

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E AVALIAÇÃO DA EDUCAÇÃO
PÚBLICA

PATRÍCIA DE LIMA PAULA

**A INFLUÊNCIA DA ELABORAÇÃO DE ITENS NA EFICÁCIA DAS AVALIAÇÕES
EM LARGA ESCALA**

JUIZ DE FORA

2018

PATRÍCIA DE LIMA PAULA

**A INFLUÊNCIA DA ELABORAÇÃO DE ITENS NA EFICÁCIA DAS AVALIAÇÕES
EM LARGA ESCALA**

Dissertação apresentada como requisito parcial para a qualificação do Mestrado Profissional em Gestão e Avaliação da Educação Pública, da Faculdade de Educação, Universidade Federal de Juiz de Fora.

Orientador: Prof. Dr. Tufi Machado Soares

JUIZ DE FORA

2018

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Paula, Patrícia de Lima.

A INFLUÊNCIA DA ELABORAÇÃO DE ITENS NA EFICÁCIA DAS AVALIAÇÕES EM LARGA ESCALA / Patrícia de Lima Paula. -- 2018.

100 f. : il.

Orientador: Tufi Machado Soares

Dissertação (mestrado profissional) - Universidade Federal de Juiz de Fora, Faculdade de Educação/CAEd. Programa de Pós Graduação em Gestão e Avaliação da Educação Pública, 2018.

1. Avaliação. 2. TRI. 3. Testes de Proficiência. I. Soares, Tufi Machado , orient. II. Título.

AGRADECIMENTOS

A realização desta dissertação de Mestrado contou com importantes apoios e incentivos de diversas pessoas.

Aos meus pais, por me terem dado educação e valores. Por me apoiarem em todas as decisões que tomei e por serem fonte de um amor incondicional na minha vida.

Ao CAEd pela oportunidade de realizar esta pesquisa e poder contribuir um pouco com o imenso trabalho que a instituição realiza.

Ao Prof. Dr. Tufi Machado Soares, pela sua orientação, disponibilidade, ensinamentos e opiniões na execução deste trabalho.

Ao Prof. Dr. Marcelo de Oliveira Santos, pelo incentivo e acompanhamento dessa minha trajetória. Pela competência, profissionalismo e dedicação tão importantes.

Aos membros da banca examinadora, Prof^a Dr^a Carla da Silva Machado e Prof^a Dr^a Naira da Costa Muylaert Lima, que tão gentilmente aceitaram participar e colaborar com esta dissertação.

À Luisa Vilard do PPGP, simplesmente por tudo que fez por mim durante a produção desta dissertação. Pela paciência com todas minhas indagações e dúvidas, pela dedicação em pesquisar sobre o tema mesmo não sendo da sua área, por toda orientação quando já não sabia mais o que fazer. Obrigada de coração!

Ao meu marido, Sandro Mendes pela sua incansável boa vontade em me ajudar, por ser meu revisor e por perder e fins-de-semana ao meu lado, compartilhando meus ideais e incentivando-me a prosseguir. Obrigada pelo presente de cada dia, pelo seu sorriso e por saber me fazer feliz.

Ao Me. Vinícius da Silva Carvalho, pela amizade, companheirismo e ajuda com o que foi preciso na execução deste trabalho.

Gostaria de expressar toda a minha gratidão e apreço a todos aqueles que, direta ou indiretamente, contribuíram para que este trabalho se tornasse uma realidade. A todos quero manifestar os meus sinceros agradecimentos.

RESUMO

A presente dissertação é desenvolvida no âmbito do Mestrado Profissional em Gestão e Avaliação da Educação (PPGP) do Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora (CAEd/UFJF). O caso de gestão estudado discutiu o processo de elaboração dos itens integrantes dos testes de proficiência de três programas de avaliação: o Sistema de Avaliação Educacional do Desempenho Educacional do Amazonas (SADEAM), o Programa de Avaliação do Espírito Santo (PAEBES) e o sistema de Avaliação Baiano da Educação (Avalia BA). Os objetivos definidos para este estudo foram baseados em compreender os fatores que interferem na qualidade dos itens de biologia nas avaliações dos Programas de Avaliação no período de 2010 a 2015. Para alcançar tal objetivo, foi necessário descrever as análises estatísticas dos relatórios gerados pelo Centro de Políticas Públicas e Avaliação da Educação (CAEd) que serviram de base para o estudo. Assumimos como hipótese o fato de vários itens dos conteúdos de biologia apresentarem baixos percentuais de acerto, além de um quantitativo elevado de itens eliminados. Para tanto, utilizamos, como metodologia de pesquisa a realização de uma análise psicométrica dos itens de Biologia do Ensino Médio com baixos percentuais de acerto dos Programas de Avaliação SADEAM, PAEBES e AVALIA BA e, a partir disso, a realização de análises pedagógicas, utilizando, como instrumentos, os relatórios estatísticos gerados pela Coordenação de Instrumentos de Avaliação (CIA) que apresentam o comportamento estatístico de cada item dos testes de proficiência aplicados.

Palavras-Chave: Avaliação. TRI. Testes de Proficiência.

ABSTRACT

The present dissertation is developed in the Mestrado Profissional em Gestão e Avaliação da Educação (PPGP) do Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora (CAEd/UFJF). The management case studied discussed the process of elaborating the items that are part of the proficiency tests of three evaluation programs: the Educational Evaluation System of the Educational Performance of Amazonas (SADEAM), the Evaluation Program of the State of Espírito Santo (PAEBES) and the Baiano Education Assessment System (Evaluates BA). The objectives defined for this study were based on understanding the factors that influence the quality of biology items in the evaluation of the Evaluation Programs in the period from 2010 to 2015. To achieve this goal, it was necessary to describe the statistical analyzes of the reports generated by the Centro de Políticas Públicas e Avaliação da Educação (CAEd), that served as the basis for the study. We assume as hypothesis the fact that several items of the contents of biology present low percentages of correctness, besides a high quantitative of eliminated items. To do so, we used as a research methodology the performance of a psychometric analysis of the items of High School Biology with low percentage of correctness of the Evaluation Programs SADEAM, PAEBES and AVALIA BA Evaluation and, from this, the accomplishment of pedagogical analyzes, using , as instruments, statistical reports generated by the Coordination of Assessment Instruments (CIA) that present the statistical behavior of each item of applied proficiency tests.

Keywords: Evaluation. TRI. Proficiency Tests.

LISTA DE ILUSTRAÇÕES

Figura 1 - Diagrama de elaboração dos itens.....	23
Figura 2 - Estrutura de um item.....	24
Figura 3 - Escala de Proficiência.....	26
Figura 4 - Recorte da Matriz de Referência de Ciências da Natureza do 3º ano do Ensino Médio do PAEBES	34
Figura 5 - Etapas do processo de elaboração de itens	36
Figura 6 - Esquema da elaboração dos itens.....	37
Figura 7 - O item e suas partes.....	38
Figura 8 - Esquema de montagem dos testes de proficiência.....	41
Figura 9 - Níveis da Escala de Proficiência em Língua Portuguesa do SADEAM.....	46
Figura 10 - Níveis de Desempenho do PAEBES.....	49
Figura 11 - Níveis de Proficiência do Avalia Ba.....	52
Figura 12 - Curva Característica do Item	56
Figura 13 - Modelo do relatório de um dos itens na aplicação de um teste	62
Figura 14 - Níveis de Complexidade da Taxonomia de Bloom.....	66
Figura 15 – Item espelho do SADEAM.....	67
Figura 16 - Relatório estatístico do Item.....	68
Figura 17 – Item espelho do PAEBES	70
Figura 18 - Estatísticas do Item do PAEBES.....	71
Figura 19 – Item espelho do Avalia Ba.....	73
Figura 20 - Estatísticas do Item do Avalie Ba.....	74
Figura 21- Item espelho do SADEAM	75
Figura 22 - Estatísticas do Item do SADEAM.....	76

LISTA DE QUADROS

Quadro 1- Detalhamento das Avaliações do SAEB	16
Quadro 2 – Habilidades da Competência 2.....	81
Quadro 3 – Domínio Matéria e Ambiente da Matriz de Referência do 3º ano do Ensino Médio de Biologia	82

LISTA DE TABELAS

Tabela 1 - Sistemas de Avaliação da Educação	33
Tabela 2 - Trajetória do SADEAM	44
Tabela 3 - Trajetória do PAEBES.....	47
Tabela 4 - Trajetória do Avalia Ba	50
Tabela 5 - Classificação de itens de acordo com a discriminação	58
Tabela 6 - Classificação dos itens de acordo com o parâmetro b	58
Tabela 7 - Percentual de acerto	63
Tabela 8 - Temas gerais dos itens com níveis de acerto de até 20%	63
Tabela 9 - Descritores utilizados na produção dos itens analisados	64
Tabela 10 – Características de Elaboração dos Itens de Citologia com baixos percentuais de Acerto	78

LISTA DE ABREVIATURAS

BIB	Bloco Incompleto Balanceado
BNCC	Base Nacional Comum Curricular
CAEd	Centro de Políticas Públicas e Avaliação da Educação
CCIs	Curva Característica dos Itens
CETPP	Centro de Estudos de Testes e Pesquisas Psicológicas da Fundação Getúlio Vargas
CIA	Coordenação de Instrumentos de Avaliação
ECIEL	Programa de Estudos Conjuntos de Integração Econômica Latino-americana
EJA	Educação de Jovens e Adultos
FNDE	Fundo Nacional de Desenvolvimento de Educação
FUNDEB	Fundo da Educação Básica
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
LDB	Lei de Diretrizes e Bases da Educação Nacional
MEC	Ministério da Educação
OECD	<i>Organisation for Economic Co-operation and Development</i>
PAEBES	Programa de Avaliação da Educação Básica do Espírito Santo
PCN	Parâmetros Curriculares Nacionais
PISA	<i>Programme for International Student Assessment</i>
PNUD	Programa das Nações Unidas para o Desenvolvimento
SABE	Sistema de Avaliação Baiano da Educação
SADEAM	Sistema de Avaliação Educacional do Desempenho Educacional do Amazonas
SAEB	Sistema de Avaliação da Educação Básica
SAEP	Sistema de Avaliação da Educação Primária
SEDUC/AM	Secretaria de Educação e Qualidade de Ensino
SREs	Secretarias Regionais de Ensino
TCT	Teoria Clássica dos Testes
TRI	Teoria de Resposta ao Item

UFJF	Universidade Federal de Juiz de Fora
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura
UNICEF	Fundo das Nações Unidas para a Infância é um órgão das Nações Unidas

SUMÁRIO

INTRODUÇÃO

1 A ELABORAÇÃO DE ITENS DE CIÊNCIAS DA NATUREZA PELO CENTRO DE POLÍTICAS PÚBLICAS E AVALIAÇÃO DA EDUCAÇÃO	13
1.1 AS AVALIAÇÕES EXTERNAS.....	13
1.1.1 As Matrizes de Referência, Itens e Escala de Proficiência	20
1.2 AS AVALIAÇÕES EXTERNAS NO CONTEXTO DAS CIÊNCIAS DA NATUREZA	27
1.3 O PROCESSO DE ELABORAÇÃO DE ITENS PELO CENTRO DE POLÍTICAS PÚBLICAS E AVALIAÇÃO DA EDUCAÇÃO.....	34
1.4 MONTAGEM E APLICAÇÃO DOS TESTES DE PROFICIÊNCIA.....	40
1.5 OS PROGRAMAS DE AVALIAÇÃO ANALISADOS.....	42
1.5.1 SADEAM.....	43
1.5.2 PAEBES.....	46
1.5.3 AVALIE BA.....	49
2 A ANÁLISE DOS ITENS DE BIOLOGIA	54
2.1 A APLICAÇÃO DA TEORIA DE RESPOSTA AO ITEM	54
2.2 OS PARÂMETROS ESTATÍSTICOS	57
2.3 METODOLOGIA DE PESQUISA	59
2.3.1 Instrumentos Metodológicos	61
2.3.2 Os Relatórios estatísticos.....	62
2.3.3 Análise dos itens	63
3 PLANO DE AÇÃO	79
3.1 Reformulação dos descritores das Matrizes de Referência	79
CONSIDERAÇÕES FINAIS	86
REFERÊNCIAS.....	89

INTRODUÇÃO

Esta dissertação, desenvolvida no Programa de Mestrado Profissional da Faculdade de Educação da Universidade Federal de Juiz de Fora, vincula-se à linha de pesquisa em Avaliação, Currículos e Desenvolvimento Profissional de Gestores e Professores da Educação Básica.

