fundamental e da 3ª série do Ensino Médio, por meio de provas cognitivas nas áreas de Língua Portuguesa com Redação e Matemática. Em 2007, o sistema sofreu diversas mudanças, dentre as quais se destaca a inclusão da avaliação de ciências da natureza e ciências humanas para os estudantes do Ensino Médio (SÃO PAULO, 2008).

Os testes de Ciências da Natureza são aplicados bianualmente e são, como os de outras disciplinas, construídos e analisados por meio da TRI. Os resultados gerados são apresentados por relatórios, que são amplamente divulgados por meio de uma escala de proficiência comum com a do SAEB para Língua Portuguesa e

Matemática e comum entre séries para Ciências da Natureza e de padrões de desempenho, os quais classificam os alunos em Abaixo do Básico, Básico, Adequado e Avançado. As Matrizes de Referência foram elaboradas a partir do Novo Currículo do Estado de São Paulo que é a base comum às escolas estaduais e dos documentos que balizam as avaliações nacionais e internacionais (SÃO PAULO, 2008).

O quadro 6, abaixo, mostra os níveis de proficiência de Ciências e Ciências da Natureza estabelecidos pelo SARESP em 2008, a partir das expectativas de aprendizagem (conteúdos, competências e habilidades) para cada série.

Quadro 6 - Níveis de Proficiência em Ciências da Natureza - SARESP 2008

Níveis	7ºEF	9ºEF	3ºEM
Abaixo do Básico	Menor que 200	Menor que 225	Menor que 275
Básico	Entre 200 e 250	Entre 225 e 300	Entre 275 e 350
Adequado	Entre 250 e 325	Entre 300 e 350	Entre 350 e 400
Avançado	Acima de 325	Acima de 350	Acima de 400

Fonte: SÃO PAULO (2008, p.22)

Os estudantes que se encontram no nível Abaixo do Básico demonstram insuficiência na consolidação de competências e habilidades esperadas para a série que se encontram; os que se encontram no Básico, mostram desenvolvimento parcial dos conteúdos; os classificados no nível Adequado demonstram domínio das habilidades desejáveis para a etapa de escolarização e os alunos no nível Avançado apresentam competências além das requeridas (SÃO PAULO, 2008).

A tabela 2, abaixo, mostra as médias de proficiência de Ciências e Ciências da Natureza alcançadas pelos estudantes paulistas do primeiro ao último ciclo de aplicação.

Tabela 2 - Médias de Proficiência dos estudantes da rede pública estadual de São Paulo em Ciências da Natureza no SARESP de 2008, 2010 e 2012

Série	2008	2010	2012
7º ano EF	226,9	222,5	220,9
9º ano EF	250	247,9	248,8
3º ano EM	274	274,4	272,3

Essa tabela mostra que os estudantes paulistas do Ensino Médio estão em situação ainda mais alarmante do que os do ensino fundamental no que diz respeito aos conhecimentos apresentados em ciências, situando-se no Abaixo do básico, enquanto aqueles se encontram no Básico. Nota-se, também, que não houve avanço nas médias de proficiência de um ciclo para o outro.

Seguindo a tendência nacional, em 2011, o Programa de Avaliação da Educação Básica do Espírito Santo – PAEBES, existente desde 2004, passa a avaliar ciências naturais no terceiro ano do ensino médio bianualmente, alternando essa avaliação com a de ciências humanas. Sua matriz de referência para os testes difere do padrão seguido pelos outros sistemas estaduais de avaliação da aprendizagem. A matriz de Biologia estrutura os descritores dentro de seis grandes temas, que correspondem aos eixos apresentados no currículo básico do Espírito Santo, a matriz de Física segue os programas das disciplinas mais frequentes nos livros didáticos adotados pela Secretaria de Educação, utilizando a designação convencional das áreas: Mecânica, Termodinâmica, Ótica, Eletromagnetismo e Física Moderna e a de Química é organizada em treze grandes temas, que constam do programa curricular da disciplina.

Coordenado pelo CAEd, o programa capixaba apresenta os mesmos eixos que os outros: TRI, escalas de proficiência, padrões de desempenho, questionários contextuais, divulgação e discussão de resultados em oficinas. O PAEBES classifica seus estudantes nos padrões de proficiência Abaixo do Básico, Básico, Proficiente e Avançado. O quadro 7, seguinte, apresenta as competências científicas na área de biologia desenvolvidas pelos estudantes do Espírito Santo no primeiro ano de aplicação dos testes dessa disciplina.

Quadro 7 - Padrão de Desempenho em Biologia dos estudantes capixabas - PAEBES 2011

Nível/Proficiência	Habilidades
Abaixo do Básico Até 250	Apresentam habilidades de baixa complexidade, como identificar ações antrópicas que causam impactos negativos ao ambiente e suas consequências; reconhecer formas de prevenção de algumas viroses; identificar mecanismos de transmissão de algumas doenças; diferenciar níveis tróficos de uma cadeia alimentar.
Básico De 250 até 325	Apresentam habilidades que requerem maior poder de abstração em situações que possuem maior complexidade, como identificar reagentes, produtos e etapas da fotossíntese; identificar a presença da flor como vantagem adaptativa; identificar características das teorias evolucionistas e o papel da seleção natural no sucesso adaptativo das espécies; relacionar os órgãos do sistema genital humano às suas funções; reconhecer a decomposição como processo responsável pela ciclagem da matéria orgânica; reconhecer impactos positivos da utilização de transgênicos; relacionar vertebrados às suas classes; compreender a interferência de fatores ambientais na fisiologia dos peixes.
Proficiente De 325 até 375	Desenvolveram habilidades mais complexas, que exigem processos cognitivos mais elaborados, como reconhecer a ancestralidade de espécies do gênero Homo em árvore filogenética; relacionar funções de órgãos e sistemas envolvidos na distribuição de energia para as células; associar características adaptativas das aves à capacidade de voo; diferenciar hipóteses sobre o surgimento da vida; diferenciar teorias evolucionistas; diferenciar os tipos de reprodução; reconhecer representantes dos diferentes reinos.
Avançado Acima de 375	Realizaram tarefas com maior exigência cognitiva, como identificar etapas do desenvolvimento embrionário; reconhecer fisiologia das plantas com sementes; reconhecer as fases da meiose; identificar aquisições evolutivas de um grupo em relação a outro; associar a estrutura bioquímica da membrana plasmática à permeabilidade seletiva; relacionar a ocorrência da eritroblastose fetal ao genótipo dos pais, resolver problemas relacionados à herança ligada ao sexo, analisar gráficos referentes ao ponto de compensação fótico; compreender a técnica do DNA recombinante, assim como as etapas e finalidades do processo de clonagem.

Fonte: ESPÍRITO SANTO (2011)

A proficiência média dos estudantes avaliados pelo PAEBES em 2011 nas disciplinas de biologia, física e química foram, respectivamente, de 249,1; 248,9 e 248,7, o que os situa no padrão Abaixo do Básico em Ciências da Natureza, estando muito aquém do esperado para concluintes do ensino médio.

Também em 2011, o estado do Rio de Janeiro começou a avaliar as competências científicas de seus estudantes por meio do programa SAERJINHO, criado no âmbito do Sistema de Avaliação da Educação Básica do Rio de Janeiro – SAERJ, existente desde 2008. O SAERJINHO é um programa bimestral de avaliação diagnóstica do processo de ensino e aprendizagem que surgiu com o objetivo de acompanhar mais de perto e de forma mais rápida o desempenho

estudantil, para que medidas de intervenções, como o reforço na aprendizagem e a capacitação docente, possam ser tomadas, visando melhorar os índices educacionais produzidos pelo SAERJ e IDEB¹¹. A Matriz de Referência para os testes de Ciências foram elaboradas a partir do Currículo Mínimo do estado, documento que serve como referência para as escolas, apresentando os conteúdos básicos que devem estar contidos nos planos de aula dos docentes.

As escolas que apresentam melhores resultados no SAERJINHO e conseguiram atingir a meta estipulada nas médias da avaliação recebem recursos extras e seus professores e funcionários logram uma bonificação, que é um complemento salarial. Os estudantes que apresentaram melhoras no desempenho também são premiados com *notebooks*.

1.5 O AVALIA BH e o processo de inclusão dos testes de Ciências

Descrever os sistemas municipais de avaliação da educação, a exemplo do realizado com os estados neste trabalho, torna-se inviável: uma breve busca na internet e a revisão de literatura revelam que embora muitos municípios brasileiros tenham imergido nas avaliações educacionais em larga escala, criando sistemas próprios de aferição da aprendizagem, poucas são as publicações a respeito do desenho e resultados desses programas.

É em uma publicação anual do Inep intitulada Prêmio Inovação em Gestão Educacional, que se pode conhecer, mesmo que de forma rasa, alguns programas municipais de avaliação da educação que se destacaram no Brasil. Esse documento relata experiências exitosas de municípios em todo o território brasileiro quanto à gestão da educação e, em grande parte, referem-se à implementação de avaliação sistêmica (BRASÍLIA, 2007, 2012). Porém, a maioria dessas iniciativas, se não todas, ainda se concentram na avaliação de Língua Portuguesa e Matemática, o que não é relevante nesse estudo que enfoca a avaliação de habilidades científicas.

Nesse sentido, voltemos, agora, para o objeto deste estudo, o programa de avaliação da educação da prefeitura de Belo Horizonte-MG. Esta seção procura descrever, de forma mais detida, o desenho do Avalia BH, para, em seguida, versar sobre seus efeitos.

¹¹ Disponível em: <<http://www.rj.gov.br/web/seeduc/exibeconteudo?article-id=616581>>. Acesso em: 12 dez. 2013.

1.5.1 Antecedentes

A universalização do acesso ao Ensino Fundamental no Brasil, ocorrida na década de 1990 em cumprimento às exigências da nova Constituição Federal, trouxe grandes desafios para os sistemas de educação no país, sobretudo os municipais, cuja ação deve ser prioritária sobre essa etapa de escolarização. É recorrente, na literatura científica, o discurso de que a ampliação das matrículas foi acompanhada do declínio da qualidade do ensino nas escolas públicas, que não se encontravam preparadas para receber alunos provenientes de classes sociais menos favorecidas, com vivências e culturas bem diferentes das pertencentes aos alunos de classe média, para os quais estiveram tradicionalmente voltadas.

Nesse contexto situava-se Belo Horizonte. Miranda (2007) relata os graves problemas detectados na estrutura educacional do município diante da universalização:

- as propostas pedagógicas vigentes não facilitavam a progressão das crianças das classes populares: continuavam elitistas e antipopulares;
- as crianças das classes sociais mais baixas que estavam ingressando na rede não eram capazes de superar as diferenças culturais – acabavam por ser expulsas das escolas [...];
- as práticas pedagógicas escolares reforçavam a repetência [...];
- a evasão aumentava à medida que os alunos se aproximavam da idade de trabalhar;
- os repetentes perdiam a autoestima por terem de conviver com colegas muito mais jovens;
- [...] as escolas não respeitavam o ritmo dos alunos, já que todos tinham de aprender os mesmos conteúdos, simultaneamente e na mesma sequência, em tempos cada vez mais curtos;
- o trabalho dos professores era muito isolado, cada um cuidava de sua disciplina e não se preocupava com a interdisciplinaridade, ou com o trabalho de seu colega. (p. 62-63)

Ainda de acordo com Miranda (2007), esse diagnóstico ocasionou uma inquietação em um grupo de docentes da rede, o qual já participava de movimentos de renovação pedagógica e debates a respeito da nova Lei de Diretrizes e Bases da Educação - LDB, em voga naquele período. Esses professores começaram a exigir da gestão municipal, mudanças significativas na qualidade de ensino, aspirando por uma proposta que rompesse com aquela lógica de organização escolar elitista. Dessa forma, foram realizadas diversas discussões, que incluíam o prefeito e seus secretários, para se chegar a um consenso sobre o tipo de projeto a ser adotado que fosse capaz de promover a melhoria da aprendizagem, reduzir a reprovação e a

evasão escolar, tornar a escola um centro cultural e atrativo para os estudantes e comunidade (MIRANDA, 2007). Daí, nasce o projeto Escola Plural.

A Escola Plural foi implementada em 1994, apresentando quatro grandes núcleos de ação: 1) os eixos norteadores, que serviam de base para as práticas escolares,

- uma intervenção coletiva mais radical;
- a sensibilidade em relação à totalidade da formação humana;
- a escola como tempo de vivência cultural;
- a escola como espaço de produção cultural;
- as virtualidades educativas da materialidade da escola;
- a vivência de cada idade de formação sem interrupção;
- a socialização adequada a cada idade – ciclo de formação;
- uma nova identidade da escola, uma nova identidade do seu profissional (BELO HORIZONTE, 2002, p. 9-22);

2) a reorganização dos tempos escolares, que determinou a ampliação do tempo escolar do ensino fundamental de oito para nove anos e substituiu a seriação pela organização em ciclos; 3) os processos de formação plural, que propôs a ressignificação do processo de ensino e aprendizagem, trabalhando a interdisciplinaridade e questões de interesse social; e 4) a avaliação, que sobrepujou a costumeira forma de avaliar classificatória, excludente e centrada no aluno, adotando uma concepção de avaliação como processo formativo, contínuo e focado no sistema educacional (MIRANDA, 2007).

Essas concepções, especialmente as referentes à avaliação, influenciarão sobremaneira a Gerência de Avaliação de Políticas Educacionais – GAPED, órgão pertencente à Secretaria Municipal de Educação de Belo Horizonte - SMED, na criação do programa Avaliação do Conhecimento Apreendido ou, como é mais conhecido, o Avalia BH, quatorze anos depois.

O Avalia BH surge, de acordo com entrevista cedida por membros da GAPED, da necessidade de um olhar próprio sobre o sistema de ensino municipal. Belo Horizonte participava, desde 2003, da avaliação estadual externa SIMAVE/PROEB, que já expunha o desempenho precário dos estudantes dos 5º e 9º anos do ensino fundamental em Língua Portuguesa e Matemática, mas não apontava uma forma de intervenção rápida e precisa. Nesse contexto, a SMED solicita à GAPED a arquitetura de um programa de avaliação em larga escala que possibilitasse o intermédio pontual e imediato sobre grupos de alunos com maiores déficits de aprendizagem, para que conseguissem avançar em seus estudos, de forma equalitária, de acordo com as prerrogativas da Escola Plural.

Assim, em 2008 é inaugurado o Avalia BH, que, conforme os documentos oficiais propalam, foi desenvolvido com o objetivo de garantir aos gestores municipais, a partir da análise dos resultados dos testes, informações fidedignas sobre a realidade educacional e, assim, obter possibilidades reais para a formulação e definição de políticas públicas capazes de criar um sistema mais justo e inclusivo, voltado para os desafios do novo século (BELO HORIZONTE, 2012).

1.5.2 O Desenho do Sistema

O Avalia BH é o sistema de avaliação educacional em larga escala da Prefeitura de Belo Horizonte. É anual e censitário, consistindo de testes de desempenho para os alunos do 3º ao 9º ano do ensino fundamental diurno e da 2ª à 8ª séries do ensino fundamental noturno, além de questionários contextuais que permitem identificar fatores que interferem na aprendizagem e que são respondidos, também, pelos professores, coordenadores e gestores escolares. Em 2008 e 2009, avaliou conhecimentos apreendidos em Língua Portuguesa e Matemática. A partir de 2010, começa a avaliar, também, as competências desenvolvidas em Ciências da Natureza, o que constituiu uma de suas inovações em relação aos demais sistemas avaliativos (BELO HORIZONTE, 2008).

Outras inovações do Avalia BH são a identificação nominal dos estudantes com baixo desempenho, o que possibilita o seu acompanhamento durante todo o ensino fundamental e a implementação de uma avaliação diagnóstica semestral, que complementa a avaliação externa, oportunizando aos docentes, de forma instantânea, o conhecimento do patamar em que se encontram seus alunos, a fim de que tenham um suporte na escolha da didática melhor apropriada para o trabalho em sala de aula (BELO HORIZONTE, 2008).

Dessa forma, o Avalia BH é subdividido em dois programas: o Programa de Avaliação Externa do Desempenho Escolar e o Programa de Avaliação Diagnóstica do Desempenho Escolar. O primeiro compreende os testes aplicados ao final de cada ano letivo, os quais são construídos e analisados sob a luz da Teoria de Resposta ao Item. Os itens que os compõem são pré-testados, isto é, aplicados a uma amostra de alunos de outro município para que os que apresentem melhores resultados estatísticos, sejam selecionados para o teste final a ser aplicado no município de Belo Horizonte, enquanto os que apresentam mau comportamento no

pré-teste, como a incapacidade de distinguir os estudantes que desenvolveram determinada habilidade dos que não desenvolveram, são eliminados.

A avaliação externa inclui o questionário contextual que coleta informações socioeconômicas e a opinião dos estudantes, professores e gestores sobre aspectos do processo de ensino-aprendizagem. Os dados produzidos por esses questionários são cruzados com os de desempenho, de modo a promover uma ampliação do conhecimento sobre a relação entre o ensino e as condições em que ocorre a aprendizagem. Os cadernos de provas do Avalia BH apresentam três blocos, contendo 9 itens, cada, os quais são combinados de forma a que um grupo de alunos responda itens correspondentes a todos os descritores da matriz de referência, segundo a metodologia dos Blocos Incompletos Balanceados – BIB¹².

As matrizes de referência do Avalia BH, elaboradas em conjunto pelos gestores e professores da rede municipal de educação, consistem em um recorte do currículo trabalhado na sala de aula, o qual é fundamentado nos Parâmetros Curriculares Nacionais – PCN e nas Proposições Curriculares para o Ensino Fundamental da Rede Municipal de Belo Horizonte, sendo que as de Língua Portuguesa e Matemática contemplam, ainda, alguns descritores do SAEB, com o objetivo de se manter a comparabilidade entre resultados municipais e os nacionais.