A avaliação em larga escala foi implantada em diversos níveis de ensino e tem sido disseminada no Brasil como uma importante ferramenta para o monitoramento da educação básica, pois apresenta um diagnóstico do sistema educacional brasileiro.

O Centro de Políticas Públicas e Avaliação da Educação (CAEd), da Universidade Federal de Juiz de Fora (UFJF), é uma instituição que trabalha com a avaliação em larga escala, cujos testes de proficiência e questionários socioeconômicos avaliam o desempenho escolar dos estudantes. Alguns destes testes são elaborados a partir da Teoria de Resposta ao Item (TRI) que analisa e avalia os resultados obtidos pelos estudantes, levando em consideração as habilidades demonstradas e os graus de dificuldade dos itens, permitindo a comparação entre testes realizados em diferentes anos.

A construção de um banco de itens que podem compor as avaliações externas realizadas pelo CAEd é feita na Coordenação de Instrumentos de Avaliação (CIA). Essa coordenação é composta por laboratórios que são divididos por áreas de conhecimento, tais como: Ciências da Natureza, Ciências Humanas, Matemática, Língua Portuguesa e Alfabetização, e tem como foco a elaboração, montagem e análise das avaliações.

No laboratório de Ciências da Natureza atuei, até 2017, como Analista de Instrumentos de Avaliação. Nesse cargo, eu coordenava e auxiliava uma equipe composta por outros analistas e auxiliares responsáveis pela elaboração dos itens que compõem os testes de proficiência das áreas de química, física e biologia.

A qualidade desses itens está atrelada a adequação deles à Teoria de Resposta ao Item (TRI) e às técnicas específicas de elaboração, tais como a utilização de suportes adequados e a plausibilidade dos distratores, para que assim o teste se torne efetivo nas suas atribuições. Os itens que não atendem aos critérios da TRI não são capazes de realizar uma boa aferição das habilidades desenvolvidas

pelos estudantes, dificultando sua manutenção no banco de itens, já que eles podem comprometer a eficácia dos testes.

O controle de qualidade dos itens é realizado através de testes que geram relatórios. Enquanto analista de instrumentos de avaliação, pude perceber, através dos relatórios gerados, que nos testes de Ciências da Natureza, os itens relacionados aos conteúdos de biologia apresentavam baixos índices de acerto. Frente a isso, o presente estudo buscará responder à seguinte pergunta: por que vários itens dos conteúdos de biologia apresentam baixos percentuais de acerto além de um quantitativo de itens eliminados?

Em função disso, essa pesquisa tem como objetivo geral compreender quais são os fatores que interferem na qualidade dos itens de biologia nas avaliações dos projetos SADEAM, PAEBES e AVALIA BA, no período de 2010 a 2015. A justificativa para a escolha desses Programas se deu em função da avaliação das etapas do Ensino Médio, matrizes e período avaliado semelhantes.

Para alcançar tal objetivo geral, elencamos os seguintes objetivos específicos:

- 1) Descrever as análises estatísticas dos relatórios gerados pelo Centro de Políticas Públicas e Avaliação da Educação (CAEd) que servirão de base para o estudo.
- 2) Realizar a análise psicométrica dos itens de Biologia do Ensino Médio com baixos percentuais de acerto dos projetos SADEAM, PAEBES e AVALIA BA e a partir disso realizar análises pedagógicas.
- 3) Propor medidas que auxiliem na melhoria da elaboração dos itens de forma que se obtenha uma melhor efetividade nos testes.

Na execução desta pesquisa, foram utilizados os relatórios estatísticos gerados pela Coordenação de Instrumentos de Avaliação (CIA) que apresentam o comportamento estatístico de cada item dos testes de proficiência aplicados. Para alcançar os objetivos propostos, o trabalho está estruturado em três capítulos: o Capítulo 1 abordará o surgimento da avaliação em larga escala no Brasil, o cenário da elaboração de itens nos testes de proficiência e o processo de construção das matrizes de referência e das escalas de proficiência. Nesse capítulo também será levantado o contexto das Ciências da Natureza nas avaliações externas através de uma abordagem histórica, a abrangência das disciplinas nessas avaliações e a presença da Biologia nos Sistemas de Avaliação Estaduais e Municipais. Na sequência, será apresentado o processo de elaboração dos itens pelo Centro de Políticas Públicas e Avaliação da Educação através da dinâmica de organização das

coordenações e sua atuação nas avaliações em larga escala, assim como o sistema organizacional no processo de elaboração de itens. Por fim, será apresentado os Programas de Avaliação do Amazonas (SADEAM), do Espírito Santo (PAEBES) e da Bahia (AVALIA Ba) que serviram de base de análise da pesquisa e atrelado a isso, serão apresentadas a elaboração de itens, de acordo com o Guia de Elaboração do Itens do CAEd, e a montagem dos testes de proficiência.

No Capítulo 2, foram analisados e discutidos os parâmetros dos itens selecionados dos Programas de Avaliação, com o objetivo de identificar os fatores que interferem na qualidade dos itens. Nesse sentido, será apresentada a aplicação da Teoria de Resposta ao Item, sua importância, metodologia, aplicação nas avaliações, sua influência na elaboração dos itens e o Guia de Elaboração utilizado como instrumento para a produção dos itens no CAEd.

Será realizada, ainda nesse capítulo, uma análise psicométrica dos itens de biologia que compõem os testes de três projetos do CAEd através de relatórios estatísticos que apresentam os parâmetros da Teoria de Resposta ao Item e a Teoria Clássica dos Testes.

No terceiro capítulo será apresentar um Plano de Ação, cuja proposta é a reformulação de alguns descritores das matrizes de referências dos Programas de Avaliação pesquisados. A reformulação desses descritores visa a melhoria da qualidade dos itens que são elaborados na equipe de Ciências da Natureza do CAEd.

1 A ELABORAÇÃO DE ITENS DE CIÊNCIAS DA NATUREZA PELO CENTRO DE POLÍTICAS PÚBLICAS E AVALIAÇÃO DA EDUCAÇÃO

Este capítulo será dividido em cinco partes com o objetivo de apresentar uma breve trajetória histórica da avaliação em larga escala no Brasil e o cenário da elaboração de itens nos testes de proficiências.

Na primeira seção, será realizado um levantamento sobre a avaliação em larga escala e sua importância. Nesse sentido, é necessário entender como funciona o processo de elaboração dos itens a partir da criação das matrizes de referência e escalas de proficiência. Na segunda seção, será configurada a apresentação das Ciências da Natureza nas avaliações externas através dos principais sistemas de avaliação em larga escala do país no âmbito nacional e regional.

Já na terceira seção será apresentado o CAEd como ambiente de estudo e sua atuação nas avaliações em larga escala. Para situarmos o contexto da elaboração dos itens, será necessário entender a organização do CAEd e sua distribuição de atividades por coordenações.

Para uma melhor compreensão sobre como ocorre a elaboração dos itens que compõem os testes de proficiência das avaliações externas será realizada uma breve elucidação sobre a trajetória dessas avaliações na educação.

Na quarta seção, será apresentado o processo da montagem e da aplicação dos testes pelo CAEd a partir da elaboração de itens, e por último, na quinta seção, serão apresentados os Programas de Avaliação que serão analisados, bem como as suas matrizes, escalas de proficiência e padrões de desempenho.

1.1 AS AVALIAÇÕES EXTERNAS

As avaliações externas, também chamadas de avaliação em larga escala, são instrumentos para a elaboração de políticas públicas nos Sistemas de Ensino, já que elas geram diagnósticos sobre a qualidade da educação nas redes.

Gatti (2002) afirma que as avaliações externas no Brasil passaram a ser discutidas recentemente, em relação a outros países, e, com isso, sua valorização e desenvolvimento como campo teórico também sofreu os efeitos dos ressentimentos,

do desprezo e da crítica ideológica. Apesar disso, as preocupações específicas sobre os processos avaliativos baseados em instrumentos iniciaram-se na década de 1960.

Segundo a autora, em 1966, ocorreu a criação do Centro de Estudos de Testes e Pesquisas Psicológicas da Fundação Getúlio Vargas (CETPP) que foi responsável por estudos dos diferentes testes educacionais e pela criação de instrumentos de avaliação compostos por provas objetivas nas áreas de linguagem, matemática, ciências físicas e naturais e estudos sociais.

Já na década de 1970 foi realizado um estudo avaliativo de várias etapas de escolarização e características pessoais e socioeconômicas pelo Programa de Estudos Conjuntos de Integração Econômica Latino-americana (ECIEL). Esses estudantes foram submetidos a um questionário para levantamento de dados sobre a situação socioeconômica, atitudes relacionadas com o processo escolar e aspirações, e um exame de compreensão de leitura e ciências.

Segundo a Gatti,

Não houve, nos anos subsequentes, outras iniciativas visando estabelecer avaliações mais abrangentes, embora alguns ensaios localizados tenham sido realizados. Quanto à formação na área, nos cursos de Pedagogia desenvolvia-se uma disciplina sobre medidas educacionais que, ao longo do tempo, foi sendo suprimida. Algumas instituições ainda mantiveram, em sua grade curricular, uma disciplina mais geral sobre avaliação educacional. (GATTI, 2002. p. 22).

Na década de 1980, o país passava por um processo gradual de redemocratização que gerava tanto um clima de busca da universalização e qualificação da educação, quanto de percepção da necessidade de se construir instrumentos para a Avaliação em Larga Escala (GOMES NETO; ROSEMBERG, 1995). Nesse sentido, o Banco Mundial incentivou, nesse período, o desenvolvimento de estudos sobre a Avaliação Educacional o que contribuiu para a criação do Sistema de Avaliação da Educação Básica (SAEB).

Em 1988 o Ministério da Educação instituiu o Sistema de Avaliação do Ensino Público Primária – SAEP com o objetivo de oferecer subsídios para a formulação, reformulação e monitoramento de políticas públicas, contribuindo, dessa maneira, para a melhoria da qualidade do ensino brasileiro (BRASIL, 2008). No mesmo período, a ampla difusão e discussão das teorias da reprodução cultural e social expuseram a função seletiva e excludente da educação brasileira. A reflexão crítica marcava os debates na elaboração do novo texto constitucional (COELHO, 2008).

A Constituição Federal de 1988 foi importante no contexto das avaliações externas por garantir os padrões mínimos de qualidade que podem ser observados no seu art. 206 que prevê que o ensino será ministrado nos seguintes princípios:

- I - igualdade de condições para o acesso e permanência na escola;
- II - liberdade de aprender, ensinar, pesquisar e divulgar o pensamento, a arte e o saber;
- III - pluralismo de ideias e de concepções pedagógicas, e coexistência de instituições públicas e privadas de ensino;
- IV - gratuidade do ensino público em estabelecimentos oficiais;
- V - valorização dos profissionais da educação escolar, garantidos, na forma da lei, planos de carreira, com ingresso exclusivamente por concurso público de provas e títulos, aos das redes públicas;
- VI - gestão democrática do ensino público, na forma da lei;
- VII - garantia de padrão de qualidade;
- VIII - piso salarial profissional nacional para os profissionais da educação escolar pública, nos termos de lei federal (BRASIL, 1988).

No mesmo sentido, o art. 214 estabelece o plano nacional de educação como um dos resultados pretendidos à melhoria da qualidade do ensino (BRASIL, 2017).