É importante frisar que a avaliação externa tem seu enfoque no desempenho escolar e o seu resultado é uma medida de proficiência que fornece o aporte para os gestores na implementação de políticas públicas (BELO HORIZONTE, 2009). A proficiência é apresentada em uma escala construída pelo SAEB para Língua Portuguesa e Matemática, a qual ordena, em um *continuum*, o desempenho dos estudantes, do nível mais baixo ao mais alto, descrevendo o número e a complexidade das habilidades desenvolvidas dependendo do intervalo onde se encontram nessa escala. Os resultados do Avalia BH também são apresentados por meio de quadros padrões de desempenho, definidos pela secretaria municipal de educação, que classificam os alunos em Abaixo do Básico, Básico, Satisfatório e Avançado, de acordo com a proficiência. Segundo a SME,

esses padrões proporcionam uma interpretação pedagógica das habilidades desenvolvidas pelos alunos e oferecem à escola o

¹²Blocos Incompletos Balanceados: delineamento em que cada um dos t tratamentos é repetido r vezes e ocorre no máximo uma vez em cada um dos b blocos.

entendimento a respeito do nível em que eles se encontram. Por meio deles, é possível analisar a distância de aprendizagem entre os alunos que se encontram em diferentes níveis de desempenho, do mais baixo ao mais elevado. É importante atentar-se para os alunos que estão nos padrões mais baixos, pois são eles os mais vulneráveis à evasão e ao insucesso escolar. (BELO HORIZONTE, 2012, p.15)

Já as avaliações diagnósticas têm como foco o aluno e objetivam prevenir a detecção tardia de dificuldades de aprendizagem, além de buscar conhecer interesses e competências enquanto pré-requisitos para futuras ações pedagógicas (BELO HORIZONTE, 2009). Essas avaliações ocorrem no início de cada semestre letivo, sendo construídas e corrigidas pelo próprio professor de cada turma. Apresentam 24 itens de cada área do conhecimento que são selecionados, de acordo com a habilidade que afere, em um banco de itens disponibilizado pelo CAEd, órgão responsável pelo desenvolvimento da metodologia do Avalia BH, desde 2009.

Os resultados das avaliações diagnósticas são aferidos por meio de uma planilha de Excel criada e cedida aos professores da rede municipal pela GAPED. Nessa planilha de correção, os docentes lançam as respostas a cada item e obtém, com o somatório dos acertos, a classificação de seus alunos dentro dos padrões de desempenho que se encontram naquele momento, além do percentual de acerto por descritor da turma como um todo. Dessa forma, o professor consegue apontar, de forma imediata, os alunos com maiores problemas de defasagem no aprendizado, podendo, desde já, encaminhá-lo para projetos de intervenção e reconhecer os descritores, ou habilidades, que precisam ser melhor trabalhadas na sala de aula.

A divulgação dos resultados do Avalia BH é realizada por meio de Oficinas de Apropriação dos Resultados oferecidas, anualmente, por especialistas do CAEd. Essas oficinas têm como público-alvo os gestores da rede municipal de educação (diretores, coordenadores pedagógicos, supervisores e técnicos da secretaria de educação) que após a capacitação, deverão atuar como multiplicadores do conhecimento apreendido nas suas escolas e comunidade escolar. Com carga horária de 16 horas, o curso objetiva mostrar como se dá o método avaliativo, desde a construção das matrizes de referência até a montagem dos cadernos de provas, orientar os gestores em relação à leitura dos dados e inculcar a importância da apropriação desses para a melhoria do processo ensino-aprendizagem.

Materiais impressos são produzidos e enviados a cada escola da rede para facilitar o acesso aos dados produzidos pelo Avalia BH, são eles: o Boletim do Sistema de Avaliação, que apresenta o programa, sua abrangência, as matrizes de referência, a metodologia de construção e análise dos testes; a Revista do Gestor, que contém informações sobre a participação, as proficiências médias e a distribuição do percentual de alunos nos padrões de desempenho definidos pela SME das nove Regionais de Belo Horizonte; o Boletim da Escola que mostra os resultados de cada escola e a sua interpretação pedagógica através da Escala de proficiência, dos padrões de desempenho e análise de itens do teste; e a Revista Contextual que descreve o perfil socioeconômico dos estudantes¹³. Esses materiais também são disponibilizados na internet, tanto no site da SME quanto no *site* do CAEd.

Merece destaque, ainda, na disseminação dos resultados do Avalia BH o Portal da Avaliação. Nesse domínio da internet, usuários cadastrados (gestores, professores, estudantes e seus pais) podem acessar seus resultados acadêmicos em todas as edições do Avalia BH, além das matrizes de referência.

1.5.3 A inclusão dos testes de proficiência em Ciências

Boletim Pedagógico da primeira edição do Avalia BH, publicado em 2009, sinalizava a inclusão de ciências da natureza nos ciclos de avaliação futuros. Em entrevista, a GAPED explica que, como não havia um banco de itens pré-testados disponíveis para utilizar na avaliação, nem tempo hábil para construí-lo, os testes de ciências, embora previstos, só puderam acontecer em 2010, ainda assim, apenas para os estudantes das 5ª e 9ª séries. Em 2011, passaram a ocorrer para as demais etapas do ensino fundamental, a partir da 3ª série.

A iniciativa da Secretaria Municipal de Educação de incluir ciências no Avalia BH partiu de uma demanda das escolas municipais que, habituadas com as avaliações sistêmicas de Língua Portuguesa e Matemática desde 2000 (SAEB, PROEB), indagavam o porquê de avaliar somente essas duas disciplinas e reclamavam um olhar externo sobre o processo de ensino aprendizagem em ciências. Essa iniciativa também sofreu forte influência do PISA: compreendeu-se, a

¹³ Disponíveis em: <<http://www.avaliabh.caedufff.net/diagnosticabh/colecao2012.faces>>. Acesso em: 25 nov. 2013.

partir dessa avaliação internacional, a importância de se aferir o nível de letramento científico dos estudantes mineiros desde os anos iniciais de escolarização, a partir de um currículo mais ajustado a realidade local.

A SME exigiu da GAPED, então, um redesenho do Avalia BH que ficou a cargo do órgão vencedor da licitação para a gestão do programa, o CAEd. O pontapé inicial foi a construção das matrizes de referência. Esses documentos foram criados por especialistas que balizaram o trabalho na Proposição Curricular para o Ensino Fundamental da RME-BH de Ciências, elaborada nos anos de 2007 e 2008 por professores da Rede Municipal de Educação de Belo Horizonte e da Universidade Federal de Minas Gerais,

tendo como referência o projeto político-pedagógico da Rede Municipal de Educação, os documentos oficiais – Parâmetros Curriculares Nacionais (PCN's), Diretrizes Curriculares, Sistema Nacional de Avaliação do Ensino Básico (SAEB), Sistema Mineiro de Avaliação (SIMAVE), Programa Nacional do Livro Didático (PNLD) – e a literatura da área de pesquisa em Educação e em ensino de Ciências. (BELO HORIZONTE, 2012, p. 6)

Com uma média de 20 descritores, as matrizes de Ciências do Avalia BH apresentam habilidades incluídas em quatro competências, a saber: Vida e Ambiente, Ser Humano e Saúde, Terra e Universo, Tecnologia e Sociedade. A primeira competência envolve habilidades de compreensão das características dos ambientes, dos seres vivos e suas relações, além do reconhecimento do equilíbrio ambiental. No eixo Ser Humano e Saúde é exigida a identificação de doenças e seus processos de prevenção e imunização, assim como, de hábitos saudáveis de vida. No eixo Terra e Universo, os estudantes devem reconhecer astros e fenômenos do Sistema Solar, além de reconhecer os processos químicos e geofísicos no planeta e no domínio Tecnologia e Sociedade, o aluno deve ser capaz de reconhecer processos de transformação da energia, compreender recursos tecnológicos e estabelecer relações entre as atividades humanas e seu reflexo no meio ambiente.

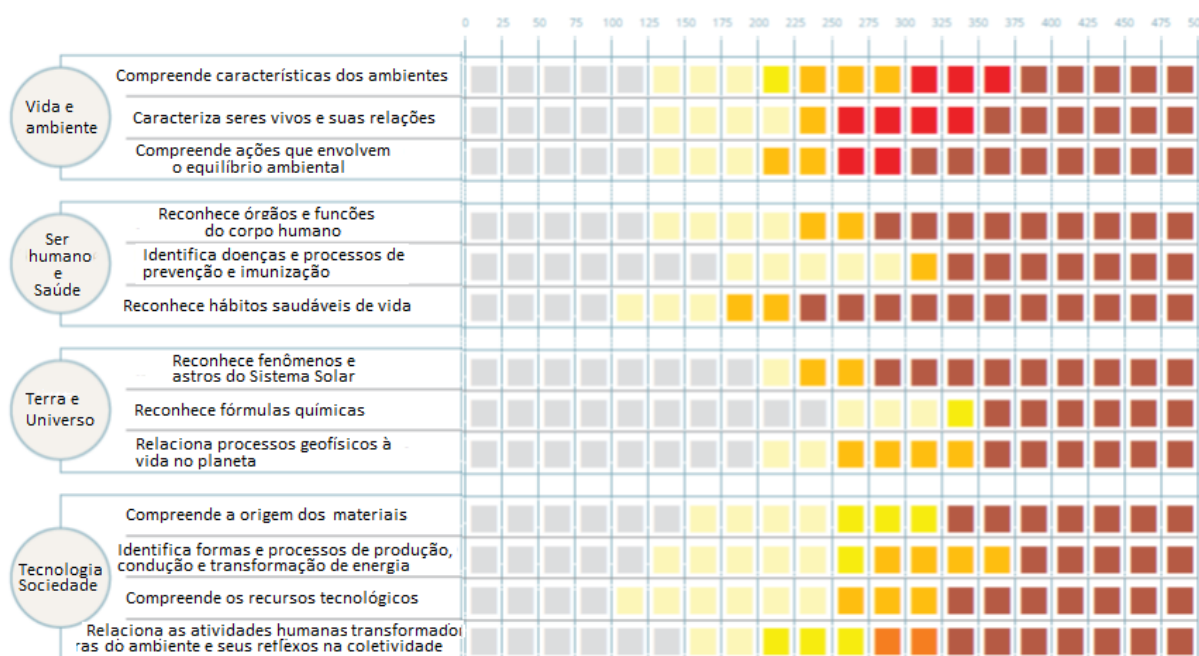
Esses eixos, frequentes nos currículos brasileiros de um modo geral, possibilitam a comunicação dos vários campos de conhecimentos que compõem a área de Ciências Naturais no Ensino Fundamental, a saber, Física, Química, Biologia, Geologia e Astronomia. De acordo com a SMEDBH, eles levam em conta os princípios específicos das disciplinas de ciências como historicidade, inter-relação, intencionalidade, aplicabilidade e provisoriedade; os conhecimentos prévios

dos educandos e a aproximação da vivência dos estudantes. Leva em conta, ainda, a relação espaço-tempo, os processos de transformação, as relações de diversidade, adaptação, energia entre outros (BELO HORIZONTE, 2010).

Anualmente, as matrizes de ciências da natureza são revistas pelos professores da rede junto a especialistas do CAEd, com o fim de melhor ajustá-la em relação à pertinência das habilidades cobradas e a capacidade de os descritores gerarem bons itens.

O Avalia BH também apresenta uma Escala de Proficiência de Ciências que foi elaborada a partir dos resultados dos primeiros testes e vem sendo complementada de acordo com as informações fornecidas pelos resultados, a cada ciclo de avaliação. O quadro 8, abaixo, retrata essa escala.

Quadro 8 - Escala de Proficiência de Ciências da Natureza do Avalia BH



Fonte: BELO HORIZONTE, 2012

Na primeira coluna dessa escala são apresentados os grandes domínios do conhecimento no campo das Ciências da Natureza, sendo agrupamentos de competências (segunda coluna) que, por sua vez, agregam as habilidades presentes na Matriz de Referência (BELO HORIZONTE, 2012). Para cada competência existem diferentes graus de complexidade, os quais são demonstrados através da gradação de cores. Dessa forma, a cor amarelo-claro indica o primeiro nível de complexidade da competência, passando pelo laranja, vermelho e chegando ao nível

mais complexo, representado pela cor marrom (BELO HORIZONTE, 2012). A Escala de Proficiência de Ciências do Avalia BH constitui um ineditismo do programa.

1.6 Retomada do problema de pesquisa

A descrição privilegiada, neste primeiro capítulo, dos principais sistemas de avaliação educacional que abrangem a aferição dos estudantes em relação ao desenvolvimento de competências e habilidades científicas, com o enfoque no programa municipal mineiro Avalia BH, traz ao debate alguns apontamentos, que, aqui, são resumidos em três: a grande maioria dos programas de avaliação educacionais aferem apenas habilidades de linguagem e matemática sob o argumento de que essas disciplinas são suportes para a aprendizagem de outras. Segundo essa ótica, o ensino de ciências é secundário, a despeito de sua importância para a formação de indivíduos capazes de se posicionarem em uma sociedade tecnológica, consumista e negligente em relação à sustentabilidade.

Um segundo ponto, mais otimista, relaciona-se ao fato de que, embora ainda relegado ao segundo plano, o ensino de ciências passa por um processo de valorização, haja vista a recente inclusão de testes relacionados a esse campo do conhecimento, desencadeado pela influência de avaliações internacionais. Por último, cabe destacar que a incipiência dos processos avaliativos do ensino de ciências aponta para a necessidade de estudos mais aprofundados sobre esse movimento: até que ponto ele tem contribuído para a diagnose dos problemas e elaboração de políticas que visem saná-los?

Essa necessidade é reforçada diante da informação primária que as avaliações em larga escala oferecem por meio de resultados traduzidos em médias de proficiência e padrões de desempenho: a deficiência no processo de ensino aprendizagem em ciências da natureza e suas tecnologias. Deficiência essa, que não era imprevista e sequer constitui novidade, a julgar pelo histórico e frustrado esforço brasileiro, em diferentes períodos e contextos socioeconômicos, para fazer do ensino de ciências o alicerce do desenvolvimento tecnológico ou da formação de cidadãos letrados cientificamente no sentido que o PISA, como visto, atribuiu a esse termo.

Nesse sentido, o próximo capítulo intenciona ampliar o espectro de informações sobre o Avalia BH, refinar o olhar ao seu desenho, mostrando os

bastidores do processo de implementação das avaliações em ciências e seus efeitos no sistema educacional de Belo Horizonte.

2. O AVALIA BH SOB ANÁLISE: RESULTADOS E IMPLICAÇÕES DOS TESTES DE CIÊNCIAS

O Avalia BH, como visto, propõe-se a produzir informações verossímeis sobre a realidade educacional, tornando possível a intervenção sobre essa realidade de forma pontual e relativamente rápida, por meio de políticas públicas ou projetos locais. O presente capítulo, nesse sentido, pretende responder às seguintes perguntas: tem o referido programa conseguido corresponder às suas expectativas em relação à avaliação da aprendizagem de Ciências? E, caso afirmativo, em que medida? Qual a percepção dos professores e gestores da rede sobre a avaliação em larga escala de competências científicas? Em suma, o esforço tem valido a pena?

A primeira seção trata da metodologia da pesquisa de campo e documental para conseguir respostas a tais indagações, enquanto as demais se dedicarão a trazer os resultados dessa pesquisa na tentativa de responder às indagações do parágrafo de abertura desse capítulo e, no limite, dar sustentação às proposições no capítulo 3.

2.1 Percurso metodológico

Eleveu-se a entrevista semiestruturada com membros da Gerência de Avaliação de Políticas Educacionais (GAPED), a aplicação de questionários a professores participantes de oficinas de elaboração de itens do Avalia BH, a análise de questionários de *feedback* previamente aplicados a gestores educacionais participantes de oficinas de apropriação de resultados do Avalia BH e a análise de documentos oficiais publicados pelo CAEd como procedimentos para a coleta de dados na presente pesquisa.

A entrevista foi realizada com membros da GAPED, órgão subordinado à Secretaria Municipal de Educação de Belo Horizonte, responsável pelo desenho e monitoramento do Avalia BH. O roteiro preconizou perguntas a respeito da motivação para a implementação dos testes de Ciências, do encadeamento desse processo, dos desafios surgidos e das mudanças dele decorrentes para a gestão do próprio programa, gestão escolar, professores da rede pública municipal, docência e ensino.

A aplicação dos questionários foi direcionada a um grupo de 23 docentes da rede pública municipal de Belo Horizonte que participavam de oficina de elaboração de itens, promovida pela GAPED em parceria com o CAEd em 2013. Essa metodologia, que envolveu oito perguntas fechadas e quatro questões discursivas, pretendeu colher dos professores a sua percepção do Avalia BH, em particular da avaliação de Ciências, e compreender a influência do programa em sua prática pedagógica e na escola em que leciona.

Os questionários de *feedback* foram aplicados, também em 2013, para participantes de Oficina de Apropriação dos Resultados de Língua Portuguesa, Matemática e Ciências do Avalia BH de 2012, um total de 99 profissionais da educação, a maioria deles, gestores. A análise desses questionários permitiu obter suas percepções sobre a importância da avaliação em larga escala e a inclusão de Ciências no programa, saber se eles compreenderam os objetivos e conteúdos da oficina, e a sua apreciação e sugestões sobre processos utilizados no Avalia BH.

Os documentos oficiais utilizados foram os boletins de divulgação de resultados e boletins pedagógicos publicados pelo CAEd e pela SMEd nos anos de 2008, 2009, 2010, 2011 e 2012.

2.2 A percepção dos docentes sobre a inclusão de ciências no AVALIA BH

Sabe-se, como afirma Castro (*apud* Souza, 2005), que a resistência ainda é encontrada, principalmente entre os professores, quando se trata de avaliações educacionais sistêmicas. Contrariando esse dado, porém, percebeu-se por meio de entrevista à GAPED e das respostas aos questionários, que os testes de Ciências em larga escala foram bem recebidos pelos gestores e docentes da rede municipal de educação de Belo Horizonte.

Membro da GAPED relatou em entrevista:

a implantação de qualquer avaliação sistêmica vai ser sempre muito complicada, vai ter resistência, sempre tem aquela crítica muito forte de que a avaliação veio para checar o professor, mas com Ciências foi diferente, porque os professores ficavam nos perguntando: porque só Português e Matemática?

Ao que outro membro complementou

Ciências entrou por ser uma das avaliações do PISA e os professores gostaram de poder comparar os nossos resultados com os das avaliações internacionais.

e continua

Já tínhamos avaliações de Português e Matemática no PROEB, na PROVA BRASIL e nas duas primeiras versões do Avalia BH, então, quando entrou Ciências, eles [gestores, coordenadores, docentes] gostaram: 'Porque só a de português e matemática? A gente quer saber mais'. Então, com Português e Matemática nós tivemos mais problemas, porque foi a implantação ainda. Quando chegou Ciências, a cultura já tinha sido implantada.

Os questionários aplicados aos professores que participavam de oficina de elaboração de itens também mostrou que os professores acharam importante a aferição das competências científicas dos estudantes. As respostas à pergunta: "Você considera relevante para a educação municipal a inclusão dos testes de Ciências no Avalia BH?" foram, em sua totalidade, respondidas de forma positiva. Selecionou-se algumas delas:

(Professor A) Sim. É uma disciplina importante na formação humana e cobrada em testes internacionais e que precisa de serem feitas reflexões sobre ela.

(Professor B) Sim. É necessária que seja realizada uma avaliação para que se tenham conhecimentos, da forma mais clara possível, de como está sendo desenvolvido o trabalho dos diversos conteúdos de ciências.

(Professor C) É essencial. Há tempos a escola vem priorizando o ensino de Língua Portuguesa e Matemática em detrimento das demais disciplinas que são tão importantes quanto para a formação do educando.

(Professor D) Sim, não podemos ficar presos à Língua Portuguesa e Matemática, é preciso reconhecer a importância das Ciências na formação do sujeito.

Constata-se, nos limites dessa pesquisa, que a implementação da avaliação em larga escala de competências científicas consistia mais do que em uma expectativa da classe docente, mas na demanda desses profissionais, que, habituados aos diversos ciclos de avaliações em larga escala das disciplinas de Língua Portuguesa e Matemática, consideraram a importância de conhecer mais profundamente o trabalho com a disciplina de Ciências, julgada por muitos, com a mesma importância daquelas. Essa constatação é alentadora já que, não existindo a barreira da resistência, os resultados do Avalia BH para ciências podem encontrar terreno fértil para gerar ações e estratégias de melhoria educacional.

2.3 A Repercussão da Inclusão e dos Resultados dos Testes de Ciências no Avalia BH

Esta seção apresenta os dados obtidos na pesquisa a respeito da repercussão da inclusão dos testes de ciências no Avalia BH, destacando os desafios advindos da ausência de uma referência nacional, um parâmetro a ser seguido ou que ao menos que sirva de comparação. Trata também dos resultados dos testes expressos em médias de proficiência e padrões de desempenho, mostrando o que o Avalia BH revelou sobre os estudantes do Ensino Fundamental da rede pública de Belo Horizonte.

2.3.1 O desafio da definição dos pontos de cortes na Escala de Proficiência pela Gerência de Avaliação de Políticas Educacionais (GAPED)

Os padrões de desempenho agrupam níveis da Escala de Proficiência que, por sua vez, traduzem as habilidades demonstradas pelos estudantes no momento da avaliação. Eles são definidos a partir de cortes numéricos, com base em metas educacionais estabelecidas pelo Avalia BH (Belo Horizonte, 2012). Compete às autoridades educacionais nos diferentes setores (gestores, supervisores e docentes) determinar esses cortes, julgando em que padrão os alunos deveriam se encontrar para que, dessa forma, seja possível analisar a distância entre o encontrado e o desejável (Fontanive *apud* Souza, 2005). Em Belo Horizonte, esses cortes foram definidos pela GAPED, com a orientação do CAEd.

O primeiro desafio relatado pela GAPED na implementação dos testes de ciências foi justamente essa definição de cortes para os padrões de Ciências da Natureza devido à falta de parâmetros. Para os testes de Língua Portuguesa e Matemática, o Avalia BH apropriou-se da escala de proficiência produzida pelo SAEB e também de seus pontos de corte, definidos e redefinidos ao longo de duas décadas de avaliação de conhecimentos apreendidos nessas disciplinas, porém, para Ciências, teve que inaugurar esse trabalho.

A GAPED relatou em entrevista:

Em ciências nós não tínhamos um parâmetro de escala. Então, testamos os cortes até achar os mais ideais. Em 2009 foi um corte, em 2010 vimos que não deu certo, então ajustamos os cortes... e você sabe que esse ano na Prova Brasil terá algumas questões de ciências. Talvez, então, agora a gente tenha, se o INEP começar a

avaliar ciências, [...] que fazer uma escala que seja comparável à escala do SAEB.

Outro membro desta Gerência reafirma:

O mais difícil é ajustar a escala de desempenho, porque [...] aqui em Belo Horizonte, por exemplo, nós não temos nenhum interesse em construir uma escala que valorize o pouco desempenho. Porque você pode construir uma escala e falar: o nível satisfatório dela está lá em baixo. Nós não temos interesse em construir uma escala assim”.

E pondera:

“Mas, no nosso caso, temos quatro níveis de desempenho, então decidir aonde cortar cada nível de forma que você não jogue para baixo o que irá considerar satisfatório, nem jogue para cima o que irá considerar baixo e que seja um corte justo, adequado para cada ano que avaliamos, acho que essa é a parte mais complicada.

A insegurança pela falta de parâmetros de comparação para as análises dos resultados de ciências e o receio de estar trilhando um caminho que pode conduzir ao erro é percebida durante todo o momento da entrevista: *“Eu sinto a avaliação de ciências como se a gente estivesse pisando em ovos. Com bastante cuidado, sabe? Com medo de estar errando”*, afirmou membro da GAPED, complementando:

É tudo muito novo e em tudo que você faz, você pode estar errando. Daqui a algum tempo, pode ser que a gente perceba que deveríamos ter feito outro corte, que isso aqui não está mostrando a realidade. Mas podemos estar acertando, ainda não sabemos, não temos como comparar. Português e matemática você não precisa enfrentar, já tem um lugar em que você se baseia. Então, eu acho que ciências vão ser só daqui a alguns anos, talvez com uns cinco anos, possamos voltar e dizer: isso aqui deu certo ou isso daqui eu tenho que mudar completamente.

Alicia Bonamino & Creso Franco já haviam denunciado, em 1999 e com relação ao SAEB, as controvérsias na interpretação das habilidades e competências associadas a diferentes pontos da escala de proficiência e, pautando-se em Gaviria & Klein, s.d., afirmaram que a correta interpretação das habilidades e competências “é um objetivo complexo e que só pode ser garantido a partir de um processo de estabilização em sucessivos exercícios de avaliação” (p. 125) e, além disso, requer qualidade e adequação dos testes ao perfil dos alunos (BONAMINO & FRANCO, 1999).

Nesse contexto, denota-se que, com apenas três ciclos de avaliação de ciências da natureza, o Avalia BH encontra-se, ainda, distante da estabilização, necessitando ser estudado, rediscutido e reajustado para que os resultados ganhem, com o tempo, maior fidedignidade.

2.3.2. Os padrões de desempenho e as médias de proficiência em Ciências

A tabela 3, abaixo, apresenta os cortes na escala de proficiência de Ciências da Natureza, definidos pela GAPED, que determinam os padrões de desempenho por série avaliada, enquanto a tabela 4, seguinte, apresenta as médias de proficiências em Ciências da Natureza dos estudantes nos ciclos avaliativos do AVALIA BH de 2011 e 2012.

Tabela 3 - Cortes na Escala de Proficiência que definem os padrões de desempenho por série

Níveis	3ºEF	4ºEF	5ºEF	6ºEF	7ºEF	8ºEF	9ºEF
Abaixo do Básico	Até 125	Até 150	Até 150	Até 175	Até 175	Até 200	Até 200
Básico	De 125 a 175	De 150 a 200	De 150 a 225	De 175 a 225	De 175 a 250	De 200 a 250	De 200 a 275
Satisfatório	De 175 a 250	De 200 a 275	De 225 a 300	De 225 a 300	De 250 a 325	De 250 a 325	De 275 a 350
Avançado	Acima de 250	Acima de 275	Acima de 300	Acima de 300	Acima de 325	Acima de 325	Acima de 350

Tabela 4 - Médias de Proficiência, por série, em Ciências, dos estudantes da rede municipal de Belo Horizonte - MG nos ciclos do AVALIA BH de 2011 e 2012¹⁴

Ano	Médias de proficiência por série						
	3EF	4EF	5EF	6EF	7EF	8EF	9EF
2011	200,57	219,74	226,4	236,04	233,42	235,46	254,21
2012	200,18	222,47	227,67	237,93	242,07	240,74	251,38

A partir da análise dessas tabelas, constata-se que, contrapondo ao observado no primeiro capítulo deste trabalho em relação aos resultados obtidos pelos estudantes no PISA e em programas de avaliação estaduais brasileiros, a maioria dos alunos da rede municipal de Belo Horizonte tem apresentado resultados considerados satisfatórios nos testes de Ciências, principalmente os que se

¹⁴ As médias de proficiência podem ser consultadas no portal da avaliação, disponível em: <<http://www.avaliabh.caedufjf.net/diagnosticabh/selecao.faces>>. Acesso em: 12 out. 2013.

encontram nas séries iniciais do Ensino Fundamental. É importante, no entanto, ressaltar que os estudantes encontram-se com média de proficiência próxima ao limite inferior do padrão de desempenho, isto significa que são poucas as habilidades que apresentam a mais do que os alunos que se encontram no padrão Básico.

Constata-se, também, que de um ciclo para o outro, não houve mudanças significativas em relação ao avanço ou retrocesso das médias de proficiência, o que era esperado, devido ao curto espaço de tempo entre esses ciclos.

O cruzamento dos dados da tabela 3 com os da tabela 4 mostra, ainda, que até o 6º ano do Ensino Fundamental, os estudantes encontram-se no nível considerado Satisfatório, enquanto nos anos finais do Ensino Fundamental, encontram-se no nível Básico. Por Satisfatório, entende-se o padrão em que se encontram alunos que demonstraram um vasto leque de habilidades que têm maior grau de complexidade e, assim, exigem um refinamento dos processos cognitivos; por Básico, entende-se o padrão em que se encontram alunos que estão no início de um processo de sistematização e domínio de habilidades consideradas básicas e essenciais ao período de escolarização que cursam (BELO HORIZONTE, 2012). A SMED destaca a necessidade de investimento de esforços nesse último grupo, a fim de que possam desenvolver habilidades mais elaboradas (BELO HORIZONTE, 2012).

Em boletim pedagógico do Avalia BH de 2012, a SME explicita que a expectativa primordial para os estudantes concluintes do 3º ano do ensino fundamental é que sejam capazes de utilizar os conhecimentos alfabéticos nas diferentes práticas sociais, sendo que os conhecimentos na área de ciências da natureza exercem importante função nesse aspecto (BELO HORIZONTE, 2012). Esse mesmo documento relata a importância do trabalho com a saúde na alfabetização, destacando a necessidade de desenvolver na criança comportamentos e atitudes de cuidado com o próprio corpo e que visem, também, a saúde coletiva:

a saúde dos alunos deve ser considerada como um bem individual e comum; dessa forma, ao mesmo tempo em que os indivíduos têm que se cuidar, alimentando-se bem, escovando os dentes, tomando banho, dormindo cedo, realizando atividades físicas, deve-se considerar que atitudes como não jogar papel no chão, não poluir o rio, dentre outras, contribuem para a saúde do próximo também, pois

o lixo pode favorecer o aparecimento de doenças (BELO HORIZONTE, 2012, p. 55).

O quadro 9 classifica os alunos do 3º ano do Ensino Fundamental em diferentes padrões de desempenho de acordo com as habilidades em ciências que apresentaram no momento do teste.

Quadro 9 - Padrões de Desempenho em Ciências dos avaliandos do 3º ano do Ensino Fundamental no AVALIA BH 2012

Padrão de desempenho	Média de Proficiência	Habilidades apresentadas nos testes
Abaixo do Básico	Ate 125	Identificam alimentos saudáveis e não saudáveis.
Básico	De 125 a 175	Identificam, a partir de imagens, tipos de ambientes terrestres, o ambiente aquático e os construídos pelo homem; reconhecem que os seres vivos necessitam de recursos naturais para a sobrevivência; discriminam as partes fundamentais de um vegetal; relacionam animais ao seu hábitat; identificam hábitos elementares de higiene e hábitos saudáveis de alimentação; identificam materiais recicláveis.
Satisfatório	De 175 a 250	Identificam processos de transformação do ambiente pelo homem; relacionam o tratamento de água à retirada de impurezas nela presentes; identificam as etapas do processo de transformação da matéria-prima na confecção de produtos industrializados; reconhecem as fases de vida de um vegetal; reconhecem a origem dos alimentos consumidos pelo homem; reconhecem a troca de denteição como sinal de que a criança está crescendo; identificam o açúcar no ápice da pirâmide alimentar, inferindo que o seu consumo deve ser reduzido.
Avançado	Acima de 250	Identificam as principais características que definem um ser vivo; identificam características do lixo orgânico e do inorgânico; compreendem a função da troca de denteição; reconhecem a importância dos fósseis para o estudo dos seres vivos e identificam a matéria-prima de objetos produzidos pelo homem.

Fonte: Belo Horizonte, 2012. Adaptado.

Como observado, a média de proficiência dos estudantes do 3º ano do ensino fundamental nos testes de ciências os classificam no padrão Satisfatório. Nesse sentido, os resultados do Avalia BH mostram que esses alunos correspondem às expectativas da SMEd em relação ao desenvolvimento de habilidades no campo da saúde, como reconhecer hábitos de higiene e de alimentação saudável, a importância do tratamento da água e a importância da troca de denteição, bem como

interpretar uma pirâmide nutricional, inferindo a respeito dos alimentos que devem ser consumidos em maior ou menor quantidade.

As propostas curriculares para o segundo ciclo do ensino fundamental, que corresponde às 4^a, 5^a e 6^a séries, estabelecem que

na área das Ciências da Natureza, o grande objetivo é oferecer aos educandos a oportunidade de aprender Ciências, ampliar suas curiosidades, incentivá-los a levantar hipóteses e a construir conhecimentos sobre os fenômenos químicos e físicos, sobre os seres vivos e sobre a relação entre o homem e a natureza e entre o homem e a tecnologia. O ensino de Ciências deve levar o educando a utilizar os conhecimentos aprendidos, a desenvolver atitudes para uma transformação pessoal, favorecendo sua integração na sociedade, na natureza, demonstrando conhecer e entender seu papel dentro do mundo dos seres vivos. (BELO HORIZONTE, 2009, p. 44)

Nesse sentido, essa etapa da escolarização parece também atender ao esperado pela SMEd quanto ao desenvolvimento de habilidades em ciências da natureza. Os alunos desse ciclo, que, em média, encontram-se no nível Satisfatório, demonstraram ter desenvolvido um maior número de habilidades científicas e, essas, com maior grau de complexidade e de inter-relação com a vivência deles. Essas habilidades são descritas no quadro 10, a seguir, nos padrões Abaixo do básico, Básico e Satisfatório.

Quadro 10 - Padrões de Desempenho em Ciências dos avaliandos do 2º ciclo do Ensino Fundamental no AVALIA BH 2012

Nível	4EF	5EF	6EF	Habilidades
Abaixo do Básico	Até 150	Até 150	Até 175	Reconhecem que o desmatamento, as queimadas e a poluição por deposição de lixo são ações humanas que prejudicam o solo; reconhecem, a partir de imagens, os instrumentos de medida desenvolvidos pelo homem; reconhecem o choque elétrico como manifestação da energia elétrica.
Básico	150 a 200	150 a 225	175 a 225	Reconhecem uma cadeia alimentar; as fases do ciclo de vida de um animal e os animais vertebrados, a partir de imagens. Identificam as consequências do descarte inadequado do lixo e compreendem a importância do tratamento do esgoto para a qualidade de vida. Relacionam os órgãos de sentido à sua função; a presença das características sexuais secundárias à atividade hormonal; o hábito de fumar ao aparecimento de doenças humanas. Reconhecem alimentos indicados para uma alimentação saudável; as atividades físicas que causam a aceleração da respiração e dos batimentos cardíacos e as estruturas responsáveis pela sustentação do corpo humano. Associam instrumentos do cotidiano ao tipo de informação que eles proporcionam; identificam as transformações físicas da água; reconhecem as principais matérias-primas dos produtos industrializados; identificam planetas do Sistema Solar; diferenciam corpos celestes luminosos dos iluminados; reconhecem as condições para que ocorra um eclipse. Reconhecem a importância da energia elétrica para o funcionamento de equipamentos utilizados no cotidiano e diferenciam materiais recicláveis dos não recicláveis.
Satisfatório	200 a 275	225 a 300	225 a 300	Reconhecem a função de órgãos e sistemas do corpo humano. Reconhecem ações coletivas que diminuem impactos ambientais. Compreendem o conceito de biodiversidade; discriminam as funções das partes fundamentais dos vegetais; diferenciam animais vertebrados de invertebrados. Conhecem as principais características e relações dos astros Sistema Solar, identificam alguns processos de transformação de energia em aparelhos elétricos.
Avançado	Acima de 275	Acima de 300	Acima de 300	Diferenciam solos como próprios e impróprios para a agricultura; exemplificam organismos decompositores; reconhecem que a urina é formada por água e substâncias retiradas do sangue através dos rins; identificam as situações que provocam desequilíbrios em teias e cadeias alimentares e reconhecem o processo de transformação de energia luminosa em química nos vegetais durante a fotossíntese.

Fonte: Revista Pedagógica de Ciências da Natureza - 2º ciclo

Encontrar-se nos padrões de desempenho mais elevados indica, sem dúvida, que o ensino de ciências na rede municipal de educação de Belo Horizonte tem trilhado um caminho acertado. No entanto, a própria SMEd salienta que mesmo os alunos posicionados no Satisfatório e Avançado necessitam de atenção e estímulos para que progridam cada vez mais (BELO HORIZONTE, 2012). Outro dado importante, é que os resultados apresentados até agora trabalham com média aritmética, o que significa que, apesar de a maioria dos estudantes do terceiro ano e do segundo ciclo do ensino fundamental encontrar-se no nível Satisfatório, ainda existem muitos alunos nos padrões Abaixo do Básico e Básico.