Gatti afirma que

No fim dos anos 80, no Ministério da Educação, discutia-se sobre vários problemas que ocorriam nos sistemas educacionais apontados pelos pesquisadores da área de educação. No centro da discussão, estava o fracasso escolar brutal que se constatava pelas estatísticas e estudos no interior das escolas. Uma das questões levantadas é que não se possuíam dados sobre o rendimento escolar dos alunos quanto aos sistemas e fatores a ele associados (GATTI, 2002. p. 24).

Em função disso, o MEC promoveu uma oficina de trabalho com um grupo de educadores para debater a questão. A proposta era de analisar a possibilidade de uma nova estrutura política em relação aos ensinos fundamental e médio. Isso levou o Ministério da Educação a intervir menos diretamente nestes níveis e a propiciar orientações e subsídios técnicos, assumindo um papel avaliador como referência para políticas na área da educação básica (GATTI, 2002).

Com as alterações da Constituição de 1988, que ampliaram os direitos e garantias individuais, assim como o direito à saúde e à educação, o SAEP passa a chamar SAEB, Sistema de Avaliação da Educação Básica que teve sua primeira avaliação em 1990. Com participação amostral, essa aplicação ocorreu nas escolas que ofertavam as 1ª, 3ª, 5ª e 7ª séries do Ensino Fundamental em escolas públicas da rede urbana e os estudantes foram avaliados em Língua Portuguesa, Matemática e Ciências. As 5ª e 7ª séries também foram avaliadas em redação e este formato se manteve na edição de 1993 (BRASIL, 2015).

A segunda aplicação, em 1993, foi organizada em três eixos de estudo: (1) rendimento do aluno; (2) perfil e prática docentes; (3) perfil dos diretores e formas de gestão escolar. Nesse ano, os testes das avaliações do SAEB foram elaborados a partir dos conteúdos mínimos comuns às propostas curriculares estaduais. Algumas questões exigiram apenas a memorização do conteúdo, outras, a compreensão do que era solicitado e, em alguns casos, a resolução de um problema ou situação por meio da aplicação de algo já aprendido (BRASIL, 1995). Os resultados dessas aplicações foram gerados a partir das médias de acertos dos estudantes em uma escala de 0 a 100 pontos caracterizando a Teoria Clássica dos Testes (TCT).

A partir de 1995, foi incorporada uma nova metodologia estatística conhecida como Teoria de Resposta ao Item, TRI, que tem permitido, entre outras coisas, a comparabilidade dos diversos ciclos de avaliação (BRASIL, 2008).

Atualmente o SAEB é composto por três tipos de avaliação apresentados no Quadro 1.

Quadro 1- Detalhamento das Avaliações do SAEB

(continua)

Avaliação	Descrição	Período de aplicação
<p style="text-align: center;">Avaliação Nacional da Educação Básica (Aneb)</p>	<p>Aplicada de maneira amostral para alunos das redes públicas e privadas do país, matriculados na 4ª série/5º ano e 8ª série/9º ano do Ensino Fundamental e no 3º ano do Ensino Médio. Seu principal objetivo é avaliar a qualidade, a equidade e a eficiência da educação brasileira. Os resultados são apresentados como um todo (país), das regiões geográficas e das unidades da federação.</p>	<p style="text-align: center;">Bianualmente</p>
<p style="text-align: center;">Avaliação Nacional de Rendimento Escolar (Anresc)</p>	<p>Também conhecida por “Prova Brasil”, é uma avaliação baseada em uma coleta de dados para a constituição de uma análise, aplicada para alunos da 4ª série/5º ano e 8ª série/ 9º ano do Ensino Fundamental das escolas públicas das redes municipais, estaduais e federal. Seu objetivo é avaliar a qualidade do ensino ministrado nas escolas públicas. Para participar desta avaliação é necessário que exista um número mínimo de 20 alunos matriculados nas séries/ano. Os resultados são divulgados por escola e por unidade federativa.</p>	<p style="text-align: center;">Bianualmente</p>

Quadro 1- Detalhamento das Avaliações do SAEB

Avaliação	Descrição	Período de aplicação
<p>Avaliação Nacional da Alfabetização</p>	<p>Assim como a Prova Brasil, também é uma avaliação de caráter censitário que envolve alunos do 3º ano do Ensino Fundamental das escolas públicas. Seu principal objetivo é avaliar os níveis de alfabetização e letramento em Língua Portuguesa, alfabetização Matemática e condições de oferta do Ciclo de Alfabetização das redes públicas. Sua incorporação ao Saeb é recente, através da Portaria nº 482, de 7 de junho de 2013.</p>	<p>Anualmente</p>

Fonte: MARTINS, 2014, p. 26.

A Lei de Diretrizes e Bases da Educação Nacional (Lei 9394/96) assegurava a implementação de Políticas de Avaliação para a melhoria da qualidade da educação, já que, em seu Art. 9º, apontava que a União incumbir-se-ia de assegurar processo nacional de avaliação do rendimento escolar no ensino fundamental, médio e superior, em colaboração com os sistemas de ensino, objetivando a definição de prioridades e a melhoria da qualidade do ensino (BRASIL, 1996).

Com a implementação do SAEB, em 1990, surge o desafio da melhoria da qualidade do ensino e essa preocupação é apresentada na Conferência Mundial sobre Educação para Todos que ocorreu em Jomtien na Tailândia em 1990. Essa conferência teve a participação da UNESCO, UNICEF, PNUD e Banco Mundial que criaram um documento com definições e novas abordagens sobre as necessidades básicas e os compromissos dos Governos e outras entidades participantes. Dessa forma, em sequência à Conferência Mundial, os países foram incentivados a elaborar Planos Decenais de Educação Para Todos, em que as diretrizes e metas do Plano de Ação da Conferência fossem contempladas (MENEZES; SANTOS, 2001).

Assim, a Avaliação Educacional em Larga Escala tem sido utilizada no Brasil como uma importante ferramenta para o monitoramento da educação básica (FERNANDES *et al.*, 2010), pois tem como objetivo diagnosticar o sistema educacional. Apesar de o Brasil ter avançado na aplicação da avaliação em larga escala, ainda há muito a ser feito para que se desenvolvam mecanismos nos quais

professores e gestores encontrem caminhos para a melhoria do ensino (BECKER, 2010).

Essas avaliações devem ser compreendidas como um processo de avaliação diagnóstica que produz informações significativas sobre a realidade educacional do país. Processo este essencial para promover o debate público e favorecer a promoção de ações orientadas para a democratização do ensino, garantindo igualdade de oportunidades educacionais a todos (CAEd, 2012).

Entre as ações que visam fortalecer o ensino, a partir do diagnóstico apresentado pelos testes de larga escala, estão a gestão educacional, a formação de professores e dos profissionais de serviço e apoio escolar, a implantação de práticas pedagógicas e avaliação e a infraestrutura física e dos recursos pedagógicos que se desdobram em diferentes ações, cada um dos quais divididos em áreas de atuação, indicadores e critérios (WERLE, 2011).

Nesse sentido, as informações produzidas pelas avaliações educacionais podem auxiliar os gestores na implementação de várias políticas como o planejamento escolar, o acompanhamento e a regulação do sistema educacional, a prestação de contas à sociedade dos resultados alcançados, a implementação e o monitoramento das políticas públicas educacionais voltadas para a melhoria da qualidade e para promoção da equidade na educação; além da implementação de políticas de responsabilização institucional, tornando os gestores públicos, os profissionais da educação e as famílias dos estudantes corresponsáveis pelos resultados alcançados pelas escolas das redes públicas de ensino (CAEd, 2012).

A partir dessa premissa, Cária e Oliveira (2015) concluem que a avaliação em larga escala, enquanto ferramenta de monitoramento do processo educacional, associada à gestão, pode contribuir com a qualidade do ensino, bem como a adoção de compromissos assumidos por todos os atores envolvidos no processo educacional.

Em contrapartida, a avaliação em larga escala é alvo de críticas. Uma delas diz respeito às políticas de bonificação, adotadas por algumas Secretarias Estaduais de Educação que utilizam os resultados de desempenho das escolas e provocam discussões acerca dos objetivos das avaliações em larga escala. Essa política resume-se em destinar uma bonificação aos profissionais de educação que alcançam as metas estabelecidas, na maioria das vezes, por meio de 14^o, 15^o e 16^o salários, como exemplo do Município de Costa Rica – MS. Vale ressaltar que esses

valores não são incorporados aos vencimentos dos profissionais e não contribuem para o avanço dos planos de carreira do magistério.

Nesse caso, as avaliações em larga escala passam a estabelecer rankings escolares que podem ser interpretados como punições, provocando um efeito inesperado, como o treinamento dos alunos para responder o teste, um aumento do número de simulados que causam a perda de aulas, estímulo aos piores alunos a faltarem as aulas, entre outros (RAVICH, 2010).

Esquinsani e Dametto (2017) indicam que boa parte dos debates sobre a educação nacional são conduzidos pelos resultados das avaliações e, por isso, deixam de ser apenas um nível de avaliação ou um dos possíveis indicadores para demarcar a qualidade da educação limitando-se à própria política educacional.

Segundo os autores

O foco estabelecido, aparentemente (e apenas), no produto mensurável dessas avaliações (notas individuais, pontuações e rankings institucionais), pode obstaculizar outras análises e aplicações para os indicadores produzidos, limitando a potencialidade de uma avaliação em larga escala e minimizando a possibilidade de conter um caráter positivamente transformador ante a construção de uma agenda educacional (ESQUINSANI, 2012).

Freitas (2005) e Coelho (2008) acreditam que as avaliações em larga escala promovem o engessamento no currículo, pois acabam por definir o conteúdo escolar e gerar práticas escolares como treinamento da leitura, da escrita e do cálculo e, com isso, segundo Esteban e Fetzner, (2015) destituem

a) a leitura, a escrita e o cálculo de sua função social, de sua utilização no desenvolvimento da sociedade, transformam em objeto da aprendizagem o que seria um instrumento da aprendizagem, ou seja, ao insistir na abordagem mecânica destas atividades e com a utilização de exercícios maçantes, descolam estes conteúdos do seu sentido social: a comunicação de ideias e a organização das informações, por exemplo.

b) a escola de sua função social. Ao definir externamente o conteúdo da escola (por meio da política de exames), impede-se a escola de pensar-se, fazer-se, propor-se, enquanto um projeto coletivo da comunidade, das famílias e dos educadores que dela participam. A escola é ocupada por treinamentos para a prova, com exercícios que não apresentam a discussão de questões que possam ser consideradas interessantes pelos alunos ou relevantes para a comunidade. Não há espaço para a reflexão, mas para o exercício. As provas externas chegam prontas da Secretaria de Educação do município e, fora isto, os alunos são chamados aos simuladões que buscam preparar para a Prova Brasil e ao preenchimento dos cadernos de reforço escolar e do livro didático. O bom desempenho nas provas transforma-se em objetivo e fim da escola (ESTEBAN; FETZNER, 2015. p.86).

Esses fatores interferem nas práticas docentes que se tornam engessadas e voltadas apenas para o sucesso das avaliações externas, sobretudo nas disciplinas de língua portuguesa e matemática, que são as disciplinas mais frequentemente avaliadas.

Há, portanto, uma multiplicidade de percepções sobre as avaliações externas, mas é necessário que haja a responsabilização dos atores escolares para que essas avaliações possam, de fato, atingir os objetivos de apontar a realidade do ensino e assegurar a qualidade de educação, fortalecendo o direito de educação de qualidade para todos.

Na subseção a seguir, apresentaremos a estrutura de uma matriz de referência e como ocorre sua elaboração junto ao CAEd, mostrar como ocorre a elaboração de itens de acordo com o Guia de Elaboração do Itens do CAEd e a importância desse processo para a eficácia da avaliação. Além disso, serão apresentadas as escalas de proficiência e sua interpretação nas avaliações em larga escala.