A rede municipal de educação de Belo Horizonte almeja que seus alunos alcancem pelo menos o nível Básico e que a maior parte se situe nos níveis Satisfatório e Avançado: “Logo, não importa apenas elevar a média, importa também tomar cuidado para não deixar nenhuma criança para trás nem permitir que se ampliem as desigualdades” (BELO HORIZONTE, 2012, p. 9). É necessário, destaca a SMEd, refletir sobre estratégias pedagógicas e ações educacionais voltadas para ampliar o número de crianças nos níveis mais elevados: “Deseja-se que não haja alunos com dificuldades de aprendizagem ou de acompanhar os conteúdos previstos para cada série, ou seja, que o percentual “Abaixo do básico” esteja cada vez menor até anular-se” (BELO HORIZONTE, 2012, p. 10).

Para o terceiro ciclo, que corresponde às 7ª, 8ª e 9ª séries do ensino fundamental, as proposições curriculares da rede municipal de Belo Horizonte enfatizam os desafios do trabalho com estudantes que se encontram na fase da adolescência. Para a SMED, o 3º ciclo é o tempo propício para a síntese dos conteúdos ministrados no ensino fundamental e consolidação de capacidades, sendo caracterizado por uma organização melhor sistematizada e uma fundamentação teórica mais ampla (BELO HORIZONTE, 2009). Nesse ciclo, “é possível incorporar práticas que possibilitam relacionar as dimensões formadoras dos adolescentes, tornando os estudos mais significativos para eles, como o trabalho em torno de problemas e questões do nosso tempo” (p. 20).

Os alunos dos anos finais do ensino fundamental, em média, situam-se no padrão Básico. Isto é, ainda estão em um patamar aquém do esperado. O quadro 11, que se segue, apresenta as habilidades desenvolvidas por esses estudantes, de acordo com a média de proficiência obtida e a classificação nos padrões de desempenho.

Quadro 11 - Padrões de Desempenho em Ciências dos avaliandos do 3º ciclo do Ensino Fundamental do AVALIA BH 2012

Nível	7EF	8EF	8EF	Habilidades
Abaixo do Básico	Até 175	Até 200	Até 200	Reconhecem a importância das etapas do processo de tratamento da água; formas de prevenção comuns a diversas doenças contagiosas; a capacidade de um ímã de atrair objetos metálicos. Relacionam qualidade de vida ao controle de condições prejudiciais ao ambiente e o descarte inadequado de óleo à contaminação dos rios.
Básico	175 a 250	200 a 250	200 a 275	Localizam, em representação do corpo humano, órgãos dentro de um sistema e reconhecem, em um esquema do ciclo menstrual, o período fértil da mulher. Identificam os modos de transmissão e prevenção de algumas doenças; reconhecem as principais características de alguns avanços científicos e suas aplicações na saúde e na alimentação. Reconhecem as causas antrópicas de desequilíbrios ambientais; os tipos de soros adequados contra diferentes peçonhas; as diversas formas e processos de transformação de energia; e o gás oxigênio como produto da fotossíntese. Identificam comportamentos que promovam a preservação dos recursos naturais; os produtos derivados do petróleo; os mecanismos gerais de funcionamento de máquinas simples e suas aplicações e as substâncias hidrossolúveis e lipossolúveis. Relacionam troca de calor com as transformações físicas da água e os movimentos da Terra aos dias, noites e estações do ano.
Satisfatório	250 a 325	250 a 325	275 a 350	Compreendem a integração dos sistemas do corpo humano, o processo da digestão. Relacionam células e funções e identificam os componentes do sangue. Diferenciam soros de vacinas; reconhecem o conceito de gravidade e o funcionamento de um circuito elétrico simples. Reconhecem fórmulas de substâncias comuns; as propriedades do som; conceitos básicos da dinâmica aplicados ao cotidiano; a relação entre frequência, comprimento de onda e a formação de sons graves ou agudos. Interpretam gráficos sobre a composição do ar e diferenciam substâncias orgânicas das inorgânicas envolvidas na fotossíntese.
Avançado	Acima de 325	Acima de 325	Acima de 350	Identificam que o peso é dependente da massa e da gravidade; fórmulas de ácidos; e as células sanguíneas que transportam gases. Reconhecem as condições necessárias para a propagação do som e os códigos usados para descrever os mecanismos de herança genética. Compreendem a condensação da água e diferenciam misturas de reações químicas.

Fonte: Belo Horizonte, 2012. Adaptado.

Observa-se que esses alunos, apesar de já terem iniciado o processo de sistematização de habilidades científicas relacionadas à sua vivência, como a de reconhecer o período fértil da mulher em um esquema do ciclo menstrual, o que é imprescindível para meninos e meninas que iniciaram a vida reprodutiva, eles ainda apresentam déficits de aprendizagem quando se trata de habilidades relacionadas à física, por exemplo, como reconhecer os conceitos básicos da dinâmica aplicados ao cotidiano.

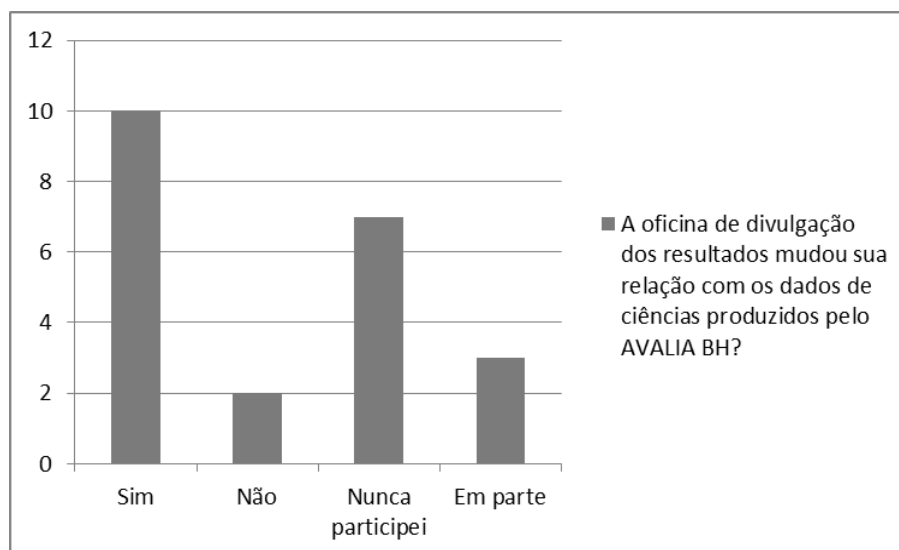
2.3.3 A divulgação dos resultados do Avalia BH em Ciências

Nesse momento da pesquisa, procurou-se compreender como os dados da avaliação em larga escala de ciência têm chegado até os gestores e professores. Como já informado na descrição realizada no primeiro capítulo, os resultados produzidos pelos testes do Avalia BH são divulgados pela internet, no Portal da Avaliação, e por meio de oficinas anuais de divulgação/apropriação de resultados oferecidas pelo CAEd a gestores, técnicos e docentes da rede municipal de Belo Horizonte.

O questionário aplicado aos professores que participavam de oficina de elaboração de itens trouxe duas perguntas com o fim de investigar se o docente acompanha esses resultados, seja pela internet, boletins pedagógicos, reuniões na sua escola ou participação em oficinas de divulgação; e se a sua forma de divulgação por meio de oficinas mudou a relação docente com os dados de avaliações em ciências, isto é, se as informações têm chegado de forma clara e compreensível até esses profissionais.

Dos 23 professores respondentes, 12 afirmaram sempre acompanhar os resultados, 8 responderam acompanhar quase sempre e 3 deles disseram que quase nunca acompanham os resultados de ciências do Avalia BH. Em relação à segunda pergunta, o gráfico 1, abaixo, reproduz as respostas dos docentes.

Gráfico 1 - Respostas dos docentes à pergunta: A oficina de divulgação dos resultados mudou sua relação com os dados de Ciências produzidos pelo Avalia BH?



Observa-se que dez professores afirmaram ter mudado seu olhar sobre os dados de avaliação em larga escala de ciências a partir de informações fornecidas nessas oficinas; 7 professores alegaram nunca ter participado desses eventos; 3 reconheceram que as oficinas afetaram sua relação com dados de avaliação de forma parcial e 2 professores disseram que não, as oficinas não mudaram suas relações com resultados da avaliação em ciências do Avalia BH.

A partir dessas respostas, infere-se que: os professores têm interesse pelos resultados de Ciências, já que a maioria procura acompanhá-los; muitos professores não têm sido contemplados para a participação nas oficinas de divulgação dos resultados; existe um desencontro entre as expectativas dos docentes em relação à divulgação dos resultados e a forma como eles são explanados, já que uma parte considerável (5 professores, isto é, aproximadamente 22% do grupo pesquisado) não conseguiu compreendê-los totalmente ou os compreenderam apenas parcialmente nas oficinas voltadas para esse objetivo.

As duas últimas inferências são preocupantes já que a falta de compreensão pode estar relacionada à inapropriação ou apropriação de forma insuficiente desses resultados. Diana Cerdeira (2012) lembra que notícias divulgadas na mídia, objeções propaladas pelos sindicatos de professores e artigos acadêmicos mostram que um dos impactos das políticas de avaliação educacional é a formação de movimentos de resistência e que uma das hipóteses para explicar essa rejeição aos indicadores de desempenho está atrelada à “falta de compreensão ou uma compreensão superficial

dos sistemas de avaliação que caracterizaria um ‘analfabetismo técnico’ por parte de professores e gestores, o que os manteriam distantes e alheios a esses sistemas” (p. 127).

Sérgio Eustáquio da Silva (2010) em dissertação de mestrado em que estudou a “Utilização e apropriação do Avalia BH pelos diretores, coordenadores pedagógicos e professores da rede municipal de educação de Belo Horizonte” chegou às seguintes conclusões:

- a grande maioria dos professores entrevistados [...] demonstram muito pouco conhecimento e apropriação das matrizes avaliativas, dos descritores e, principalmente, de como utilizar tais resultados como ferramenta pedagógica [...];
- ao que parece indicado pelas entrevistas, até mesmo o processo primário de informar os professores sobre os resultados demonstra-se falho [...] muitos professores deixam claro que desconhecem os objetivos da avaliação, os aspectos metodológicos envolvidos na construção de tais instrumentos e até mesmo desconhecem os resultados enviados às escolas em forma de boletins pedagógicos (p.82).

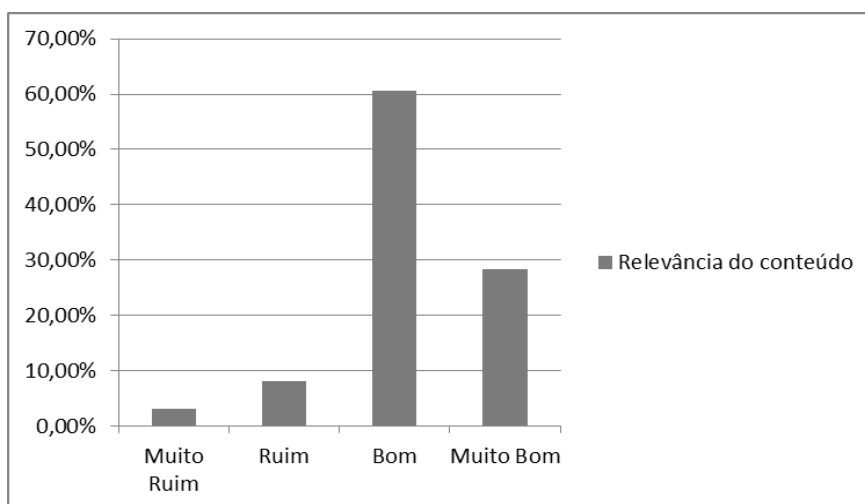
É premente, portanto, não só a necessidade de inserir os docentes nos processos de avaliação, propiciando-lhes a vivência em oficinas de divulgação/apropriação dos resultados do Avalia BH, como a de que tanto essas oficinas como outros meios de divulgação superem a mera informação dando lugar à discussão de indicadores oficiais de desempenho para que esses profissionais consigam transformar números em ações na sala de aula.

Quando se analisa os questionários respondidos pelos 99 participantes da Oficina de Apropriação dos Resultados do Avalia BH de 2012, em sua maioria, gestores, mais uma vez se observa que os profissionais da educação se interessam pelos resultados da avaliação externa (gráfico 2), mas, em contrapartida, que uma parcela considerável da amostra pesquisada não consegue compreender em sua totalidade os resultados de avaliação em larga escala, mesmo após a oficina (gráficos 3 e 4).

O gráfico 2, a seguir, apresenta o percentual das respostas Muito ruim, Ruim, Bom e Muito bom dos gestores em relação à relevância do conteúdo da oficina de divulgação dos resultados do Avalia BH 2012. Segundo *portfolio* da oficina, o seu conteúdo tratou das características das avaliações em larga escala, da descrição do programa Avalia BH, da apresentação e discussão das Matrizes de Referência de Língua Portuguesa, Matemática e Ciências, da análise dos resultados de

participação e dos resultados de desempenho nessas três disciplinas, por meio da leitura e interpretação das Escalas de Proficiência e dos Quadros Padrões de Desempenho.

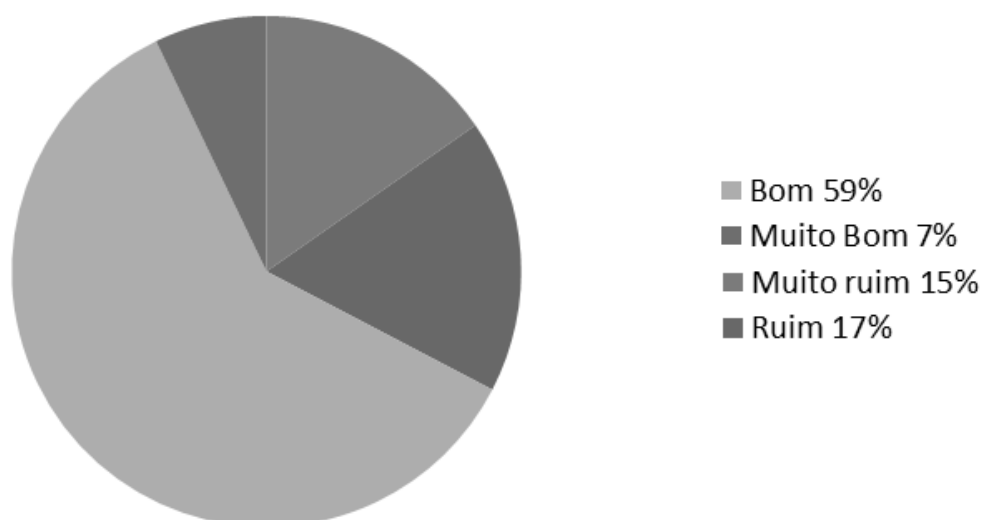
Gráfico 2 - Respostas dos gestores em relação à relevância do conteúdo da oficina de divulgação de resultados de Língua Portuguesa, Matemática e Ciências do AVALIA BH 2012



Esse gráfico mostra que quase 90% do total de respondentes consideraram o conteúdo da oficina como Bom ou Muito bom. Isto é, o tema das avaliações em larga escala é relevante para os profissionais da educação. Essa informação é útil no sentido de mostrar que os possíveis problemas de compreensão dos dados de avaliação em larga escala não estão associados à negligência dos educadores em relação a eles. Silva (2010) encontrou cenário semelhante em sua, já citada, pesquisa, em que a entrevista com gestores e coordenadores pedagógicos revela que a grande maioria concorda com a avaliação sistêmica, o que indica “certa sedimentação” por parte dos gestores acerca da importância dos testes de desempenho padronizados.

O próximo gráfico, no entanto, que irá mostrar as respostas dos gestores em relação à percepção do preparo oferecido pela oficina para a apropriação dos resultados, apresentará um acréscimo no número de respostas Muito ruim e Ruim.

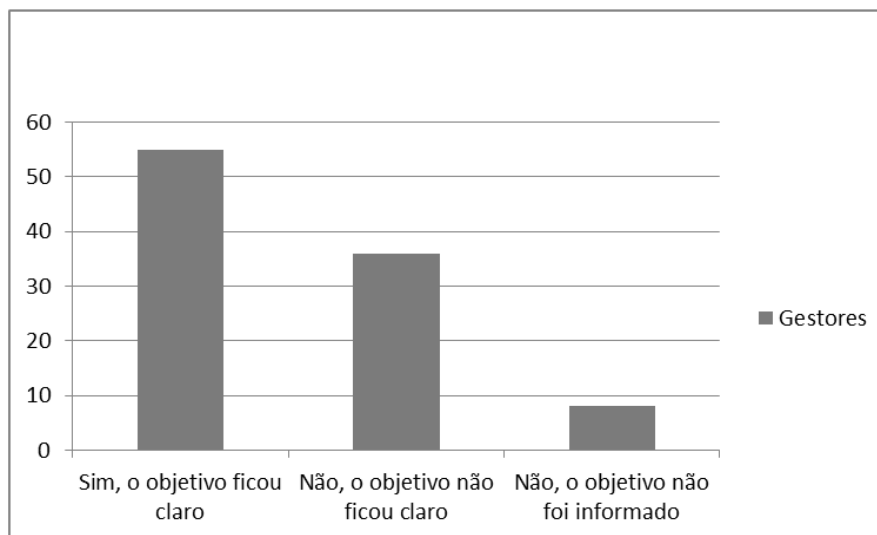
Gráfico 3 - Respostas dos gestores quando perguntados sobre o que acharam do preparo oferecido pela oficina para a apropriação dos resultados de Língua Portuguesa, Matemática e Ciências dos testes do Avalia BH de 2012.



Observa-se que, embora o número de cursistas que se julgam preparados para a apropriação de resultados dos testes do Avalia BH seja maior, 66%, a quantidade que considerou a preparação ineficiente é considerável, formando um total de 32% de gestores que, provavelmente, não estão aptos a trabalharem com resultados de avaliação em larga escala para gerar melhorias educacionais.