1.1.1 As Matrizes de Referência, Itens e Escala de Proficiência

O Centro de Políticas Públicas e Avaliação da Educação (CAEd) da Universidade Federal de Juiz de Fora (UFJF) é uma instituição que operacionaliza programas estaduais e municipais destinados a mensurar o rendimento de estudantes das escolas públicas. Esse Centro atua junto ao Governo Federal, estados, municípios, instituições e fundações na realização de avaliações em larga escala com a produção de medidas de desempenho e na investigação de fatores intra e extraescolares associados ao desempenho. O objetivo principal do CAEd é oferecer dados e informações capazes de subsidiar as ações de melhoria da qualidade da educação e equidade nas oportunidades educacionais (UFJF, 2017).

O CAEd trabalha para implementar a avaliação em larga escala, que tem objetivos e procedimentos diferenciados das avaliações internas, isto é, são organizadas a partir de um sistema de avaliação cognitiva dos estudantes e são aplicadas de forma padronizada para um grande número de pessoas, entre os quais estão alunos, professores, diretores e coordenadores. Estas avaliações em larga escala utilizam instrumentos, tais como testes de proficiência e questionários, que

permitem avaliar o desempenho escolar e os fatores, de dentro e de fora do meio escolar (UFJF, 2017).

No CAEd, as avaliações em larga escala são formadas a partir das Matrizes de Referência. A princípio, formuladas em um diálogo entre os especialistas em educação do CAEd, profissionais da rede de ensino e gestores. As matrizes de referência apresentam o objeto de uma avaliação e são constituídas por um grupo de descritores que apresentam as habilidades que são esperadas dos alunos em diferentes etapas de escolarização e passíveis de serem aferidas em testes padronizados de desempenho.

Os conjuntos de questões, chamados itens, compõem o Bloco Incompleto Balanceado (BIB) que corresponde a um esquema otimizado para o rodízio de blocos com aplicações em diversas áreas. Esse rodízio é importante devido às aplicações apresentarem um grande número de itens, sendo necessário limitar a quantidade que é submetida a cada estudante em um valor aceitável e adequado de tempo de prova (BEKMAN, 2001). Portanto, o BIB corresponde a uma técnica utilizada na estruturação dos cadernos de testes que permite medir um número maior de habilidades e conteúdos.

Os itens são elaborados com base nos descritores das Matrizes de Referência das disciplinas avaliadas nos testes de proficiência, que, por sua vez, reúnem o conteúdo a ser avaliado em cada etapa escolar e disciplina, apresentando, assim, o que se espera do aluno em termos de desempenho escolar.

Esses itens têm como objetivo aferir a proficiência dos estudantes em determinada área do conhecimento em várias etapas de escolarização. Com isso, é importante que haja a definição das habilidades e competências a serem avaliadas em cada área de conhecimento para que se possam elaborar os itens a serem utilizados na composição dos testes (CAEd, 2008).

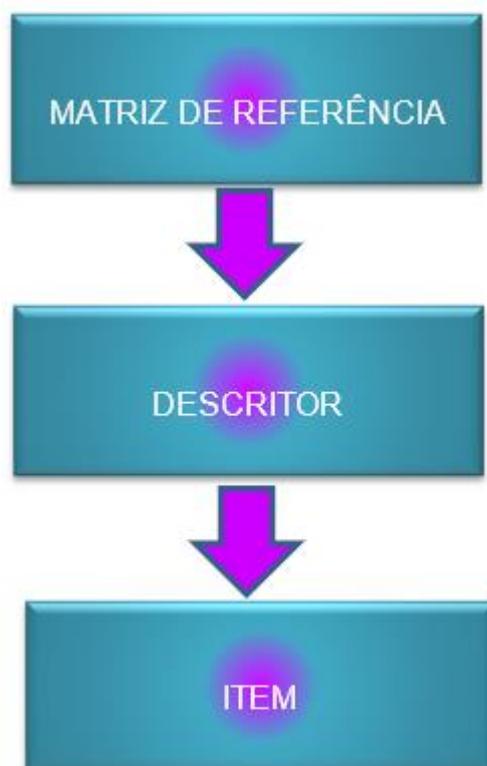
Segundo o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira as matrizes de referência,

Trata-se de uma referência para a construção do instrumento de avaliação, sendo diferente de uma proposta curricular ou programa de ensino, que são mais amplos e completos. A Matriz de Referência foca habilidades essenciais de alfabetização e letramento, ou seja, avalia os conhecimentos esperados que os alunos tenham adquiridos após o início do processo de alfabetização, mas de forma alguma deve substituir o currículo da escola em relação às habilidades de Língua Portuguesa e Matemática (BRASIL,2015).

As matrizes de referência devem, preferencialmente, ser criadas a partir de um debate com professores, especialistas e pesquisadores baseado nos Parâmetros Curriculares, livros didáticos, propostas curriculares do Estado ou Município e outros documentos que norteiam as avaliações nacionais e/ou estaduais e municipais. Essas matrizes são compostas por um conjunto de descritores, termos padronizados que descrevem as habilidades que são avaliadas por meio dos itens que compõem os testes, que abrangem o conteúdo a ser avaliado em cada etapa de escolarização e o nível de operação mental necessário para alcançar a habilidade avaliada. Esses descritores devem ser elaborados, junto com as matrizes, de forma que cada um possa ser avaliado por meio de itens de múltipla escolha.

Para a elaboração dos itens, é necessária a utilização das matrizes de referências que servem de base para a construção dos testes, uma vez que cada item perpassa por uma habilidade da matriz. Os itens são instrumentos de um teste e são elaborados a partir das matrizes de referências que, por sua vez, são compostas de descritores em determinada área do conhecimento. Esses descritores representam as habilidades que serão avaliadas nos testes de proficiência.

A Figura 1 apresenta um diagrama que ilustra o processo de elaboração dos itens dos testes de proficiência das avaliações em larga escala.

Figura 1 - Diagrama de elaboração dos itens

Fonte: CAEd, 2014.

Os itens são compostos de quatro elementos. O enunciado corresponde ao estímulo que mobiliza os recursos cognitivos dos estudantes e pode ser apresentado como uma situação problema, um texto e/ou imagens, os dois últimos são chamados de suportes, pois têm como objetivo auxiliar o estudante na resolução do item. Depois de apresentado o suporte, o item deve conter um comando para resposta que pode ser dada sob a forma de complementação ou de interrogação que deve estar atrelado ao descritor e à habilidade avaliada.

De acordo com o Guia de Elaboração de Itens (2014), esses itens devem apresentar as alternativas de resposta que podem ser quatro, quando forem itens do Ensino Fundamental, ou cinco nos itens de Ensino Médio. Dentre as alternativas, a correta é denominada gabarito, enquanto as demais são denominadas distratores.

A Figura 2 apresenta um exemplo de item de biologia para o Ensino Médio.

Figura 2 - Estrutura de um item

(B120025G5) **Observe as imagens abaixo.** → Enunciado

Golfinho

Ictiossauro

Tubarão

→ Suporte

Disponível em: <<http://www.colegioweb.com.br/wp-content/uploads/8793.jpg>>. Acesso em: 8 nov 2014.* Adaptado para fins didáticos.

Que tipo de processo de evolução pode ser identificado nesses seres vivos? → Comando

A) Analogia.
 B) Convergência adaptativa.
 C) Homologia.
 D) Irradiação adaptativa.
 E) Seleção artificial.

→ Alternativas

Fonte: Espírito Santo, 2015.

As escalas de proficiência são instrumentos capazes de traduzir medidas em diagnósticos qualitativos do desempenho escolar. Segundo Klein (2003), as escalas de proficiências correspondem a um conjunto de números ordenados que são obtidos pela Teoria de Resposta ao Item e que aferem a habilidade de uma determinada área de conhecimento. Assim, a probabilidade de se acertar um item aumenta à medida que a proficiência aumenta.

Nesse sentido, as escalas de proficiência apresentam os resultados do teste em uma régua na qual os valores de proficiência obtidos pelos estudantes são ordenados e categorizados em intervalos, indicando, assim, o grau de desenvolvimento das habilidades dos estudantes que alcançaram determinado nível de desempenho (ESPÍRITO SANTO, 2013), ou seja, cada intervalo de proficiências corresponde a um nível chamado de âncora.

De acordo com Oliveira, Franco e Soares (2007) a construção da escala de proficiência dos programas de avaliação educacional envolve dois procedimentos básicos. O primeiro é a identificação dos itens âncora (aqueles característicos ou representativos de cada nível âncora), que norteiam todo o trabalho de produção e interpretação das escalas.

Segundo Andrade, Tavares & Valle,

Para um item caracterizar um determinado nível âncora da escala, precisa ser respondido corretamente por pelo menos 65% dos indivíduos do nível de habilidade e por no máximo 50% dos indivíduos do nível de habilidade imediatamente anterior. Além disso, a diferença entre a proporção de indivíduos com esses níveis de habilidade que acertam o item deve ser de pelo menos 30%. Assim, para um item ser âncora ele deve ser um item “típico” daquele nível, ou seja, bastante acertado por indivíduos com aquele nível de habilidade e pouco acertado por indivíduos com um nível de habilidade imediatamente inferior (ANDRADE; TAVARES & VALLE, 2000).

Para que um item seja classificado como âncora a um determinado nível âncora (Z), pode-se simplesmente dizer que o item pertence ao nível âncora Z. Para isso, o item deve ser acertado por pelo menos 65% dos indivíduos com proficiências no intervalo correspondente ao nível Z e por o máximo de 50% dos indivíduos com proficiências no nível anterior. Além disso, a diferença entre as proporções de indivíduos com proficiência no nível âncora Z e no nível anterior deve ser pelo menos 30%.

Nesse sentido, a definição dos itens âncoras pode ser representada da seguinte forma:

$$P(U=1/\theta \in I(Z)) \geq 0,65$$

$$P(U=1/\theta \in I(Z-1)) < 0,65$$

$$P(U=1/\theta \in I(Z)) - P(U=1/\theta \in I(Z-1)) \geq 0,30$$

Onde

- U=1 representa o acerto ao item;
- $\theta \in I(Z)$ significa que a proficiência θ está no intervalo correspondente ao nível âncora Z;
- Z – 1 é o intervalo anterior a Z.

Já o segundo procedimento está relacionado à interpretação pedagógica desses itens. Com base nos resultados e na interpretação das escalas de proficiência, é possível produzir um diagnóstico de desempenho escolar indispensável para a melhoria da qualidade da educação nas escolas, pois as

apresentam os conhecimentos presentes na escala de proficiência e os descritores na Matriz de Referência.

Os conhecimentos estão dispostos nas linhas da escala e a gradação de cores indica a complexidade de cada capacidade. A partir disso, é possível realizar uma leitura horizontal na qual a cor amarelo-claro indica o primeiro nível de complexidade, passa pelo amarelo-escuro, laranja-claro, laranja-escuro e chega ao nível mais complexo representado pela cor vermelha. Essa leitura permite analisar como os estudantes desenvolvem as habilidades e realizar uma interpretação que auxilie no planejamento dos professores, assim como as intervenções pedagógicas.

É possível realizar uma leitura vertical da escala a partir dos padrões de desempenho, caracterizados por intervalos numéricos divididos em faixas de 25 pontos que podem ir de zero a quinhentos ou de zero a mil, conforme o projeto de avaliação externa.

Os padrões de desempenho permitem 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 apresentam um maior grau de desenvolvimento nas habilidades avaliadas e aqueles que apresentam menores graus de desenvolvimento nas referidas habilidades, permitindo se pensar em ações e políticas educacionais destinadas à promoção da equidade (CAEd, s/d).

A leitura vertical, por meio dos padrões de desempenho, permite identificar o panorama do desenvolvimento dos estudantes em um determinado intervalo, sendo possível relacionar os conhecimentos adquiridos com o percentual situado em cada padrão.

Após apropriar-se dos instrumentos de avaliação, e para melhor conhecer as políticas analisadas, será apresentado um breve histórico da área de Ciências da Natureza nas avaliações, mostrar como o ensino tem avançado e as iniciativas dos estados brasileiros em incluí-lo em suas políticas de avaliação educacional.

1.2 AS AVALIAÇÕES EXTERNAS NO CONTEXTO DAS CIÊNCIAS DA NATUREZA

A presença da disciplina de Ciências da Natureza, em um exame de avaliação externa no Brasil, iniciou-se junto ao Sistema de Avaliação da Educação

Básica (Saeb) cuja primeira edição, em 1990, avaliou de forma amostral os estudantes nas disciplinas de Língua Portuguesa, Matemática e Ciências.

Porém, em outros países, como nos Estados Unidos, a educação em ciências, nas escolas, ganhou grande relevância após o impacto do Sputnik soviético em 1957. Esse evento fez com que os EUA se sentissem pressionados a aumentar a pesquisa e o desenvolvimento em ciências para alcançar a URSS e, hoje, ela se expressa em um grande número de organizações científicas, profissionais com publicações especializadas em várias partes do mundo. É o caso da revista *Science Education* e instituições como *Association for Science Education* na Grã Bretanha, a *Association for the Promotion a Advancement of Science Education* no Canadá e a *National Science Teachers Association* nos Estados Unidos (SCHAWARTZMAN; CHRISTOPHE, s.d).

Segundo Schwartzman e Christoph (s.d.), o número reduzido de jovens que escolhem carreiras das áreas de natureza científica e tecnológica, até mesmo em países mais desenvolvidos, como a União Europeia, França e Estados Unidos, realça a preocupação com o ensino de ciências e a necessidade de introduzi-la nos primeiros anos da escola. Nesse sentido, os programas de educação em ciências, que fazem uso da indagação, experimentação, prática e trabalho em grupo, apresentam a ideia de que as pessoas deveriam ter um mínimo de competência e conhecimentos científicos, assim como os conceitos de alfabetização ou letramento utilizados nas avaliações de língua portuguesa e matemática.

Segundo os autores

Em relação aos conteúdos, há um debate permanente, na literatura especializada, a respeito de se a formação inicial deveria ser interdisciplinar ou por disciplinas, e se deveria se orientar por estandartes – conteúdo específicos que deveriam ser atingidos a cada etapa ou mais abertos. Em grande parte, este debate reproduz a oposição entre as correntes construtivistas, que dominaram o campo do ensino de ciências nos anos 60 e 70 nos Estados Unidos, e a tendência mais recente de reforçar os conteúdos e as disciplinas científicas (SCHWARTZMAN; CHRISTOPH, s.d.).

No Brasil, os Parâmetros Curriculares Nacionais (PCNs), publicados em 1997, apresentam as expectativas em relação à formação científica e tecnológica dos jovens. Os PCNs são documentos elaborados pelo Governo Federal que expressam as referências para o Ensino Fundamental e Médio do país, cujo objetivo é garantir a todas as crianças e jovens brasileiros, o direito de usufruir do conjunto de

conhecimentos reconhecidos como necessários para o exercício da cidadania (BRASIL, 1998).

De acordo com os Parâmetros Curriculares, o ensino de Ciências Naturais deve se organizar de forma que, ao final do ensino fundamental, os alunos tenham as seguintes capacidades:

- Compreender a natureza como um todo dinâmico, sendo o ser humano parte integrante e agente de transformações do mundo em que vive;
- Identificar relações entre conhecimento científico, produção de tecnologia e condições de vida, no mundo de hoje e em sua evolução histórica;
- Formular questões, diagnosticar e propor soluções para problemas reais a partir de elementos das Ciências Naturais, colocando em prática conceitos, procedimentos e atitudes desenvolvidos no aprendizado escolar;
- Saber utilizar conceitos científicos básicos, associados a energia, matéria;
- Transformação, espaço, tempo, sistema, equilíbrio e vida;
- Saber combinar leituras, observações, experimentações, registros, etc., para coleta, organização, comunicação e discussão de fatos e informações;
- Valorizar o trabalho em grupo, sendo capaz de ação crítica e cooperativa para a construção coletiva do conhecimento;
- Compreender a saúde como bem individual e comum que deve ser promovido pela ação coletiva;
- Compreender a tecnologia como meio para suprir necessidades humanas, distinguindo usos corretos e necessários daqueles prejudiciais ao equilíbrio da natureza e ao homem (BRASIL, 1997).

A partir disso, surgem as temáticas a serem estudadas de forma interdisciplinar e emerge o conceito de competência e a crítica ao ensino conteudista, com excesso de informações pouco significativas para os estudantes (SILVA, 2014). É também, nessa década que o movimento da avaliação em larga escala teve início 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 *Programme for International Student Assessment* (PISA) é um programa de avaliação internacional coordenado pela *Organisation for Economic Co-operation and Development* – OECD e, segundo o Ministério da Educação (2016), diferentemente das demais avaliações, permite ao Brasil aferir conhecimentos e habilidades dos estudantes de 15 anos em leitura, matemática e ciências,

contrastando com resultados do desempenho de alunos dos países membros da OCDE, além de 35 países/economias parceiras (BRASIL, 2015).

Esse programa influenciou a participação das Ciências da Natureza nos Sistemas de Avaliação Nacionais, uma vez que abrange as áreas do conhecimento – Leitura, Matemática e Ciências – havendo, a cada edição do programa, maior ênfase em cada uma dessas áreas.

O PISA dialoga com o SAEB, pois além de abordar o panorama da educação brasileira sob o foco do PISA, este relatório descreve as principais diferenças metodológicas e estruturantes entre esse programa e o Sistema de Avaliação da Educação Básica (SAEB). Propõe uma reflexão sobre as semelhanças e diferenças entre seus marcos referenciais e apresenta, ainda, um estudo da correlação entre os resultados recentes das duas avaliações. Entender as particularidades de cada uma delas, seus fundamentos, métodos e resultados é fundamental para melhor uso das informações pelos atores educacionais, bem como pelos gestores e pesquisadores da área (BRASIL, 2016).

Com isso, esse programa se caracteriza como um modelo para as políticas de avaliação, uma vez que aborda conceitos presentes nos discursos sobre avaliação no Brasil, como os de habilidades e competências. Esses conceitos são utilizados em vários programas de avaliação nacional, inclusive o SAEB. Além disso, o PISA utiliza a Teoria de Resposta ao Item (TRI) como modelo estatístico, assim como o SAEB, com a finalidade de comparar os dados dos resultados dos diferentes testes aplicados.

O Brasil teve sua participação no PISA desde a implementação do programa no ano de 2000 sob a responsabilidade do (INEP) e os relatórios gerados apresentam os resultados nacionais e regionais em leitura, matemática e ciências, além de expor dados sobre o contexto nacional e comparar os resultados com outros países participantes e entre as 27 unidades da Federação. O sistema de referência para o PISA, no país, é o Censo Escolar no qual uma lista de escolas com possíveis estudantes elegíveis (grupo de estudantes definido) é encaminhada ao consórcio internacional a cada edição.

Os instrumentos de avaliação do PISA aplicados no Brasil com maior destaque na área de ciências foram das edições de 2006 e 2015. De acordo com o Ministério da Educação (2016), a compreensão de ciência e tecnologia é indispensável na formação de um jovem para a vida na sociedade moderna e

segundo as diretrizes do PISA, tornar-se letrado cientificamente envolve a ideia de que os propósitos da educação na ciência devem ser amplos e aplicados; portanto, o conceito de letramento científico se refere tanto ao conhecimento da ciência como ao da tecnologia pautada na ciência.

Letramento científico, de acordo com a Organização para a Cooperação e Desenvolvimento Econômico (OCDE, 2016), é a capacidade de se envolver com questões relacionadas à ciência e com a ideia da ciência, como um cidadão reflexivo. Assim, uma pessoa letrada cientificamente está disposta a participar de uma discussão fundamentada sobre ciência e tecnologia que exige competências para explicar os fenômenos científicos, avaliar e planejar as investigações científicas e interpretar dados e evidências científicas.

Os resultados de ciências do PISA mostram que as escolas brasileiras não estão conseguindo transmitir, para grande maioria dos estudantes, as habilidades e competências científicas mínimas para o uso cotidiano das pessoas e, tampouco, está conseguindo formar uma pequena elite que possa, depois, profissionalizar-se como cientistas e pesquisadores plenos (SCHWARTZMAN; CHRISTOPH, s.d.).

Segundo Schawartzman e Christophe,

Ao contrário do que muitas vezes se pensa, os estudantes brasileiros não precisam ser estimulados a valorizar a ciência e reconhecer sua importância. Eles já fazem isto naturalmente, seja pelos estímulos que recebem na escola, seja, mais provavelmente, pelo que observam e captam nos ambientes em que vivem, circundados por toda parte pelos produtos das modernas tecnologias. Isto não traduz, no entanto, em competências e conhecimento efetivos para entender e fazer uso das abordagens e conhecimentos científicos em sua vida cotidiana e profissional. Isso significa que o que faz falta para os estudantes é a educação em ciências enquanto tal, muito mais do que atividades que estimulam o interesse pela ciência, que podem ser redundantes (SCHAWARTZMAN; CHRISTOPHE, s/d).

É importante ressaltar que a Base Nacional Comum Curricular, em debate no cenário educacional atual, ao abordar os conteúdos de Ciências da Natureza é bastante clara ao elencar os benefícios e alguns problemas relacionados a forma como a humanidade, por meio da ciência, tem gerido a vida na Terra, conforme seu texto:

A sociedade contemporânea está fortemente organizada com base no desenvolvimento científico e tecnológico. Desde a busca do controle dos processos do mundo natural até a obtenção de seus recursos, as ciências influenciaram a organização dos modos de vida. Ao longo da história, interpretações e técnicas foram sendo aprimoradas e organizadas como

conhecimento científico e tecnológico, da metalurgia, que produziu ferramentas e armas, passando por motores e máquinas automatizadas até os atuais chips semicondutores das tecnologias de comunicação, de informação e de gerenciamento de processos. No entanto, o mesmo desenvolvimento científico e tecnológico de notáveis progressos na produção e nos serviços também pode promover impactos e desequilíbrios na natureza e na sociedade, que demandam outras sabedorias, não somente científicas, para serem compreendidos e tratados (BRASIL, 2016).

Percebe-se, nessa introdução, que o desenvolvimento científico e tecnológico gera impactos e desequilíbrios relacionados à atuação da população nas questões ambientais e nos problemas de natureza social. Com isso, há a preocupação em promover um futuro do qual as crianças tenham consciência e compromisso ético na busca de melhores condições de vida percebendo-se como parte de um ser vivo que se relaciona com outros seres vivos. Esses são alguns dos principais desafios dos professores da área de Ciências da Natureza.

Com o intuito de conquistar esses desafios, a BNCC em ciências da Natureza propõe como objetivos imediatos e de grande importância:

- Compreender a ciência como um empreendimento humano, construído histórica e socialmente.
- Apropriar-se de conhecimentos das Ciências da Natureza como instrumento de leitura do mundo.
- Interpretar e discutir relações entre a ciência, a tecnologia, o ambiente e a sociedade.
- Mobilizar conhecimentos para emitir julgamentos e tomar posições a respeito de situações e problemas de interesse pessoal e social relativos às interações da ciência na sociedade.
- Saber buscar e fazer uso de informações e de procedimentos de investigação com vistas a propor soluções para problemas que envolvem conhecimentos científicos.
- Desenvolver senso crítico e autonomia intelectual no enfrentamento de problemas e na busca de soluções, visando transformações sociais e construção da cidadania.
- Fazer uso de modos de comunicação e de interação para aplicação e divulgação de conhecimentos científicos e tecnológicos.
- Refletir criticamente sobre valores humanos, éticos e morais relacionados com a aplicação dos conhecimentos científicos e tecnológicos. (BRASIL, 2017)

Através desses objetivos, pode-se observar que existe a necessidade de aprender e utilizar a ciência a partir de um processo de alfabetização e letramento científico que permita entender os aspectos técnicos das diferentes ciências e suas interações com os demais saberes em sua aplicação na realidade.