O gráfico 4, a seguir, reproduz as respostas dos gestores quando questionados se o objetivo da oficina de divulgação ficou claro para eles. Segundo cartas convites elaboradas pelo CAEd para os cursistas que participaram das oficinas de divulgação dos resultados do Avalia BH, o propósito desse evento é “capacitar os participantes nos temas concernentes à avaliação em larga escala e à interpretação e compreensão dos resultados dessas avaliações.” Essas cartas afirmam, ainda, que os participantes aprenderão a analisar os resultados gerais de participação e proficiência dos estudantes obtidos nos testes de Língua Portuguesa, Matemática e Ciências para que as instâncias gestoras possam planejar a execução de políticas públicas, criar metas de qualidade e equidade educacionais e promover mecanismos de formação continuada e para que os gestores das unidades escolares possam elaborar a sua avaliação institucional e o projeto pedagógico da escola, bem como monitorar a qualidade do ensino ofertado.

Gráfico 4 - Respostas dos gestores participantes de oficina de divulgação dos resultados de Língua Portuguesa, Matemática e Ciências de 2012 em relação ao objetivo dessa oficina.

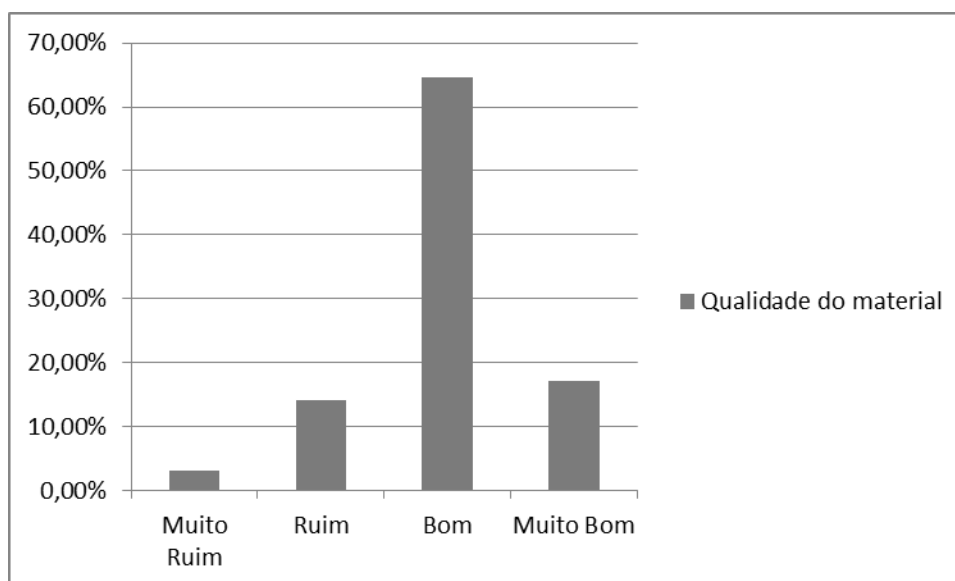


Como se observa, apenas pouco mais da metade dos gestores conseguiu compreender os objetivos da oficina da qual participaram, 36 alegaram não tê-lo compreendido e 8 afirmaram que sequer foram informados sobre. Isto é, um total de 44% dos gestores participou da oficina sem ao menos reconhecer a sua função.

A análise dos gráficos que se seguem pode apontar o momento em que está ocorrendo a lacuna entre o objetivo das oficinas de divulgação de resultados e a demanda do profissional da educação a respeito dos dados de avaliação educacional externas. Os gráficos 4, 5 e 6 apresentam o julgamento dos gestores em relação aos seguintes aspectos da oficina: esclarecimentos e explicações proporcionados por especialistas em avaliação educacional nas áreas de Língua Portuguesa, Matemática e Ciências; atividades propostas pelos especialistas com o fim de dinamizar a aprendizagem e estimular a participação e surgimento de ideias e a qualidade do material disponibilizado que inclui a apresentação dos temas em *slides*, cartazes contendo gráficos e tabelas explicativos e as revistas pedagógicas.

O gráfico 5 apresenta o percentual das respostas Muito ruim, Ruim, Bom e Muito bom dos gestores em relação à qualidade do material fornecido durante a oficina.

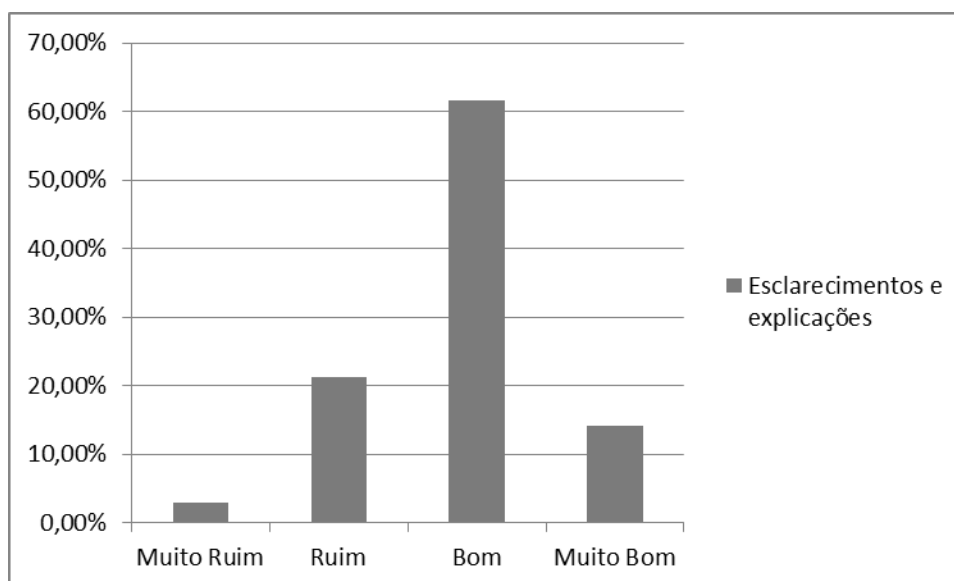
Gráfico 5 - Respostas dos gestores em relação à qualidade do material oferecido na oficina de divulgação de resultados do Avalia BH 2012 para a compreensão dos dados de avaliação em larga escala em Língua Portuguesa, Matemática e Ciências



Observa-se que os gestores, em sua maioria, julgam como Bom ou Muito Bom quando se trata da qualidade do material de divulgação, demonstrando, dessa forma, que o que separa a divulgação da compreensão adequada dos resultados por parte dos gestores não se relaciona diretamente ao material utilizado com esse fim.

O gráfico 6 apresenta o percentual das respostas Muito Ruim, Ruim, Bom e Muito Bom fornecidas pelos gestores em relação aos esclarecimentos e explicações dos especialistas em avaliação externa do CAEd sobre os dados de avaliação educacional externa.

Gráfico 6 - Percentual de respostas dos gestores participantes de oficina de divulgação dos resultados de Língua Portuguesa, Matemática e Ciências do Avalia BH 2012 em relação aos esclarecimentos e explicações sobre dados de avaliação externa.



Nota-se que o percentual de respondentes insatisfeitos, os que marcaram Ruim e Muito ruim como respostas, começa a crescer, chegando a quase 25%. Soma-se a esse dado, as inquietações que alguns cursistas demonstraram por meio de recados escritos nos versos dos questionários. Um gestor sugeriu: “A oficina foi muito boa, mas sugiro que haja um levantamento das questões nas escolas junto aos coordenadores sobre avaliação externa porque muitas dúvidas não foram/puderam ser respondidas”. Outro participante da oficina foi mais contundente na exigência por uma maior interação entre especialistas em avaliação e escola: “Antes de falar, de apresentar algo para alguém, é preciso saber ouvir. Não adianta dar sobremesa para uma pessoa que tem fome de comida. Dos resultados já tomamos conhecimento, mas e o trabalho dos elaboradores da avaliação externa em conjunto com os coordenadores e professores?”.

Algumas reclamações foram também observadas em relação às explicações sobre procedimentos e metodologias da avaliação em larga escala, assim como seus usos: “Senti falta de maiores explicações da Escala de Proficiência”; “Gostaria que trouxessem profissionais para responder sobre os diversos processos da avaliação (itens, escolhas, elaboração, seleção, análise dos níveis dos itens)”; “Seria mais produtivo se trouxessem apresentações e estudos dos resultados das regionais e o que significam esses resultados”; “poderiam ter preparado um material com

itens, com maior índice de erros, com sugestões de abordagens”, “os especialistas se detiveram em ministrar aula de ciências em vez de discutir os resultados”.

Depreende-se dessas reclamações que considerável parte dos gestores tem sede de conhecer e entender mais profundamente como funciona a avaliação em larga escala e que essa demanda não foi suprida ou foi remediada de forma precária na oficina de divulgação dos resultados. E, ainda, que esse fato pode estar relacionado ao de que os avaliadores, com receio de não conseguirem se fazer entender ao explicarem termos complexos que envolvem a avaliação sistêmica, como a Teoria de Resposta ao Item (TRI), levam para as oficinas apenas aspectos teóricos superficiais ou, pior, disparam em cima dos gestores um acúmulo de tecnicismos, através de uma linguagem inadequada para tratar com um público ainda leigo no assunto.

Nilma Fontanive (2005) destaca que

embora a apresentação clara dos resultados da avaliação e sua apropriação pelas autoridades educacionais, professores e pais não seja condição suficiente para produzir mudanças, ela é, sem dúvida, uma condição necessária. Portanto, os grupos envolvidos em conduzir as avaliações externas devem ter o compromisso de realizar esforços no sentido de tornar os resultados acessíveis ao público não acostumado a leituras técnicas, especializadas nesse campo. (p. 47)

Para essa autora, os relatórios técnicos pedagógicos destinados aos diretores das escolas e aos professores devem ser capazes de fornecer informações inteligíveis, que sugiram mudanças na prática cotidiana da sala de aula, de forma clara, simples e atraente (FONTANIVE, 2005).

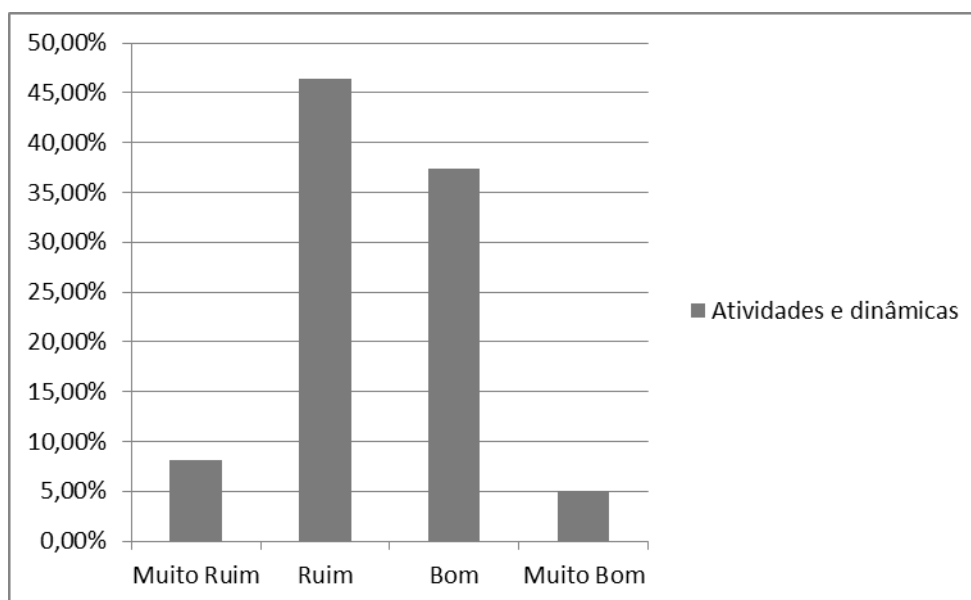
Outro autor também chama a atenção para esse fato, afirmando que, embora a avaliação da educação exija o domínio de técnicas que podem ser estatisticamente complexas, é possível que um leigo consiga compreender os resultados:

Qualquer autor responsável escrevendo para um público mais amplo tem a obrigação de verter o jargão técnico em termos compreensíveis para os não iniciados. Não é impossível e nem particularmente difícil para um leigo em avaliação entender os resultados de um estudo escrito de forma clara e sem tecnicismos. (p. 247).

Sendo assim, é forçoso que os avaliadores conheçam a real necessidade dos gestores e professores e adequem a linguagem utilizada em encontros com esses profissionais. Para isso, antes de tudo, é preciso ouvi-los, como afirma Freire (1998), “não é falando aos outros, de cima para baixo, sobretudo, como se fôssemos os portadores da verdade a ser transmitida aos demais, que aprendemos a escutar, mas é escutando que aprendemos a falar com eles” (p.127).

A seguir, tem-se uma informação mais enfática a respeito dessa discussão. Reproduzindo o julgamento dos gestores em relação às atividades e dinâmicas oferecidas na oficina de divulgação dos resultados, o gráfico 6 mostra que: 8% desses profissionais da educação consideraram-nas Muito Ruim; 46% as consideraram Ruim; 37% marcaram Bom nos questionários e 5% consideraram Muito Bom.

Gráfico 7 - Percentual de respostas dos gestores em relação às atividades e dinâmicas oferecidas na oficina de divulgação de resultados do Avalia BH 2012.



Isto é, mais da metade dos participantes da oficina, 56%, considerou que o evento realizado para a divulgação e apropriação dos resultados do Avalia BH não cumpriu o papel que se propôs a realizar no sentido do que se conceitua como oficina. E, mais uma vez, os comentários deixados nos versos dos questionários podem elucidar o sentimento de grande parte dos gestores ao final da oficina: “Para mim, o que foi feito não é formato de oficina, e sim, uma formação teórica do Avalia BH”; “Sugiro que as próximas oficinas sejam de fato práticas”; “a oficina se ateve a

questões históricas e aprofundou pouco nos resultados”; “sugiro que na próxima oficina exista uma explanação com perguntas/dúvidas frequentes, que exista trabalho em grupo e sugestões de atividades para trabalhar com os resultados”.

Maria Figueirêdo *et al* (2006) explicam que a proposta metodológica de oficina pedagógica objetiva a apreensão do conhecimento através da inter-relação entre a teoria e a prática. Os autores afirmam que a oficina deve constituir

um espaço de construção coletiva do conhecimento, de análise da realidade, de um confronto e troca de experiências. A atividade, a participação, a socialização da palavra, a vivência de situações concretas através de sociodramas, análise de acontecimentos, a leitura e a discussão de textos, [...] são elementos fundamentais na dinâmica das oficinas pedagógicas (p. 3).

E, citando Candau (1995), ela destaca, ainda, que as oficinas são “unidades produtivas de conhecimentos a partir de uma realidade concreta, para serem transferidas a essa realidade a fim de transformá-la” (p.3).

Nesse sentido, conclui-se que a transformação da realidade educacional por meio das avaliações sistêmicas depende não somente de um trabalho sério na produção de dados mas, também, na forma de repassá-los ao público formado por gestores, coordenadores, professores e comunidade escolar em geral.

2.3.4 Os usos dos resultados do Avalia BH em Ciências

Esse tópico se dedica a investigar se os resultados dos testes de Ciências do Avalia BH têm influenciado a prática profissional de gestores e professores, bem como estimulado a criação de projetos de intervenção ou ações no sentido de promover melhorias no ensino e aprendizagem dessa área do conhecimento.

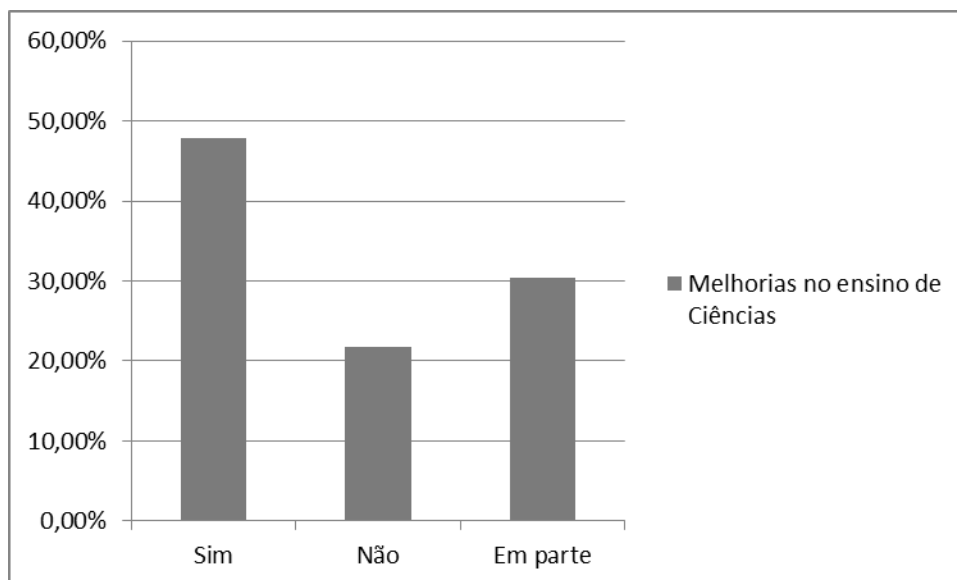
O questionário aplicado aos 23 professores que participavam de oficina de elaboração de itens trouxe as seguintes perguntas relacionadas aos usos dos resultados de ciências: Os resultados do Avalia BH em relação às competências científicas dos estudantes influenciam a sua prática pedagógica? Você vê melhorias no processo de ensino e aprendizagem em Ciências a partir da implementação da avaliação em larga dessa competência? Na escola em que leciona, existem projetos de intervenção para atender estudantes com baixo desempenho em Ciências no Avalia BH?

As respostas para a primeira pergunta foram, em sua maioria, positivas: 15 professores disseram que os resultados da avaliação em larga escala influenciam

totalmente a sua prática pedagógica nas aulas de ciências, enquanto 7 afirmaram que a sua prática é influenciada em parte e apenas 1 docente afirmou não ser influenciado em seu trabalho pelos resultados de ciências.

As respostas à segunda pergunta estão reproduzidas no gráfico 7 abaixo.

Gráfico 8 - Percentual de respostas dos professores em relação a sua percepção das melhorias no processo de ensino aprendizagem de ciências a partir da inclusão dessa disciplina nos testes do Avalia BH



Embora a maioria dos docentes da amostra pesquisada, aproximadamente 78%, tenha percebido mudanças positivas no ensino de Ciências, é significativo o percentual de docentes que não perceberam melhorias, pouco mais de 20% dos professores.

As respostas dos professores à terceira indagação confirma essa conclusão: 22 dos 23 docentes disseram não existir nas escolas em que lecionam projetos ou ações de intervenção voltados para alunos de baixo desempenho em ciências, um deles chegou a enfatizar:

Na RMEBH [Rede Municipal de Educação de Belo Horizonte], os projetos de intervenção são para Língua Portuguesa e Matemática.” O único professor que respondeu positivamente a essa questão, informou: “não é um projeto em especial, mas nos encontros com a coordenação do Ciclo discutimos e modificamos parte do planejamento de ciências.

Na entrevista realizada com membros da GAPED, um dos membros, quando perguntado sobre os projetos de intervenção implementados na rede municipal de educação como alternativas para a melhoria dos resultados do Avalia BH em relação

ao desempenho em ciências, relata: “Nós tivemos a surpresa de que em ciências os meninos já são bons. Esperávamos resultados mais baixos”. E complementa que o foco da rede é Língua Portuguesa e Matemática:

[...] o foco da nossa rede, primeiramente, é português e matemática, mais português, ainda, do que matemática. [o foco]... é os meninos lerem e entenderem o que está lendo, [...] depois vem matemática e quando a gente vencer mesmo a qualidade de português e matemática, aí sim, vamos chegando com as outras disciplinas.

Ele afirma, porém, que algumas mudanças pequenas foram observadas na gestão escolar pós-inclusão de testes de ciências no Avalia BH: “percebemos a ativação dos laboratórios de ciências nas escolas, [...] não foram mudanças extraordinárias, mudanças muito grandes, até porque é muito recente.”.

Em relação à utilização dos dados produzidos pelos testes de ciências, conclui-se, portanto, que, embora existam algumas iniciativas, como a dos professores que os levam em conta na sua prática pedagógica ou como a dos gestores que perceberam a importância dos laboratórios para o aprendizado dos alunos, as melhorias no ensino de competências científicas ainda são bastante tímidas. Esse fato vem a confirmar o que Castro (2009) ressalta:

Mas, se é verdade que o Brasil avançou na montagem e consolidação dos sistemas de avaliação, é também verdade que ainda não aprendemos a usar, de modo eficiente, os resultados das avaliações para melhorar a escola, a sala de aula, a formação de professores. Este, aliás, é um dos grandes desafios das políticas educacionais, sem o qual o objetivo principal da política de avaliação perde sentido para os principais protagonistas da educação: alunos e professores (p. 167).

Silva (2010) corrobora com a constatação nesse trabalho afirmando que ainda falta um longo caminho para que os resultados do Avalia BH, mesmo os dos testes de Língua Portuguesa e Matemática que são inclusos há mais tempo no programa, sejam apropriados pelos gestores e docentes como um instrumento de gestão e reflexão pedagógica e sugere que esse problema deva ser sanado o mais urgente possível pelos implementadores da política de avaliação educacional externa.

É importante destacar, no entanto, um projeto que não foi desenhado a partir do Avalia BH, mas que, com ele, tem mantido estreita relação, atualmente, e produzido alguns avanços: o Projeto de Intervenção Pedagógica Municipal (PIP) que consiste no reforço escolar nas disciplinas de Língua Portuguesa e Matemática. Os

alunos que participam do reforço no período do contraturno são, em sua maioria, selecionados a partir de suas notas no Avalia BH.

A Revista do Sistema de 2012 exemplifica um dos avanços decorrentes do PIP:

A partir dos dados produzidos pelo Avalia BH foi possível refletir sobre estratégias pedagógicas e ações educacionais na direção de elevar o progresso nessa etapa [6º ano do Ensino Fundamental] e ampliar as crianças nos níveis mais elevados. Observou-se que apenas em dois anos o percentual de alunos nos níveis satisfatório/avançado elevou-se para 73,4%, uma incorporação de mais de 18% dos alunos que antes estavam nos níveis mais baixos. Enquanto isso, o percentual dos alunos no abaixo do básico caiu de 10,6% para apenas 4,4%, um progresso significativo (p.8).

Nesse sentido, a extensão desse projeto para a disciplina de Ciências poderia, também, surtir efeitos positivos no seu processo de ensino e aprendizagem, principalmente para os anos finais do Ensino Fundamental em que, como observado na seção 2.2.2 desse capítulo, o percentual de alunos nos níveis mais baixos da Escala de Proficiência é maior.

Na entrevista com os membros da GAPED também foi perguntado se o Avalia BH repercutiu em algum programa de formação continuada para os professores da rede municipal a partir dos resultados dos testes de Ciências. A Gerência considera que os programas de formação docente existentes não têm relação com o programa e não estão voltados para a avaliação. Inclusive, dessa percepção, surgiu a proposta de Silva (2010,) descrita no seu Plano de Ação Educacional, de elaborar um ambiente de formação *online* a ser gestada pela GAPED para diretores, coordenadores e professores sobre o Avalia BH:

A função dessa plataforma de EaD será possibilitar aos diretores, coordenadores pedagógicos e professores da RME ampliar seu conhecimento sobre a avaliação escolar, compreender as funções das avaliações sistêmicas na atual dinâmica da educação brasileira e entender as funções e objetivos do Avalia-BH para a política municipal de educação de Belo Horizonte (p. 100).

Dessa forma, embora muito importante, o curso de formação proposto não parece abranger uma reciclagem dos professores a partir dos resultados dos testes em disciplinas específicas, como ciências. Não há a previsão de um trabalho coletivo com profissionais da educação que envolva discussão sobre as matrizes curriculares de ciências, análise dos descritores que apresentam menores índices de acerto, proposições de estratégias para reverter os quadros de baixo desempenho.

2.4 Outros efeitos da aferição de Ciências no Avalia BH

A avaliação educacional em larga escala causa impactos por meio de seus resultados, mas também por meio de seus elementos avaliativos, aqui entendidos como os instrumentos e/ou procedimentos adotados no desenho do programa que não apenas os testes. Nesse sentido, o objetivo dessa seção é apresentar as percepções dos docentes a respeito da Matriz de Referência de Ciências, bem como investigar a sua aplicabilidade na sala de aula e conhecer o que mudou na gestão do Avalia BH com a implementação da avaliação de ciências.

2.4.1 A Matriz de Referência de Ciências e a sua utilização como ferramenta pedagógica: um efeito perverso do Avalia BH?

Como afirmou membro da GAPED, o Avalia BH foi concebido e implementado com o fim de que o sistema educacional de Belo Horizonte pudesse dispor de uma avaliação melhor adequada à realidade local, mais condizente com o currículo que se trabalha nas escolas municipais: “a avaliação tem que estar muito casada com o que se trabalha em sala de aula, se não, o professor não enxerga a avaliação como um retrato”.

Dessa forma, procurou-se saber, então, se os testes do Avalia BH têm retratado o currículo de Ciências do Ensino Fundamental segundo a ótica dos docentes, bem como investigar se e de que forma as Matrizes de Referência são utilizadas por esses profissionais na sala de aula.

A GAPED relatou que, logo após o primeiro ciclo de avaliação de Ciências, muitos professores, gestores e coordenadores regionais de ensino ligaram para a GAPED a fim de comentar a prova de Ciências. De acordo com essa gerência,

muitos reclamavam que seus alunos acharam a prova difícil demais, que estava fora das proposições curriculares, [enquanto] outros já falaram que as proposições curriculares estavam falhas, que enxergaram, através da prova, algumas falhas nas proposições que eles vão ter que corrigir.

Nesse sentido, pode-se dizer que um primeiro efeito da avaliação de Ciências foi provocar a reflexão nos docentes sobre o que ensinar, uma questão tradicionalmente conflituosa em qualquer campo do conhecimento e que merece uma atenção maior quando se estabelece uma matriz avaliativa.

Os 23 professores que participavam de oficina de elaboração de itens foram questionados se as matrizes de referência para Ciências do Avalia BH estão adequadas ao que eles consideram importante que o estudante saiba ao final de cada etapa da escolarização. Dez desses professores disseram que estão adequadas, 12 responderam que estão adequadas apenas parcialmente e 1 afirmou que não está adequada. Dessa forma, mais da metade dos docentes pondera que as matrizes necessitam de readequações.

Segundo a GAPED, essa demanda tem sido atendida. A cada ciclo avaliativo, no momento das oficinas de elaboração de itens, as matrizes de ciências são rediscutidas com os professores e alguns ajustes são realizados a fim de que se chegue a um consenso sobre o que deve ser avaliado e o que não deve ser cobrado nos testes.

Nessas ocasiões, são analisados os comportamentos dos descritores nos testes de acordo com o percentual de acerto. Na última oficina, por exemplo, realizada em 2013, analisou-se o fato de um descritor da matriz do terceiro ano do ensino fundamental apresentar um índice de acerto desprezível. O descritor exigia do estudante a habilidade de reconhecer as consequências dos movimentos de rotação e translação do planeta Terra e os docentes chegaram à conclusão de que, na prática, essa habilidade não é trabalhada no terceiro ano, mesmo que ela esteja prescrita nas proposições curriculares para essa etapa da escolarização. Dessa forma, o descritor foi retirado da matriz de referência. Em outros casos, descritores foram acrescentados ou sofreram modificações em sua redação. Importante ressaltar que esse trabalho preza pela manutenção da comparabilidade entre ciclos avaliativos e mantém pelo menos 50% da matriz intacta.

Quando perguntados se utilizam as matrizes de referência de ciências na prática em sala de aula, 22 professores dos 23 que participaram da pesquisa responderam afirmativamente. Quando perguntados de que forma essas matrizes são trabalhadas, quase a sua totalidade as relacionam ao planejamento da disciplina na sala de aula, o que pode significar que, na ânsia por obter bons resultados de seus alunos, docentes estão aproveitando a matriz de referência para os testes como matriz de referência para o ensino.

Conquanto a Matriz de Referência seja um material rico em possibilidades para o trabalho docente, a sua utilização no lugar do currículo gera um efeito perverso da avaliação em larga escala, já sinalizado por alguns autores (Casassus,

2009; Dameto, 2012, Alavarse, 2013; entre outros) que é o treinamento dos estudantes para responder os itens dos testes padronizados, o que resulta em perda de qualidade do ensino. Ocimar Alavarse (2013) afirma que é possível considerar a hipótese de que algumas escolas tendam tanto a concentrar o ensino no que constitui o objeto de avaliação, relegando o currículo oficial a segundo plano, quanto a utilizar os moldes da prova, com itens de múltipla escolha, no lugar das tradicionais avaliações internas e pontua:

se as restrições às avaliações internas podem ser superadas com formação e capacitação, por seu turno, as avaliações externas não podem e não deveriam se converter em orientadoras privilegiadas dos processos formativos, mesmo quando parecem se constituir no mais saliente traço das políticas educacionais. Distinto disso seria a tarefa de articular, por vários meios e objetivos, as avaliações externas com as internas.

Essa articulação se daria, por exemplo, por meio da inserção dos docentes nos processos avaliativos, isto é, da oportunização desses profissionais em debates sobre o que deve significar a matriz de referência na sala de aula e como esse documento poder servir de ferramenta pedagógica de forma a que não se sobreponha como diretriz em relação ao currículo, tornando o ensino reducionista.

2.4.2 O efeito sobre a gestão do Avalia BH com a inclusão de Ciências

Como órgão responsável pela gestão do Avalia BH, a GAPED relata o que mudou com a implementação de testes de Ciências no Avalia BH em 2010: “A logística aumentou demais! São três dias de aplicação de provas, temos que fazer oficinas de divulgação de resultados para maior quantidade de professores e temos que criar o banco de itens de Ciências”. Outro membro ratifica: “A parte da logística é muito afetada, o banco de dados fica mais complexo, são mais informações para correlacionar”.

A GAPED também destaca um conflito entre essa gerência e as escolas fortalecido por essa implementação:

o processo de avaliação altera muito a rotina da escola, e as escolas ressentem disso. Nem sempre conseguimos cumprir o nosso cronograma cem por cento à risca, às vezes tem um atraso. Aí muda a data da avaliação, antecipa uma semana ou posterga uma semana, às vezes choca com alguma coisa que a escola já tinha programado, como um passeio, uma viagem, uma festa... então a

gestão chega e obriga a escola a mudar toda a rotina para aplicar uma avaliação...existe uma tensão nesse processo.

2.4.3 Conclusões da pesquisa

Diante dos dados trazidos pela entrevista à GAPED, pelos questionários aplicados aos docentes e gestores e pela análise documental, conclui-se que o estudo sobre a inclusão de ciências da natureza no Avalia BH ainda apresenta muitos pontos cegos. Essa fato pode ser atribuído à incipiência de todo o processo. Não houve tempo hábil de avaliação para que os seus efeitos pudessem ser observados e explorados de forma mais aprofundada.

Ainda assim, alguns aspectos puderam ser percebidos e merecem relevo. Dentre esses aspectos, destaca-se que os testes de Ciências no Avalia BH realçaram a discussão sobre a importância de se avaliar competências e habilidades desenvolvidas nessa disciplina, além de ter promovido a reflexão dos docentes a respeito do currículo. No campo das ações, porém, constatou-se a fragilidade do programa no momento de traduzir os seus resultados em melhorias no ensino ou de, no limite, criar estratégias para o alcance desse objetivo mesmo que em longo prazo.

O salto que o programa de monitoramento da educação municipal de Belo Horizonte deu ao incorporar as avaliações de ciências parece ter sido meramente quantitativo. Ainda existe uma preterição das ciências em relação às disciplinas de Língua Portuguesa e Matemática, como se a atenção suficiente só pudesse ser concedida a esse campo do conhecimento quando todos os problemas concernentes aos primeiros estiverem resolvidos.

Outra impressão causada pela entrevista e pelos questionários é a de que a divulgação dos dados de ciências tem ocorrido de forma precária e isso pode estar intimamente relacionado ao fato de os resultados de desempenho serem subutilizados. Evidenciou-se, ainda, que os testes de ciências precisam superar o patamar da experimentação: não há certezas sobre parâmetros e instrumentos de avaliação.

Nesse sentido, arremata-se que a inclusão de ciências no AVALIA BH tem gerado um ônus que ainda não foi recompensado com efetivas ações de melhorias e

esse fato aponta para a necessidade de se repensar o programa. Não se defende, aqui, a abolição do mesmo, afinal, como afirma Simon Schwartzman (2005),

os problemas associados às avaliações parecem ser claramente preferíveis aos problemas associados à sua não existência, como a falta de parâmetros, a ausência de critérios, os desperdícios e injustiças na distribuição de recursos e a impossibilidade de estabelecer políticas consistentes para sistemas de educação de massas (p. 31).

O que se sugere é a ampliação da análise aqui realizada a fim de identificar se os objetivos não alcançados podem, em um futuro próximo e de forma eficaz, serem atingidos; se o processo de inclusão de ciências, seus métodos e instrumentos revelaram seus méritos ou necessitam passar por reformulações; se os efeitos sobre os beneficiários do programa são relevantes a ponto de justificar todos os ônus, inclusive os financeiros; se os vieses podem ser contornados. Enfim, propõe-se um estudo que traga juízo de valor sobre o Avalia BH, especificamente sobre a inclusão de ciências no programa, por meio da meta-avaliação, tema do capítulo que encerra o trabalho.

3. A META-AVALIAÇÃO COMO PLANO DE AÇÃO EDUCACIONAL PARA O AVALIA BH/CIÊNCIAS DA NATUREZA

Este último capítulo visa apresentar uma proposta que reveja os problemas apresentados pelo Avalia BH na implementação da avaliação do desempenho estudantil em ciências da natureza e, assim, contribuir para a sua solução. Observou-se, no capítulo anterior, que essa política não tem surtido o efeito que se espera ou, antes, tem gerado efeitos diversos dos objetivos da avaliação em larga escala, como a utilização da Matriz de Referência como baliza para o ensino na sala de aula.

Supõe-se que o estabelecimento de uma avaliação do Avalia BH concorre para identificar os seus principais desafios, promovendo a reflexão sobre o que está dando errado e clareando os caminhos para reverter essa situação. Uma avaliação da avaliação responderá se a decisão de incluir testes de Ciências no programa seguiu os critérios de adequação, eficácia, eficiência e relevância, isto é, se essa decisão pode ser traduzida em efeitos positivos aos beneficiários (gestores, professores, alunos) e apontar as inconsistências no delineamento do desenho do sistema avaliativo para que possam ser refinados a tempo de fazer valer os recursos gastos.

Ressalta-se que a presente proposta encontrou inspiração na tese de doutorado da pesquisadora Lys Dantas, que, em 2009, investigou as contribuições das políticas de avaliação educacional no estado da Bahia utilizando o modelo de meta-avaliação, e baseou-se em artigos publicados por Daniel Stufflebeam, pesquisador membro da *Joint Committee on Standards for Educational Evaluation – JCSEE*¹⁵, que apresenta um modelo de meta-avaliação e a defende como obrigação de todo avaliador.

3.1 O que é meta-avaliação?

O termo meta-avaliação foi utilizado pela primeira vez por Michael Scriven, em 1969, para designar a avaliação de um plano de avaliação educacional (STUFFLEBEAM, 2001). Scriven (*apud* Stufflebeam, 2001) ressaltou que a emissão de relatórios imprecisos ou tendenciosos poderia induzir o consumidor a comprar um

¹⁵ Criada em 1975, o *Joint Committee on Standards for Educational Evaluation* é uma organização sem fins lucrativos, constituída exclusivamente para o desenvolvimento de padrões de avaliação e promoção do seu uso.

produto educacional inadequado e usá-lo em detrimento de crianças e jovens. Assim, defendia que tais relatórios deveriam ser, eles próprios, avaliados para o bem-estar dos que iam utilizá-lo.