Nesse sentido, é importante concentrar esforços para a melhoria da qualidade do processo de ensino e aprendizagem de Ciências no Brasil e, nesse contexto,

alguns programas de avaliação em larga escala, ao longo dos anos, passaram a incluir essa disciplina nos testes de proficiência. A Tabela 1 apresenta a relação dos programas de avaliação externa brasileiros, em parceria com o CAEd, que avaliam as áreas de Ciências da Natureza e seus respectivos anos de criação.

Tabela 1 - Sistemas de Avaliação da Educação

Programa	Início	Disciplinas Avaliadas			
		LP	MAT	CN	CH
SISTEMAS MUNICIPAIS					
Avalia BH	2008	X	X	X	
Prova Floripa	2007	X	X	X	X
SAEMI	2013	X	X	X	X
SIMARE	2014	X	X	X	X
SISTEMAS ESTADUAIS					
Avalia Ba (SABE)	2011	X	X	X	X
PAEBES	2009	X	X	X	X
SADEAM	2008	X	X	X	X
SAERJINHO	2011	X	X	X	X
SALTO	2011	X	X	X	
SARESP	2008	X	X	X	X

Fonte: CAEd, 2017.

A partir dos dados da Tabela 1, é possível observar que a participação das Ciências da Natureza nas avaliações é mais expressiva que Ciências Humanas e que todos os sistemas de avaliação analisam as áreas de Língua Portuguesa e matemática.

O CAEd elabora testes de proficiência nas áreas de Língua Portuguesa, Matemática, Ciências Humanas e Ciências da Natureza. Esta última engloba Ciências no Ensino Fundamental e Biologia, Física e Química no Ensino Médio. Essas avaliações são produzidas de acordo com as matrizes de referência, de cada projeto, que apresentam descritores enquadrados nos domínios. Os domínios são uma subdivisão de acordo com o conteúdo, competências e habilidades que representam os eixos temáticos das áreas de conhecimento, além de permitir a interação entre os componentes curriculares e, a partir deles, elaborar os descritores que expressam habilidades, em Biologia, Física e Química.

Os domínios de Ciências da Natureza do Ensino Médio são matéria e energia, terra e universo, vida e ambiente, ser humano e saúde; e tecnologia e sociedade. Na disciplina de biologia, os domínios são representados por descritores de diversos conteúdos, como citologia, evolução, genética e ecologia. A Figura 4 mostra um

recorte da Matriz de Referência de Ciências da Natureza do 3º ano do ensino médio no projeto PAEBES.

Figura 4 - Recorte da Matriz de Referência de Ciências da Natureza do 3º ano do Ensino Médio do PAEBES

MATRIZ DE REFERÊNCIA CIÊNCIAS DA NATUREZA 3ª SÉRIE EM		
III. VIDA E AMBIENTE		
D48	B	Identificar a importância e o significado da nomenclatura e das classificações biológicas.
D49	B	Identificar os grupos de seres vivos dos reinos Monera, Protista, Fungi, Animalia e Plantae, quanto às características morfofisiológicas e evolutivas.
D50	B	Reconhecer a importância econômica e ecológica dos seres vivos dos reinos Monera, Protista, Fungi, Animalia e Plantae.
D51	B	Caracterizar a estrutura morfofisiológica dos vírus.
D52	B	Reconhecer a importância da reprodução sexuada e assexuada nos seres vivos.
D53	B	Reconhecer os processos de formação dos gametas.
D54	B	Identificar os componentes bioquímicos da célula e suas principais funções.
D55	B	Identificar a importância das organelas e do núcleo para o metabolismo celular.
D56	B	Associar características adaptativas dos animais vertebrados a diferentes ambientes.
D57	B	Reconhecer os processos de divisão celular.
D58	B	Compreender processo de síntese proteica.
D59	B	Identificar as relações ecológicas entre os seres vivos.
D60	B	Reconhecer causas de desastres ecológicos, relacionadas à ação antrópica.

Fonte: Espírito Santo, 2016.

Nesse recorte, observa-se o domínio “vida e ambiente” que engloba um conjunto de habilidades caracterizadas pelos descritores que vão do D48 ao D60 os quais, por sua vez, descrevem as habilidades que serão avaliadas por meio dos itens que compõem os testes de proficiência. Já a letra “B”, apresentada na segunda coluna da matriz, identifica os conteúdos de biologia, uma vez que a matriz de Ciências da Natureza engloba também os conteúdos de física (F) e química (Q).

Na seção seguinte, será apresentada a Coordenação de Instrumentos de Avaliação do CAEd e sua atuação nas avaliações em larga escala através da elaboração e montagem dos testes de proficiência utilizados nas avaliações em larga escala.

1.3 O PROCESSO DE ELABORAÇÃO DE ITENS PELO CENTRO DE POLÍTICAS PÚBLICAS E AVALIAÇÃO DA EDUCAÇÃO

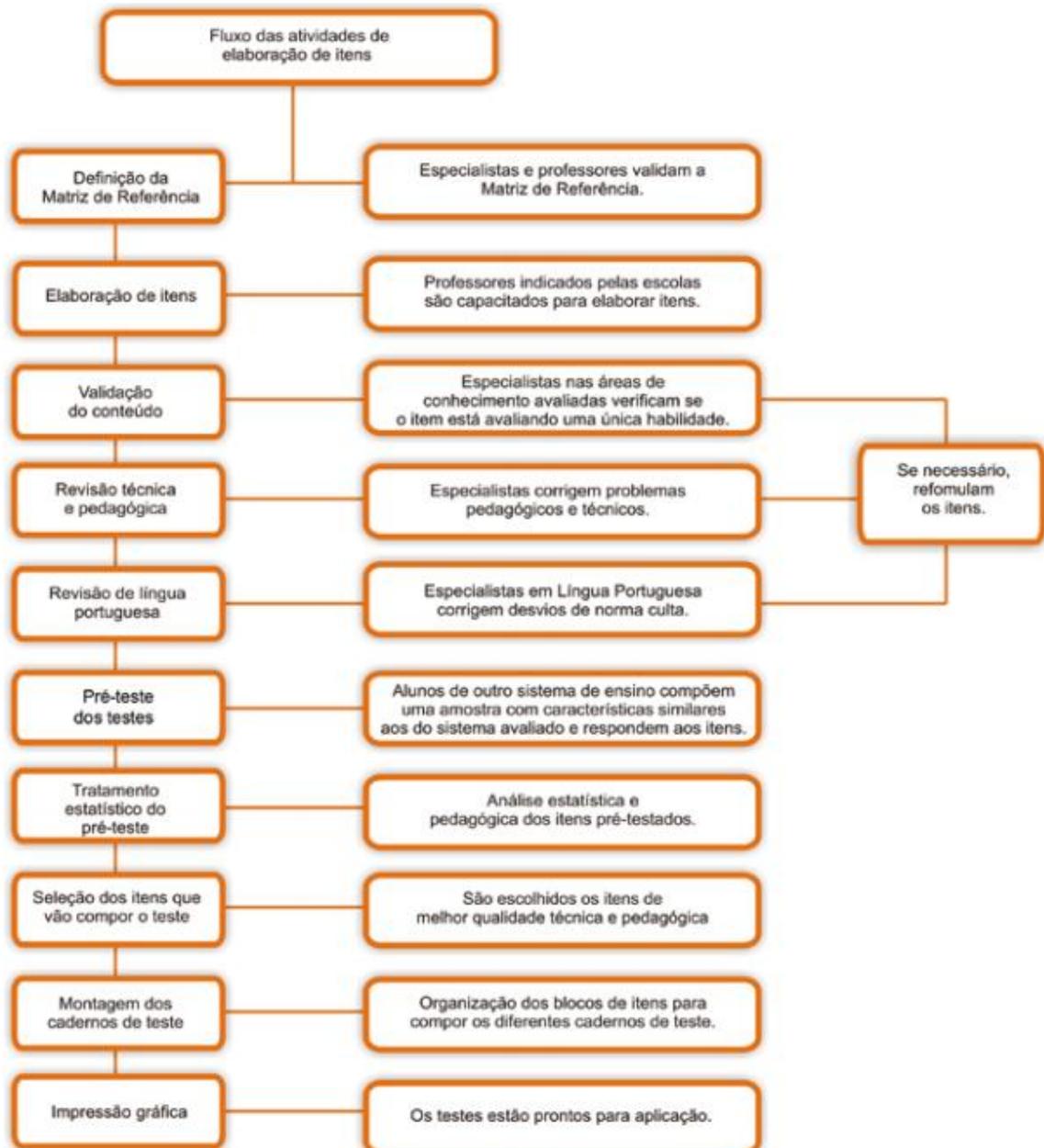
A Coordenação de Instrumentos de Avaliação (CIA) do CAEd é a equipe responsável pela elaboração dos cadernos de testes, definição dos itens que vão compor cada um e o monitoramento do processo de montagem das provas. Esta

coordenação também cuida da montagem das matrizes de referência que vão nortear todo o processo de avaliação. Essa coordenação é composta por laboratórios com profissionais especializados que atuam na elaboração dos itens de diversos projetos. Esses laboratórios são divididos por áreas como as de ciências humanas, matemática, linguagens, alfabetização e ciências da natureza.

No laboratório de Ciências da Natureza atuei, até 2017, como Analista de Instrumentos de Avaliação. Neste cargo, eu coordenava e auxiliava uma equipe composta por outros analistas e auxiliares responsáveis pela elaboração dos itens que compõem os testes de proficiência das áreas de química, física e biologia. A qualidade desses itens está atrelada à sua adequação à Teoria de Resposta ao Item (TRI) e às técnicas específicas, para que assim o teste se torne efetivo nas suas atribuições. Com isso, os itens que não atendem a TRI não são capazes de realizar uma boa aferição das habilidades desenvolvidas pelos estudantes, dificultando a manutenção dos bancos de itens, o que compromete a eficácia dos testes.

A elaboração dos itens produzidos pelo CAEd é norteadada por um documento denominado “Guia de Elaboração de Itens” que apresenta os critérios utilizados na elaboração dos itens de avaliação em larga escala, além das recomendações técnicas e pedagógicas a serem consideradas. Além disso, esse documento também apresenta os critérios para a revisão dos itens elaborados, assim como orientações sobre como proceder na revisão. A figura 5 apresenta as etapas do processo de elaboração de itens.

Figura 5 - Etapas do processo de elaboração de itens



Fonte: CAEd, 2008.

Segundo informações contidas no Guia de Elaboração de Itens (CAEd, 2008), a elaboração dos itens deve ser realizada por profissionais capacitados, com domínio na área de conhecimento a ser avaliada, ou seja, os elaboradores precisam entender o conteúdo escolar como um meio para se desenvolver habilidades e competências. Para isso, é necessário compreender os conteúdos abordados nos PCNs e suas dimensões conceituais de modo a levar o estudante a mobilizar seus recursos cognitivos (CAEd, 2008).

O elaborador precisa entender os processos de desenvolvimento e aprendizagem e saber ajustar a complexidade e o grau de dificuldade dos itens. Além disso, é necessário que esses profissionais tenham domínio da linguagem verbal utilizada nos itens, que devem ser construídos para estudantes de determinadas faixas etárias. Portanto, a linguagem deve expressar o desejado de maneira mais simples possível.

Após a elaboração das matrizes de referência, ocorre a produção dos itens, os quais são criados a partir dos descritores que traduzem as habilidades avaliadas associando os conteúdos curriculares e operações mentais desenvolvidas pelos estudantes. A Figura 6 apresenta um esquema do processo inicial da elaboração dos itens.

Figura 6 - Esquema da elaboração dos itens



Fonte: CAEd, 2008.

O item avalia, portanto, uma única habilidade, indicada pelo descritor da Matriz de Referência e é composto por cinco partes, como apresentado na Figura 7.