A meta-avaliação é uma reflexão sobre os procedimentos utilizados na avaliação, ou, como Stufflebeam (2001), mais detalhadamente, conceitua

é o processo de obter, delinear e aplicar informações descritivas e de julgamento sobre a utilidade, a viabilidade, a adequação e a precisão de uma avaliação e sua natureza sistemática, competência, integridade/honestidade, respeitabilidade e responsabilidade social para orientar a avaliação e divulgar publicamente seus pontos fortes e fracos (p. 108).

A divulgação dos pontos fortes contribui para reforçar as medidas adotadas, enquanto a dos pontos fracos colabora para a retificação do processo, quando se trata de uma meta-avaliação formativa, que aqui se julga como a mais adequada para o caso do Avalia BH. Em tempo, meta-avaliação formativa é o modelo empregado durante o desenvolvimento da avaliação, auxiliando os avaliadores no planejamento, na interpretação e na melhoria do sistema de avaliação, divergente da meta-avaliação somativa, que é realizada ao término da avaliação, focalizando o relatório produzido e tomando nota dos acertos e erros, sem perspectivas de retomá-los (Stufflebeam, 2001).

Segundo Lygia Elliot (2011), a meta-avaliação apresenta uma trajetória considerável no exterior, enquanto que no Brasil, os esforços nesse sentido são modestos. De fato, por meio de uma revisão de literatura se constata que são poucos os estudos divulgados que empregaram essa metodologia, sobretudo na área de educação. Lys Dantas (2009) já afirma que graças à expansão da avaliação educacional em larga escala no país, esse quadro tem mudado e chama a atenção para a necessidade de que isso efetivamente ocorra, pois para a autora, meta-avaliar é uma questão ética: “se algo está errado em uma avaliação, o erro não deve impactar, injustamente, o objeto da avaliação” (p. 33).

Estaria o Avalia BH impactando, de forma injusta, o ensino de ciências na rede municipal de Belo Horizonte? A presente pesquisa aponta para uma indefinição de parâmetros, a utilização inadequada das matrizes de referência, a divulgação ineficaz dos resultados, além de mostrar que, em três ciclos de avaliação, muito pouca coisa tem sido feita a partir dos resultados de proficiência em ciências da natureza. O delineamento de uma meta-avaliação para o Avalia BH Ciências, nesse

sentido, objetiva identificar os vieses, erros técnicos e má utilização, contribuindo para a melhoria dos procedimentos avaliativos e consolidação da avaliação externa como meio para a melhoria da qualidade do ensino.

3.2 Modelo de meta-avaliação para o Avalia BH

Diferentes autores propuseram, ao longo do tempo, diferentes delineamentos de meta-avaliação. De acordo com Dantas (2009), não existe um formato padrão, apenas o mais adequado para meta-avaliar determinado programa. Nesse contexto, o presente trabalho propõe para a meta-avaliação do Avalia BH um modelo que resulta de uma compilação entre as abordagens de Daniel Stufflebeam (2001) e Michael Scriven (2007).

3.2.1 A escolha do meta-avaliador e a formalização da meta-avaliação

Daniel Stufflebeam (2001) recomenda que a equipe selecionada para realizar a meta-avaliação deva ser externa, especializada e experiente para que possa fornecer informações independentes e com credibilidade. No entanto, o mesmo autor admite que nem sempre é possível ou necessário que se contrate uma equipe de meta-avaliadores e afirma:

em algumas avaliações que contam com poucos recursos e, sobretudo, nas avaliações formativas, os avaliadores podem adequadamente realizar a meta-avaliação formativa ou parte dela. Tal prática de se auto meta-avaliar é melhor do que a não realização da meta-avaliação (p. 189).

Em artigo publicado em 1974, ele havia chegado à conclusão que para meta-avaliações somativas, o ideal era que o meta-avaliador fosse externo enquanto que para meta-avaliações formativas, o meta-avaliador deveria ser interno, opinião compartilhada por Ignacio Cano (2009) em seu livro sobre avaliações de programas sociais.

Sendo assim, neste trabalho, a proposta de meta-avaliação da avaliação de ciências do Avalia BH se dirige ao órgão gestor do programa, isto é, à GAPED. Investimentos financeiros já são realizados pela prefeitura de Belo Horizonte para o pagamento dos serviços prestados pela agência avaliadora CAEd. Contratar outra agência avaliadora ou a mesma para os procedimentos de meta-avaliação implica no aumento de gastos, o que se julga impróprio para o momento. Além do mais,

acredita-se ser importante que os próprios membros da GAPED pratiquem a reflexão sobre o desenho da política que desenvolveram e que, conhecendo mais profundamente o programa, forneçam maior colaboração com sugestões para a sua melhoria e enfrentem menos resistência quanto à obtenção de informações.

Definida a equipe que realizará a meta-avaliação, a segunda tarefa seria o registro formal de sua data de início, seus objetivos, padrões que irá utilizar e outros esclarecimentos que forem necessários, além de firmar o compromisso de transparência e de divulgação do relatório final do processo. É de suma importância, também, que neste momento exista uma interação com os denominados, por Stufflebeam (2001), como os *stakeholders*, isto é, o público interessado na avaliação, no caso, os gestores educacionais, os docentes, os alunos, a família dos alunos, enfim, todas as pessoas que são afetadas direta ou indiretamente pelos testes de ciências do Avalia BH. Os padrões de qualidade da avaliação devem ser negociados com esse público, a fim de que a demanda deles seja atendida e a meta-avaliação ganhe credibilidade.

Dantas (2009) afirma, no entanto, que definir os padrões de qualidade para uma meta-avaliação não é uma tarefa simples. Para a autora, com a qual se concorda, é fácil determinar que a avaliação deva produzir dados precisos, mas é difícil definir esse grau de precisão. A boa notícia é que na literatura especializada encontram-se diferentes listas de critérios a serem considerados em uma meta-avaliação, de acordo com a sua natureza e especificidades. Em sua maioria, baseiam-se em categorias estabelecidas pela *Joint Committee on Standards for Educational Evaluation – JCSEE*, que serão adotadas, também, neste plano de ação educacional.

Antes disso, salienta-se que é necessário organizar os dados de forma sistemática para que possam ser melhor visualizados e posteriormente julgados. Michael Scriven (2007) projetou uma Chave de Avaliação para o “uso na avaliação de programas, planos, políticas e relatórios de avaliação” (p. 1). O autor defende a adoção dessa chave como bastante útil para o começo da reflexão sobre a avaliação, para estabelecer a terminologia técnica e iniciar um diário de bordo do projeto. Embora esse documento seja mais voltado para meta-avaliações somativas, é possível adaptá-lo para a meta-avaliação formativa que aqui se propõe.

O quadro a seguir resume a Chave de Avaliação de Scriven (2007).

Quadro 12 - Chave de Avaliação para realização de meta-avaliações

Preliminares	Estrutura	Subavaliações	Conclusões e implicações
Sumário executivo	Antecedentes e contexto	Processo	Síntese
Prefácio	Descrições e definições	Efeitos	Recomendações e explicações
Metodologia	Consumidores	Custos	Responsabilização
	Recursos	Comparações	
	Valores	Generalizações	

Fonte: Scriven (2007)

Na seção de Preliminares, o sumário executivo é o campo para ser preenchido com os resultados da avaliação de forma resumida, o prefácio deve informar quem é o público interessado na avaliação e a metodologia descreve o delineamento e todos os procedimentos técnicos previstos para a avaliação (SCRIVEN, 2007).

Na seção de Estrutura, preconiza-se a identificação do contexto em que se insere a avaliação; a descrição de fatores que podem ser relevantes para o seu sucesso ou fracasso; a revisão de literatura sobre avaliações semelhantes e seus resultados; a descrição do programa de acordo com diferentes fontes (oficial, usuários); a definição de termos técnicos a serem utilizados; a relação de grupos de pessoas que são afetadas pelos resultados das avaliações, a identificação dos recursos utilizados e dos que estavam disponíveis, tanto em termos financeiros como em termos de capital social (SCRIVEN, 2007).

A seção denominada Subavaliações se dedica a relatar tudo o que aconteceu antes dos resultados, no que diz respeito às operações do programa, ao cumprimento dos objetivos, os custos e a descrever todos os efeitos causados a diferentes interessados no programa (SCRIVEN, 2007).

A última seção, Conclusões e Implicações, destaca os impactos presentes e futuros da avaliação, as restrições da avaliação, a comparação entre o que se esperava da avaliação e o seu resultado, as limitações do programa e as recomendações e sugestões para a sua melhoria (SCRIVEN, 2007).

Com base nessa Chave de Scriven (2007), propõe-se uma chave de meta-avaliação para as avaliações de Ciências do Avalia BH, conforme mostra o quadro 13.

Quadro 13 - Proposta de Chave de Meta-avaliação para as avaliações de Ciências do Avalia BH

Pontos de Avaliação		Preenchimento da Chave
PRELIMINARES	Sumário executivo	Resumo dos resultados dos testes de proficiência em ciências do Avalia BH em todos os ciclos de avaliação
	Prefácio	Identificação do público alvo da política de avaliação: alunos matriculados no ensino fundamental da rede municipal de Belo Horizonte
	Metodologia	Descrição da metodologia para a construção das Matrizes de Referência e Escalas de Proficiência de Ciências, definição dos Padrões de Desempenho, elaboração dos itens, tratamento e leitura de dados, divulgação de resultados
ESTRUTURA	Antecedentes e contexto	Relato dos antecedentes da implementação de Ciências no Avalia BH
	Descrições e definições	Descrição dos objetivos e justificativas da inclusão de ciências no Avalia BH, revisão de literatura sobre avaliações de ciências em larga escala
	Consumidores	Identificação das pessoas interessadas nos resultados das avaliações de ciências: Secretaria Municipal de Educação, Gestores de escolas, docentes, discentes, família dos discentes
SUBAVALIAÇÕES	Processo	Descrição de como se tem dado o processo: acertos, erros e desafios na construção de Matrizes de Referências de Ciências; acertos, erros e desafios na construção e interpretação da Escala de proficiência de Ciências; acertos, erros e desafios na interpretação e divulgação de resultados.
	Efeitos	Descrição dos efeitos para a gestão educacional em geral e gestão escolar a partir dos resultados dos testes de desempenho em ciências, bem como para os docentes, alunos e família dos alunos; descrição dos vieses da avaliação; efeitos esperados, efeitos indesejados; utilidade da avaliação para todos os interessados.
	Custos	Descrição dos custos financeiros e ônus que couberam à GAPED com a logística, às escolas com a aplicação de mais uma avaliação, entre outros.
CONCLUSÃO	Síntese	Síntese de todos os pontos anteriores, com destaque para os efeitos e utilidade da avaliação de ciências; conclusão sobre a meta-avaliação
	Recomendações e explicações	Explicações dos pontos fracos do programa e sugestões para a sua melhoria
	Responsabilização	Divulgação dos resultados da meta-avaliação, prestação de contas aos interessados.

Finalizada a etapa de construção da Chave de avaliação, é necessário estabelecer os padrões de qualidade que servirão para o seu julgamento, tema de que trata a seção subsequente.

3.2.2 A definição dos padrões de qualidade

São cinco¹⁶ as publicações da JCSEE resultantes de esforços de diferentes pesquisadores e *stakeholders* no sentido de criar um *checklist* de padrões de qualidade para a realização de meta-avaliações. As primeiras, que datam de 1981, 1994, 2004 e 2008 estabelecem a utilidade, a viabilidade, a adequação e a precisão como principais critérios para a condução de um julgamento, enquanto o relatório de 2011 acrescenta o critério de *accountability* (ELLIOT, 2011).

A utilidade, explica Elliot (2011), refere-se à preocupação direta com as necessidades dos usuários da avaliação. De acordo com esse critério, a avaliação deve dar atenção a todos os envolvidos no processo, negociar os objetivos, indicar os valores que embasam esses objetivos, descrever de forma clara o que foi realizado, considerar as consequências que os resultados podem gerar (ELLIOT, 2011). Seguramente, é o padrão de qualidade mais importante para se verificar em uma avaliação, já que esta não faz sentido se não servir aos interessados, mesmo que atenda aos outros padrões com sucesso.

A viabilidade inclui a competência da equipe gestora da avaliação em assegurar-lhe a praticidade e a eficiência. Certos autores traduzem esse padrão como exequibilidade, como Elliot (2011) que esclarece:

Os padrões de exequibilidade são voltados para procedimentos que assegurem que a avaliação será gerenciada de modo efetivo, será prática e realística, será cuidadosa ao tratar do contexto e seus interesses políticos e culturais, será tanto efetiva quanto eficiente ao utilizar os recursos (p. 951-952).

A adequação está relacionada ao bem estar das pessoas envolvidas na avaliação ou que são afetadas pelos seus resultados. Trata do respeito aos direitos

¹⁶ *Standards for evaluations of educational programs, projects, and materials* (1981)
Standards for evaluations of educational programs, projects, and materials – 2^o edition (1994)
The program evaluation standards (2004)
The program evaluation standards – 2^a edition (2008)
The Program Evaluation Standards: a guide for evaluators and evaluation users (2011)
 Artigos disponíveis em: <<http://www.jstor.org/action/showPublication?journalCode=jeducsocwork>>.
 Acesso em: 6 jul. 2014.

dos participantes, de possíveis conflitos de interesse, da acessibilidade dos resultados, por exemplo (ELLIOT, 2011).

A precisão concentra-se nas questões técnicas da avaliação como a qualidade de implementação, do desenho do programa, dos instrumentos selecionados para a sua realização e do tratamento dos dados obtidos (DANTAS, 2009).

Por fim, o padrão mais recentemente incorporado pela *JCSEE*, *accountability*, é diretamente relacionado à meta-avaliação. Trata da comprovação de cada etapa realizada da meta-avaliação, como um atestado de qualidade da avaliação, uma prestação de contas aos *stakeholders* (ELLIOT, 2011).

A aplicação dos padrões supracitados varia de acordo com a natureza e o foco da avaliação ou da meta-avaliação. No presente trabalho, considerando que o padrão de viabilidade seria melhor observado por um meta-avaliador externo, ele será omitido. Sendo assim, para a meta-avaliação das avaliações de Ciências do AVALIA BH sugere-se que sejam usados os padrões de utilidade, adequação, precisão e *accountability*.

É importante ressaltar que cada padrão definido pela *JCSEE*, é na realidade, uma categoria abrangente que envolve critérios mais específicos. Assim, o padrão de utilidade, por exemplo, é subdividido em oito critérios: credibilidade do avaliadores, atenção aos *stakeholders*, propósitos negociados, explicitação de valores, informação relevante, processos e produtos significativos, comunicação e relatórios apropriados e no prazo, preocupação com consequências e influência (ELLIOT, 2011). Como nem sempre esses “sub-padrões” se ajustam à meta-avaliação que se sugere, eles não serão descritos neste trabalho.

Para apoiar o julgamento sobre aspectos da avaliação, Stufflebeam (2001) explica que, em seu modelo de meta-avaliação, costuma atribuir um valor de escala, o que significa, por exemplo, marcar como excelente, muito bom, bom ou ruim, em cada aspecto da avaliação, de acordo com cada padrão de qualidade. Assim, esse autor determina o grau em que o sistema de avaliação teve padrões satisfeitos e identifica os seus pontos fracos e fortes.

Apropria-se, então, das ideias de Stufflebeam (2001) para a construção do quadro 14, a seguir, proposto para a utilização dos padrões de qualidade pela meta-avaliação da avaliação de ciências do Avalia BH.

Quadro 14 - Julgamento de aspectos da avaliação de Ciências do Avalia BH de acordo com padrões de qualidade

Pontos de Avaliação		Utilidade	Adequação	Precisão	Accountability
PRELIMINARES	Sumário executivo				
	Prefácio				
	Metodologia				
ESTRUTURA	Antecedentes e contexto				
	Descrições e definições				
	Consumidores				
SUBAVALIAÇÕES	Processo				
	Efeitos				
	Custos				
CONCLUSÃO	Síntese				
	Recomendações e explicações				
	Responsabilização				

Para o preenchimento desse quadro, sugere-se o uso da seguinte escala: atende ao padrão de qualidade, atende apenas parcialmente, não atende ao padrão. É evidente que cada ponto de avaliação abrange uma gama de perguntas que deverão ser formuladas pelo meta-avaliador e respondidas dentro da apresentada escala. Esclarecendo, para julgar o ponto de avaliação “Processo”, por exemplo, quanto ao padrão de adequação, verificaria-se a construção das matrizes de referência de Ciências apresentou conflitos de interesse, se a linguagem utilizada nas oficinas de divulgação de resultados é acessível ao público leigo em terminologias de avaliação em larga escala, se essas oficinas têm ocorrido no momento adequado, entre outros questionamentos. Para cada resposta a esses questionamentos, seria atribuído um dos valores mencionados no início desse

parágrafo: atende o padrão de adequação, atende apenas parcialmente, não atende ao padrão de adequação.

O preenchimento da chave de avaliação contribuirá para a formulação das perguntas da meta-avaliação. Incube ao meta-avaliador buscar todas as informações possíveis para respondê-las, por meio de entrevistas, questionários e reuniões que envolvam os *stakeholders*.

3.2.3 O cronograma da meta-avaliação

Propõe-se que a meta-avaliação ocorra concomitantemente ao ciclo de avaliação. O delineamento paralelo da avaliação e da meta-avaliação facilita a obtenção de dados pela última, otimiza tempo e recursos, além de promover uma atitude reflexiva em relação ao processo de avaliação, já que se sabe avaliado.

3.2.4 Os resultados da meta-avaliação

Espera-se que na etapa final da meta-avaliação, os pontos críticos da avaliação sejam identificados. Nesse momento, é importante que os resultados da meta-avaliação sejam divulgados para os interessados e, mais do que isso, discutidos com o fim de obter sugestões e recomendações para a sua melhoria. Stufflebeam (2001) aconselha a preparação e a apresentação de relatórios interinos e finais por meio de *workshop*, publicações em revistas especializadas e em domínios da internet.

Stufflebeam (2001) também destaca que a meta-avaliação deve constituir um esforço colaborativo e essa colaboração é conquistada quando o objetivo é o de ajudar uma organização a reformar seu sistema de avaliação e proteger o público da falta de informação ou mal uso dos resultados. Ele afirma, ainda, que “deve se comunicar adequadamente ao público os relatórios de meta-avaliação com o fim de garantir a sua confiança, interesse, assistência, compreensão e uso dos resultados” (p. 202).