Figura 7 - O item e suas partes.

1 — Leia o texto abaixo.

2 — **Curaçao, um simpático e colorido paraíso**

Há uma lenda que explica a razão de Curaçao ser uma ilha tão colorida. Consta que um governador, há muitos anos, sentia dores de cabeça terríveis por todas as construções serem pintadas de branco e refletirem muito a luz do sol. Ele teria então sugerido algo a seus conterrâneos: colocar outras cores nas fachadas de suas residências e comércios; ele mesmo passaria a usar o amarelo em todas as construções que tivessem a ver com o governo. E assim nasceu o colorido dessa simpática e pequena ilha do Caribe.

E quem se importa se a história é mesmo real? Todo o colorido de Punda e Otrobanda combina perfeitamente com os muitos tons de azul que você vai aprender a reconhecer no mar que banha Curaçao, nos de branco, presentes na areia de cada uma das praias de cartão-postal, ou nos verdes do corpo das iguanas, o animal símbolo da ilha.

Acostume-se, aliás, a encontrar bichinhos pela ilha. Sejam grandes como os golfinhos e focas do Seaquarium, os lagartos que vivem livres perto das cavernas Hato, ou os muitos peixes que vão cercar você assim que entrar nas águas da lindíssima praia de Porto Mari. Tudo em Curaçao parece querer dar um "oi" para o visitante assim que o avista.

A ilha, porém, tem mais do que belezas naturais. Descoberta apenas um ano antes do Brasil, Curaçao também teve um histórico [...] que rendeu ao destino uma série de atrações [...], como o museu Kura Hulanda, ou as Cavernas Hato. [...]

Disponível em: <http://zip.net/bhq1CS>. Acesso em: 11 out. 2013. Fragmento. (P070104F5_SUP)

3 — (P070105F5) De acordo com esse texto, qual é o animal símbolo da ilha?

A) A foca.
 B) A iguana.
 C) O golfinho.
 D) O lagarto.

4 —

5 —

Fonte: Espírito Santo, 2015.

O número 1 é o **enunciado** que se caracteriza como um estímulo para que o estudante mobilize recursos cognitivos com o intuito de solucionar o problema apresentado. O estímulo pode conter uma imagem, um texto ou outros recursos que recebem o nome de **suporte** e auxiliam na resolução do problema levantado pelo item. Além disso, pode-se utilizar situações problemas nas quais é realizada uma contextualização com uma situação conhecida pelo estudante apresentando assim os dados e informações para a resolução do item. Nos itens de Ciências da Natureza, os suportes utilizados podem ser gráficos, tabelas, trechos de artigos, esquemas de experimentos, etc.

O número 3 apresenta o **comando de resposta** que pode ser escrito sob a forma de complementação ou interrogação. Esse comando deve estar intimamente ligado ao descritor, além de ser claro e objetivo sobre a tarefa a ser realizada. Os números 4 e 5 mostram as alternativas de resposta, sendo o **gabarito** (5) a resposta

correta e os demais são denominados **distratores** (4) que devem ser plausíveis e estarem associados ao tema do item.

A elaboração dos itens é complexa e envolve uma série de regras, a saber:

- O enunciado deve apresentar todas as informações necessárias para a resolução do item.
- Deve atentar-se às fontes utilizadas para garantir a fidedignidade das informações e garantir que esteja relacionada ao cotidiano dos estudantes. Não é permitido o uso de livros didáticos.
- Os suportes devem apresentar as referências bibliográficas completas.
- Os itens devem ser inéditos.
- Os itens não podem avaliar apenas a capacidade de memorização dos estudantes.
- Devem conter apenas um gabarito.
- Não é permitido o uso de expressões negativas.
- Os enunciados não devem induzir a resposta pelo estudante.
- Os enunciados não podem ser escritos na primeira pessoa.
- As alternativas de resposta devem apresentar paralelismo semântico e sintático.
- Não é permitido alternativas que induzam ao acerto por exclusão.
- Não se deve utilizar alternativas muito longas.

Em 2017, a equipe do laboratório de Ciências da Natureza desenvolveu um projeto de pesquisa, intitulado “Avaliação do Ensino de Ciências na segunda etapa do ensino fundamental no Avalia BH”, com o objetivo de analisar os itens das aplicações do Programa de Avaliação do município de Belo Horizonte. A partir das análises dos itens realizadas nesse projeto, foi possível perceber que alguns desses não atendiam a TRI devido a problemas na matriz de referência do projeto estudado, além da pouca experiência em elaboração de itens dos profissionais que trabalham no laboratório.

O baixo número de itens capazes de produzir uma boa medição das habilidades desenvolvidas pelos estudantes tem dificultado a elaboração e manutenção de um banco de itens, o que compromete a eficiência dos testes de proficiência. Dessa forma, a maior eficiência na produção de itens válidos pela TRI

poderá melhorar a capacidade de produção nos laboratórios de elaboração de itens em Ciências da Natureza da Coordenação de Instrumentos de Avaliação.

Para uma melhor compreensão sobre a produção dos instrumentos utilizados nos Programas de Avaliação, na próxima seção, será apresentada a dinâmica de elaboração dos testes pelo CAEd a partir da elaboração de itens.

1.4 MONTAGEM E APLICAÇÃO DOS TESTES DE PROFICIÊNCIA

Todos os procedimentos técnicos e pedagógicos na construção dos itens, mostrados anteriormente, são importantes para garantir a confiabilidade do item, seu poder avaliativo e a eficiência de um programa de avaliação.

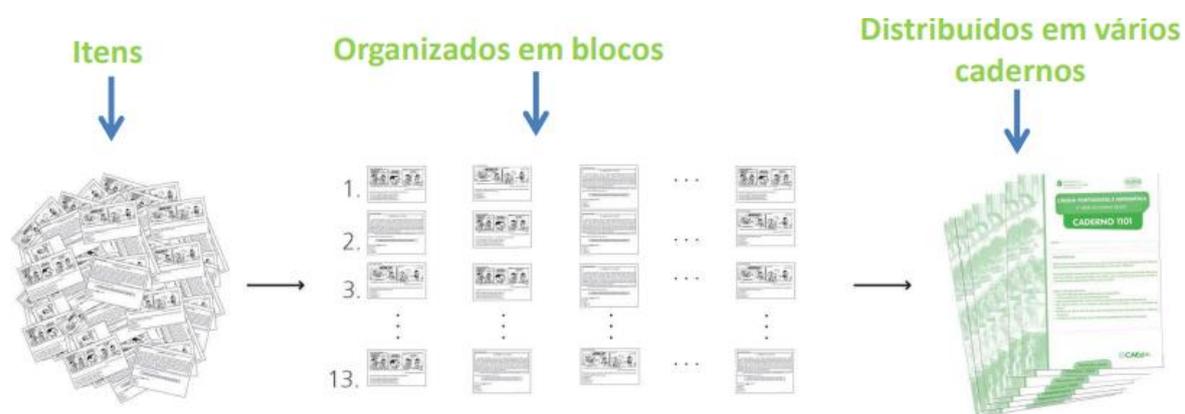
Os itens produzidos passam por uma pré-testagem de caráter amostral que tem como objetivo a calibração dos itens de acordo com a TRI garantindo assim a efetividade da avaliação.

Conforme apontado pelo boletim pedagógico do SADEAM, 2016:

A definição sobre o número de itens é crucial para a composição dos cadernos de teste. Por um lado, o teste deve conter muitos itens, pois um dos objetivos da avaliação em larga escala é medir de forma abrangente as habilidades essenciais à etapa de escolaridade que será avaliada, de forma a garantir a cobertura de toda a Matriz de Referência adotada. Por outro lado, o teste não pode ser longo, pois isso inviabiliza sua resolução pelo aluno. Para solucionar essa dificuldade, é utilizado um tipo de planejamento de testes denominado Blocos Incompletos Balanceados – BIB (CAEd, 2016. p.22).

Os blocos incompletos balanceados (BIB) podem ser configurados como um esquema otimizado para o rodízio de blocos no qual os itens são organizados em blocos, formando um caderno de teste. Nas avaliações em larga escala, os BIB são essenciais para permitir que um grande número de itens seja aplicado a um conjunto de estudantes avaliados, mesmo que cada estudante não responda a muitas questões. Assim, é possível garantir que o teste abarque toda a Matriz de Referência e, com isso, avalie todo o conjunto de habilidades esperadas, conforme Figura 8.

Figura 8 - Esquema de montagem dos testes de proficiência



Fonte: Acervo pessoal da autora.

Na Figura 8, observa-se que na montagem dos testes os itens são organizados em blocos que formam um caderno de teste. Com o uso do BIB, é possível elaborar diferentes cadernos de teste para serem aplicados a estudantes de uma mesma série. Pode-se destacar duas vantagens na utilização desse modelo de montagem de teste: a disponibilização de um maior número de itens em circulação no teste, avaliando, assim, uma maior variedade de habilidades; e o equilíbrio em relação à dificuldade dos cadernos de teste uma vez que os blocos são inseridos em diferentes posições nos cadernos, evitando, dessa forma, que um caderno seja mais difícil que outro (CAEd, 2015).

Após a montagem dos testes, ocorre a implementação das estratégias de aplicação, além de procedimentos adotados para os dias de aplicação e os de contingência diante das ocorrências e situações adversas realizadas no CAEd, pela Coordenação de Operações de Avaliação (COA). Após a aplicação dos testes, é realizada uma série de tarefas, como a receitação dos malotes, organização do material e a elaboração da base de dados consistida com os resultados que são divulgados nos boletins pedagógicos encaminhados para as escolas.

O material produzido pelo CAEd, em parceria com as Secretarias de Educação dos Estados, apresenta os resultados dos testes de proficiência com o objetivo de promover a discussão sobre a apropriação desses resultados nas escolas avaliadas. Esse material é composto por três revistas distintas:

- Revista do Sistema: possui uma visão geral dos resultados ressaltando, entre outras coisas, a importância da apropriação dos resultados e da avaliação como um indicador da qualidade da educação.
- Revista da Gestão: Direcionada à equipe gestora das escolas. Apresenta artigos e estudo de caso relacionado à gestão escolar e informa o resultado das escolas em todas as disciplinas e etapas de escolaridade avaliadas.
- Revista Pedagógica: Direcionada aos professores da disciplina/etapa de escolaridade e à equipe pedagógica da escola. Apresenta informações pedagógicas, estudo de caso relacionados à prática pedagógica dos educadores e textos que tratam de temas da área avaliada.

De posse desses materiais, os profissionais da educação têm a possibilidade de conhecer e se apropriar dos resultados da sua escola, e, com isso, promover intervenções nos projetos pedagógicos, ou seja, conhecendo a realidade de cada escola, é possível reformular as ações para que os estudantes possam avançar nas escalas de proficiência.

No CAEd, esses resultados são utilizados como instrumentos de melhoria da qualidade do processo de elaboração dos itens, pois, a partir das análises estatísticas e pedagógicas, é possível verificar se eles são eficientes e se estão cumprindo o seu papel. Nesse processo, alguns itens podem ser eliminados por motivos como problemas de elaboração (mais de um gabarito, suporte confuso, etc.) e outros podem apresentar comportamentos estatísticos diferentes do esperado, como duas ou mais bisseriais positivas, parâmetro C alto, entre outros, o que pode prejudicar o processo de avaliação. Esses conceitos serão abordados no capítulo 2.

Com o intuito de contextualizar a análise do presente trabalho, na seção seguinte serão apresentados os Programas de Avaliação analisados com a finalidade de conhecer um breve histórico de cada um e suas especificidades.

1.5 OS PROGRAMAS DE AVALIAÇÃO ANALISADOS

Nas subseções seguintes, serão apresentados os Programas de Avaliação que avaliam o Ensino Médio e que serão foco de análise destes estudos. Será realizada a descrição e análise do desenho através da trajetória de cada um dos Programas analisados e, com isso, serão apontadas as escalas e padrões de desempenho correspondentes a cada programa.