3.3 Considerações finais

Este trabalho constituiu um plano de ação educacional, como parte integrante do modelo de dissertação definido pelo programa de pós-graduação profissional. Constituiu, também, no limite, uma pretensiosa forma de contribuir para com o

ensino de Ciências no Brasil ao destacar a necessidade de se rever as políticas públicas educacionais voltadas para a sua melhoria. Sabe-se que as avaliações externas podem promover a transparência acerca da situação do ensino, produzindo o conhecimento sobre onde intervir para que deficiências sejam superadas. Isso, no entanto, não significa que o seu desenho dispense questionamentos. Muito pelo contrário: uma avaliação da avaliação é imprescindível para observar se as metas têm sido alcançadas, se os atores envolvidos no processo estão comprometidos com a eficiência e eficácia do sistema, se os beneficiários da política estão sendo impactados positivamente.

A análise, aqui realizada, do sistema de avaliação educacional da prefeitura de Belo Horizonte, o Avalia BH, permitiu verificar que apesar dos esforços em incluir os testes de Ciências - louváveis por se tratar de uma inauguração em nível municipal e, mais ainda, por ser uma iniciativa no sentido de se pensar o processo de ensino e aprendizagem de competências científicas, os resultados de desempenho nesse campo do conhecimento são subutilizados. A situação das Ciências Naturais ainda é relegada ao segundo plano, isto é, ainda espera pelo preenchimento das lacunas nos ensinamentos de Língua Portuguesa e Matemática para obter maior atenção.

Políticas de avaliação são caras, competindo com outras propostas educacionais por recursos. Além disso, alteram o calendário escolar, sobrecarregam a gestão, geram situações incômodas. Dessa forma, implementar uma avaliação em larga escala só faz sentido se essa for compensatória, capaz de gerar os benefícios esperados pelos seus principais interessados. Acentua-se, também, nesse contexto, a necessidade da reflexão sobre os instrumentos de medição empregados para a garantia da fidedignidade dos seus resultados.

Assim, espera-se que a meta-avaliação proposta seja um instrumento relevante para o Avalia BH no tocante aos testes padronizados de Ciências, atestando, quando aprovar, a qualidade do programa e de suas metodologias e evidenciando os aspectos que precisam ser reformulados. Além disso, a expectativa deste trabalho é a de que sua temática inspire novos estudos, afinal, conquanto a literatura sobre avaliação sistêmica no Brasil seja já bastante densa, o mesmo não ocorre quando se trata de meta-avaliações, a despeito de sua importância.

É forçoso, porém, reconhecer as limitações desta pesquisa, algumas próprias das abordagens qualitativas, como a dificuldade de colher do entrevistado uma

opinião completamente desvinculada do viés político e a inevitável subjetividade quando da interpretação dos dados obtidos. Admite-se, ainda, que a percepção dos docentes, talvez os principais *stakeholders*, poderia ser melhor observada caso a amostragem não tivesse se restringido aos que compareceram às oficinas e que a incipiência da inclusão das ciências nos sistemas avaliativos nacionais, de forma geral, responde pelas incertezas relacionadas a esse processo. Sobre esses dois últimos pontos, conforta o fato de que ambos, o tamanho da amostra e a incipiência da aferição da proficiência em ciências, acabam por reforçar ainda mais a demanda pela meta-avaliação.

Não se pode deixar de assentir que o plano de ação educacional apresentado é substancialmente menos arrojado do que os modelos de meta-avaliação que lhe serviram de molde. O objetivo foi o de evitar entraves diante da complexidade de processos e torná-lo o mais exequível e aberto à participação possível. Por fim, ambiciona-se que o desfecho deste trabalho concorra para a sua efetivação pela Secretaria Municipal de Educação de Belo Horizonte e que a sua contribuição para a educação seja comparável ao crescimento pessoal, acadêmico e profissional que proporcionou à sua autora.

REFERÊNCIAS

ABREU, Lenir Silva. **O desafio de formar professores dos anos iniciais do ensino fundamental para ensinar ciências**. 2008. Dissertação (Mestrado em Ensino, Filosofia e História das Ciências) – Universidade Federal da Bahia/Universidade Estadual de Feira de Santana, Salvador, Bahia, 2008. Disponível em: <<https://twiki.ufba.br/twiki/pub/PPGEFHC/DissertacoesPpgefhc/Lenir2008.pdf>>. Acesso em: 25 maio 2013.

ALAVARSE, Ocimar Munhoz. Avaliar as avaliações em larga escala: desafios políticos. **Educação**, São Paulo, 2013.

AMAZONAS. Secretaria de Estado da Educação e Qualidade do Ensino. **Revista do Sistema SADEAM – 2011**. Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd, Juiz de Fora, MG, v.1, jan./dez. 2011.

_____. Secretaria de Estado da Educação e Qualidade do Ensino. **Revista Pedagógica – Ciências da Natureza – Ensino Médio Regular e EJA SADEAM – 2012**. Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd, Juiz de Fora, JF, v. 1, jan./dez. 2012.

BAHIA. Secretaria da Educação do Estado. **SABE (Sistema de Avaliação Baiano de Educação)**. AVALIE ALFA – 2011/ Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd, Juiz de Fora, MG, v. 1, jan./dez. 2011.

BELO HORIZONTE. Prefeitura Municipal de Belo Horizonte. **Avalia BH Rede Municipal de Educação**. Belo Horizonte, MG: FGV, 2008.

_____. Secretaria Municipal de Educação de Belo Horizonte. **Avalia BH 2012 Revista Pedagógica de Ciências da Natureza - 1º ciclo (3º ano do Ensino Fundamental)**. Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd, Juiz de Fora, MG, v. 1, jan./dez. 2012.

_____. Secretaria Municipal de Educação de Belo Horizonte. **Boletim do Sistema de Avaliação Avalia BH**. Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd, Juiz de Fora, MG, v. 1, jan./dez. 2009.

_____. Secretaria Municipal de Educação de Belo Horizonte. **II Congresso Político-Pedagógico da Rede Municipal de Ensino/Escola Plural**. Belo Horizonte, MG, 2002.

_____. Secretaria Municipal de Educação de Belo Horizonte. Desafios da Formação. **Proposições curriculares Ensino Fundamental Ciências**. Belo Horizonte, MG, 2010.

BIZZO, N. Ciências biológicas. In: BRASIL. Ministério da Educação. **Orientações curriculares nacionais do ensino médio**. Brasília, DF: MEC, 2004. Disponível em: <<http://www.mec.gov.br>>. Acesso em: 22 maio 2013.

BONAMINO, Alícia & FRANCO, Creso. Avaliação e política educacional: o processo de institucionalização do SAEB. **Cadernos de Pesquisa**, São Paulo, SP, n.108, p. 101-132, nov.1999.

BRASIL. Lei nº 4.024 de 1961. Estabelece as diretrizes e bases da educação nacional. **Diário Oficial da União**, Brasília, DF, 23 dez. 1996.

_____. Lei nº 5692 de 1971. Fixa diretrizes e bases para o ensino de 1º e 2º graus, e dá outras providências. **Diário Oficial da União**, Brasília, DF, 12 ago. 1971. Seção 1, p. 6377.

_____. Ministério da Educação. Portaria nº 304, de 21 de junho de 2013. **Diário Oficial da União**, Brasília, DF, 24 jun. 2013. Seção 1, p. 33.

_____. Ministério da Educação. Secretaria de Educação Fundamental. **Parâmetros Curriculares Nacionais: Introdução aos Parâmetros Curriculares Nacionais**, Brasília, DF, 1996.

_____. Ministério da Educação. Secretaria de Educação Fundamental. **Parâmetros Curriculares Nacionais: Ciências**, Brasília, DF, 1997.

_____. Parecer CNE/CEB nº 15/98 referente às Diretrizes curriculares nacionais para o Ensino Médio. **Diário Oficial da União**, Brasília, DF, 26 jun. 1998.

BRASÍLIA. Ministério da Educação. **Inclusão de Ciências no SAEB**. 1.ed. Brasília: MEC, 2013.

_____. Ministério da Educação. **Matrizes curriculares de referência para o SAEB**. Brasília: INEP, 1997.

_____. Ministério da Educação. **PISA 2000: Relatório Nacional**. Brasília: INEP, 2001.

_____. Ministério da Educação. **Prêmio Inovação em Gestão Educacional 2006**. Brasília: INEP, 2007.

_____. Ministério da Educação. **Prêmio Inovação em Gestão Educacional 2011**. Brasília: INEP, 2012.

_____. Ministério da Educação. **Saeb SAEB 2001: Novas Perspectivas**. 1. ed. Brasília: MEC, 2002.

BROOKE, 2012 (Org.). **Marcos históricos na reforma da educação**. 1.ed. Belo Horizonte: Fino Traço, 2012.

CACHAPUZ, Antònio *et al.* **A necessária renovação do ensino de ciências**. São Paulo: Cortez, 2005.

CANO, Ignácio. **Introdução à Avaliação de Programas Sociais**. 3. ed. São Paulo, SP: FGV, 2009.

CASASSUS, Juan. Uma nota crítica sobre a avaliação estandarizada: a perda da qualidade e a segmentação social. **Sísifo: Revista de Ciências da Educação**, p. 71-79, 2009. Disponível em: <www.sisifo.fpce.ul.pt>. Acesso em: 12 maio 2014.

CASTRO, Cláudio de Moura. Avaliar não é para amadores. In: MELLO E SOUZA (Org.). **Dimensões da Avaliação Educacional**. Petrópolis, RJ: Vozes, 2005. p.246-258.

CASTRO, Maria Helena Guimarães de. Sistemas de avaliação da educação no Brasil: avanços e novos desafios. **São Paulo Perspectiva**. São Paulo, v. 23, n. 1, p. 05-18, jan./jun. 2009.

COLENCI JUNIOR, Alfredo *et al.* A falta de engenheiros, o desenvolvimento econômico e a educação no Brasil. In: XXXIX Congresso Brasileiro de Educação em Engenharia, 2011, Blumenau. **Anais...COBENGE**, 2011, p. 1-10.

DAMETTO, Jarbas. Notas preliminares sobre as avaliações em larga escala e sua incidência sobre a subjetividade docente. **Revista Espaço Acadêmico**, n.138, p. 56-63, nov. 2012.

DANTAS, Lys Marias Vinhaes. **As contribuições das políticas de avaliação educacional em larga escala: o caso da avaliação de aprendizagem na Bahia**. 2009. Tese (Doutorado em Educação) – Faculdade de educação, Universidade Federal da Bahia, Salvador, 2009. Disponível em: <<https://repositorio.ufba.br/ri/bitstream/ri/11878/1/Lys%20Dantas.pdf>>. Acesso em: 12 maio 2014.

ELLIOT, Ligia Gomes. Meta-avaliação: das abordagens às possibilidades de aplicação. **Revista Ensaio: Avaliação e Políticas Públicas em Educação**, Rio de Janeiro, v. 19, n. 73, p. 941-964, out./dez. 2011.

FERRÃO, Maria Eugênia *et al.* O SAEB – Sistema Nacional de Avaliação da Educação Básica: objetivos, características e contribuições na investigação da escola eficaz. **Revista Brasileira de Estudos de População**, Rio de Janeiro, RJ, v.18, n.1/2, jan./dez. 2001.

FIGUEIRÊDO, Maria do Amparo Caetano de. Metodologia de oficina pedagógica: uma experiência de extensão com crianças e adolescentes. **Revista Eletrônica Extensão Cidadã**, v.2, 2006. Disponível em: <http://www.okara.ufpb.br/ojs/index.php/extensaocidada/article/view/1349>. Acesso em: 26 abr. 2014.

FONTANIVE, Nilma Santos. O uso pedagógico dos testes. In: MELLO E SOUZA (Org.). **Dimensões da Avaliação Educacional**. Petrópolis, RJ: Vozes, 2005. p.139-173.

FONTANIVE, Nilma Santos & KLEIN, Ruben. Uma visão sobre o sistema de avaliação da educação básica do Brasil – SAEB. Rio de Janeiro, **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 29, n.8, p. 409-439, 2000.

FOUREZ, Gérard. Crise no ensino de ciências? **Investigações em Ensino de Ciências**, Rio Grande do Sul, RS, v.8(2), p.109-123, 2003.

GATTI, Bernardete Angelina. Avaliação educacional no Brasil: experiências, problemas, recomendações. **Estudos em Avaliação Educacional**, São Paulo, SP, n.10, p. 67-80, 1994.

_____. Avaliações de Sistemas Educacionais no Brasil. **Revista de Ciências da Educação**, São Paulo, SP, n.9, p. 7-18, maio/ago., 2009.

KLEIN, Ruben. Testes de rendimento escolar. In: MELLO E SOUZA (Org.). **Dimensões da Avaliação Educacional**. Petrópolis, RJ: Vozes, 2005. p.110-138.

KRASILCHIK, Myriam. Reformas e realidade: o caso do ensino das ciências. **São Paulo Perspectivas**. São Paulo, v.14, n.1, p. 85-93, 2000. Disponível em: <<http://www.scielo.br/pdf/spp/v14n1/9805.pdf>>. Acesso em: 27 mar. 2012.

LAUGLO, Jon. Críticas às prioridades e estratégias do Banco Mundial para a Educação. **Cadernos de Pesquisa**, São Paulo, SP, n.100, p. 11-36, mar. 1997.

MIRANDA, Glaura Vasques de. Escola Plural. **Estudos Avançados**, São Paulo, SP, v. 21, n. 60, p. 61-74, 2007.

MORTIMER, Eduardo Fleury. Construtivismo, mudança conceitual e ensino de ciências: para onde vamos?. **Investigações em Ensino de Ciências**, Porto Alegre, RS, v.1, p. 20-39, 1996.

NASCIMENTO JÚNIOR, Antônio Fernandes. **Construção de estatutos de ciência para a biologia numa perspectiva histórico-filosófica**: uma abordagem estruturante para seu ensino. 2010. Tese (Doutorado em Educação para a Ciência) - Faculdade de Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Bauru, 2010. Disponível em: <http://lesec.icb.ufg.br/uploads/263/original_Tese_Estatutos_do_Conhecimento_Biol%C3%B3gico_Nascimento_Jr.pdf>. Acesso em: 29 jul. 2014.

OECD. **Assessing Scientific, Reading and Mathematical Literacy**: A Framework for PISA 2006, OECD Publishing, 2006.

OLIVEIRA, Lina Kátia Mesquita de. **Três investigações sobre escalas de proficiência e suas interpretações**. 2008. Tese (Doutorado em Educação) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

SANFELICE, José Luís. O Manifesto dos Educadores (1959) à luz da história. **Educação & Sociedade**, Campinas, v.28, n.99, p. 542-557, maio/ago. 2007. Disponível em: <<http://www.cedes.unicamp.br>>. Acesso em: 28 mar. 2012.

SANTOS, Wildson Luiz Pereira dos. Educação Científica na perspectiva de letramento como prática social: funções, princípios e desafios. **Revista Brasileira de Educação**, Rio de Janeiro, RJ, v.12, n. 36, p. 474-550, set./dez. 2007.

SÃO PAULO. Sumário Executivo. **SARESP 2008**. São Paulo, SP, 2008.

_____. **SARESP Matrizes de Referência para a avaliação**. São Paulo, SP, 2008.

SCHWARTZMAN, Simon. A Ciência no período de pós-guerra. In: **II Módulo do Programa de Política e Administração em Ciência e Tecnologia**, CNPq, 1989. Disponível em: <<https://archive.org/details/ACienciaNoPeriodoDePosGuerra>>. Acesso em: 27 mar. 2012.

SCRIVEN, Michael. **Key evaluation checklist**. The Evaluation Center, Kalamazoo, MI, 2007. Disponível em: <http://www.wmich.edu/evalctr/archive_checklists/kec_feb07.pdf>. Acesso em: 12 jul. 2014.

SILVA, Maria Juliana de Almeida. O sistema mineiro de avaliação da educação pública: impactos na escola fundamental de Uberlândia. **REICE - Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación**, v. 5, n. 2, p. 241-253, 2007.

SILVA, Sérgio Eustáquio da. **Utilização e apropriação do AVALIA-BH pelos diretores, coordenadores pedagógicos e professores da Rede Municipal de Educação de Belo Horizonte**. 2013. Dissertação (Mestrado Profissional em Gestão e Avaliação da Educação Pública) – Faculdade de Educação da Universidade Federal de Juiz de Fora, Juiz de Fora, Minas Gerais, 2013.

STUFFLEBEAM, Daniel. The Metaevaluation Imperative. **American Journal of Evaluation**, v. 22, n.2, p. 183-209, 2001.

TENENBLAT, Ketí (Coord.). **O Ensino de Ciências e a Educação Básica: propostas para superar a crise**. Rio de Janeiro: Academia Brasileira de Ciências, 2007.

TROJAN, Rose Mari *et al.* **A influência da OCDE nas políticas de avaliação sob a perspectiva do PISA e do TALIS**. In: 5º ENCONTRO INTERNACIONAL DA SOCIEDADE BRASILEIRA DE EDUCAÇÃO COMPARADA, Belém, PA, 2012. Disponível em: <http://www.sbec.org.br/evt_2012.php>. Acesso em: 25 ago. 2013.

VIANNA, Heraldo Marelim. Avaliação do desempenho em matemática e ciências: uma experiência em São Paulo e em Fortaleza. **Estudos em Avaliação Educacional**, São Paulo, SP, n.5, p. 107-120, jan./jun. 1992.

WASELFISZ, Julio Jacobo. **O ensino das Ciências no Brasil e o PISA**. 1.ed. São Paulo: Sangari do Brasil, 2009.

WALDHELM, Mônica de Cássia Vieira. **Como aprendeu ciências na educação básica quem hoje produz ciência?** 2007. Tese (Doutorado em Educação) – Departamento de Educação, PUC-Rio, Rio de Janeiro, 2007. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=125319>. Acesso em: 29 jul. 2014.

ZOTTI, Solange Aparecida. **Sociedade, educação e currículo no Brasil** - dos jesuítas aos anos de 1980. 1.ed. Campinas, SP: Autores Associados, 2004.