1.5.1 SADEAM

O SADEAM representa um dos instrumentos da política educacional do Amazonas no qual são produzidas, regularmente, informações sobre a educação do seu estado. Esse sistema foi criado em 2008 pelo Governo do Estado do Amazonas, por meio da Secretaria de Educação e Qualidade de Ensino (SEDUC/AM). O SADEAM se apresenta como uma avaliação censitária, aplicada anualmente com o objetivo de buscar informações detalhadas dos estudantes da rede pública do ensino do Amazonas, uma vez que a Secretaria se limitava, anteriormente, às informações de desempenho da Prova Brasil e SAEB, que ocorrem de dois em dois anos (VIEIRA, 2015).

A parceria com o CAEd para a realização do SADEAM ocorreu em 2011 com a aplicação de testes de desempenho para os alunos dos 3º e 7º anos do Ensino Fundamental, em Língua Portuguesa e Matemática, incluindo os alunos que estavam cursando os anos iniciais e finais da EJA (5º e 9º anos). No Ensino Médio, avaliou-se, na modalidade regular e EJA, Linguagens, Matemática, Ciências Humanas e Ciências da Natureza (CAEd, 2017).

Esse sistema gera um diagnóstico do desempenho individual e coletivo dos estudantes e oferece subsídios aos profissionais da educação na busca da qualidade da educação básica. A partir desse diagnóstico, esses profissionais podem identificar os fatores associados ao desempenho dos alunos além de criar ações para a melhoria do desenvolvimento educacional do Estado, pois os resultados das avaliações apontam as habilidades não alcançadas que necessitam de atenção e planejamento de intervenção. A tabela 2 apresenta a trajetória do SADEAM e sua consolidação na avaliação.

Tabela 2 - Trajetória do SADEAM

Ano	Rede	Disciplinas	Etapas	Estudantes avaliados
2011	Estadual	LP, Redação, Matemática, CH e Ciências da Natureza.	3º EM, EJA anos iniciais, EJA anos finais e EJA Ensino Médio	91.623
2012	Estadual e Municipal	LP, Redação, Matemática, CH e Ciências da Natureza.	3º Ano da Alfabetização, 5º Ano do EF, EJA - Anos Iniciais, 7º EF, 9º EF, EJA - Anos Finais, 1º EM, 3º EM, EJA - Ensino Médio	201.258
2013	Estadual	LP, Matemática, CH e Ciências da Natureza.	EJA - Anos Iniciais, 7º EF, EJA - Anos Finais, 1º EM, 3º EM, EJA - Ensino Médio.	115.092
2014	Estadual e Municipal	LP, Redação, Matemática, Ciências da Natureza e CH.	5º EF, EJA - Anos Iniciais, 9º EF, EJA - Anos Finais, 3º EM, EJA - Ensino Médio	117.969
2015	Estadual e Municipal	LP, Redação, Matemática, Ciências da Natureza e CH.	3º EF, 7º EF, 1º EM, 3º EM, EJA PRESENCIAL - Anos Iniciais, EJA PRESENCIAL - Ensino Médio	189.084

Fonte: Amazonas, 2015.

A partir dos dados da Tabela 2, é possível observar que as etapas de ensino avaliadas mudam, ou seja, são diferentes a cada edição. Essa característica se dá pelo fato do SADEAM complementar informações do SAEB, pois essas avaliações são realizadas em anos intercalados e estabelecem uma linearidade na interpretação dos dados, já que ambas as avaliações são similares na forma de obtenção dos dados (AssEA/SEDUC/AM, 2015).

Assim como outras avaliações externas, o SADEAM apresenta alguns instrumentos para seu desenvolvimento, entre eles a Matriz de Referência, a Escala de Proficiência, os Padrões de Desempenho e os Níveis de Desempenho.

A Matriz de Referência do SADEAM apresenta um conjunto de habilidades relacionadas às orientações curriculares da rede de ensino. Sendo assim, não consiste em um documento organizado de modo desvinculado do trabalho realizado

pela escola, mas tem a sua origem no próprio Currículo da rede, cujo documento é o mesmo que orienta a elaboração dos Projetos Políticos e Pedagógicos de todas as escolas. As habilidades, apresentadas em cada Matriz de Referência, específicas para cada disciplina e etapa avaliadas, servem como base para a elaboração dos itens que compõem o teste (CAEd, 2017). Os itens têm por objetivo avaliar uma única habilidade apresentada por descritores que compõem a Matriz de Referência de cada disciplina avaliada.

Após a aplicação dos testes, os resultados são produzidos e disponibilizados para a rede e entre os dados gerados, é possível analisar os valores de proficiência dos estudantes.

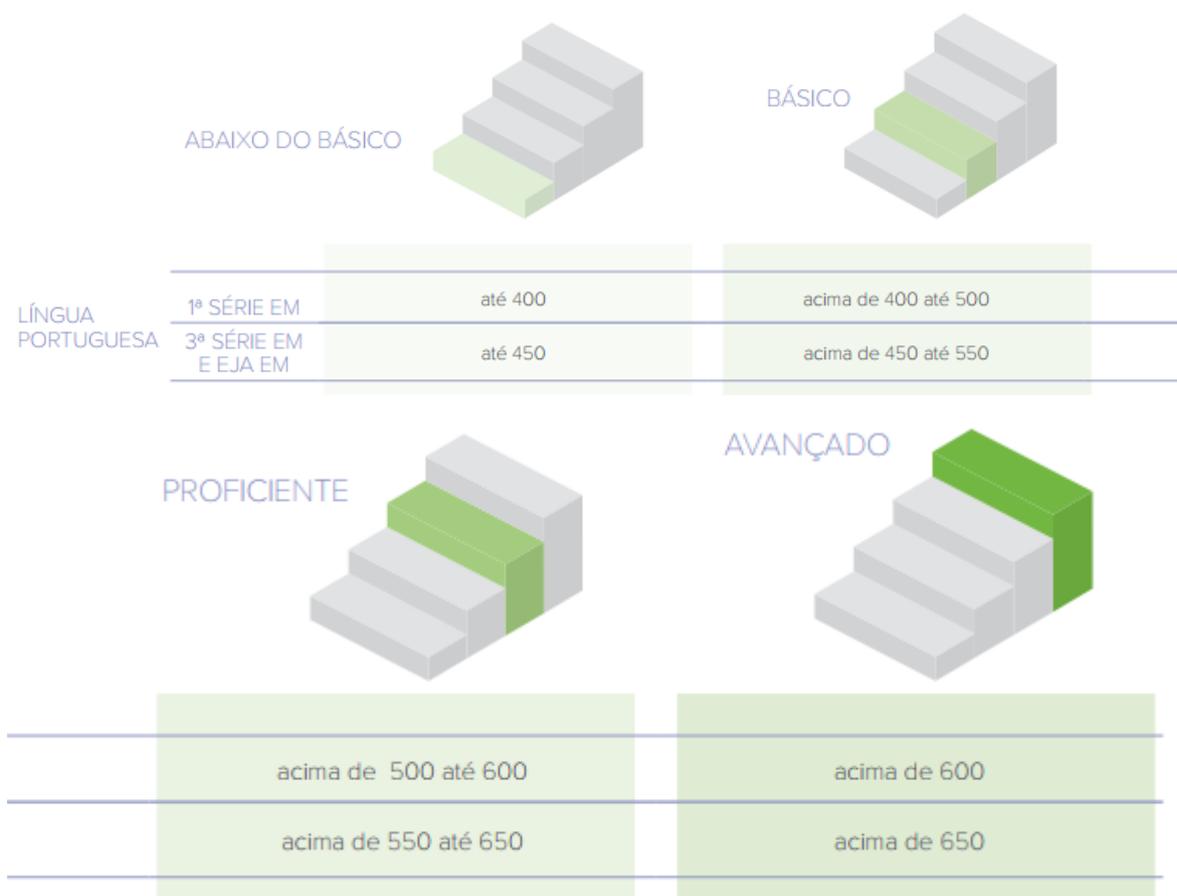
De acordo com as informações do site do SADEAM (2017):

[...] a proficiência é uma medida que representa um determinado traço latente de um aluno. Na avaliação educacional em larga escala, esse traço latente é a aptidão, o conhecimento que um aluno demonstra possuir em relação a determinado conteúdo de uma área do conhecimento que foi avaliada. As diferentes proficiências, desse modo, compõem uma escala numérica, num *continuum*. A escala associa a proficiência ao desempenho (habilidades e competências) alcançado por cada aluno ou por um grupo (turmas, escolas, dentre outros) no teste. Essa escala é identificada, na avaliação externa, como Escala de Proficiência (AMAZONAS, s.d).

As proficiências obtidas pelos estudantes nas avaliações do SADEAM são agrupadas em quatro padrões de desempenho: Abaixo do Básico, Básico, Proficiente e Avançado. Esses padrões representam divisões nas escalas de proficiência e permitem uma análise dos aspectos cognitivos que demarcam os percentuais de alunos situados nos diferentes Níveis de Desempenho.

Outra forma de representar os resultados é através da escala de proficiência que se caracteriza por uma “régua” no qual os resultados de avaliação são apresentados em níveis que são agrupados por Padrões de Desempenho (Figura 9). Sendo assim, a média de proficiência dos estudantes pode ser alocada na descrição dos intervalos da Escala de Proficiência, permitindo a realização de um diagnóstico pedagógico bastante útil (CAEd, 2015).

Figura 9 - Níveis da Escala de Proficiência em Língua Portuguesa do SADEAM



Fonte: Amazonas, 2015.

A Escala de Proficiência de Biologia do SADEAM é de (499,737, 158,516), ou seja, a média das proficiências dos estudantes é de 499,737 e o desvio padrão de 158,516 pontos que permitem a distribuição dos resultados do desempenho dos alunos no período de escolaridade avaliado, revelando assim o desempenho da avaliação.

1.5.2 PAEBES

A Secretaria de Estado da Educação (SEDU) do Espírito Santo possui dois programas de avaliação, o Programa de Avaliação da Educação básica do Espírito Santo (PAEBES) e o Programa de Avaliação da Educação Básica do Espírito Santo dos anos iniciais do Ensino Fundamental (PAEBES ALFA).

O PAEBES foi criado pela SEDU em parceria com o CAEd em 2009 e avalia os estudantes do Ensino Fundamental e Médio das escolas da rede estadual, redes

municipais associadas e escolas particulares participantes, com o objetivo de aferir o nível de desempenho estudantil de cada estudante.

Possuindo um desenho transversal, ou seja, que realiza uma coleta periódica de um determinado recorte da Educação Básica em busca de informações que, ao serem confrontadas, mostram se houve progresso escolar ao longo de um determinado espaço de tempo. O PAEBES avalia anualmente o nível de apropriação dos estudantes em Língua Portuguesa e Matemática (de todas as etapas avaliadas) e, em anos alternados, em Ciências Humanas e Ciências da Natureza (CAEd, 2017).

O PAEBES ALFA, criado também em 2009, tem como objetivo, diagnosticar o desempenho dos estudantes nos anos Iniciais do Ensino Fundamental em Língua Portuguesa (Leitura e Escrita) e Matemática. Diferente do PAEBES, esse programa de avaliação possui um desenho longitudinal, dividida em ciclos conforme as etapas de escolaridade avaliadas, assim, permitem o acompanhamento dos estudantes desde sua entrada no 1º ano do Ensino Fundamental até o fim do ciclo de alfabetização, além da ampliação da avaliação das habilidades de alfabetização no campo da Leitura, Escrita e Matemática (CAEd, 2017). A Tabela 3 apresenta a trajetória do PAEBES de 2008 a 2013.

Tabela 3 - Trajetória do PAEBES

Ano	Disciplinas	Etapas	Estudantes avaliados
2008	LP e Matemática	1ª Série do EM	14.446
2009	LP e Matemática	5º, 9º Anos do EF e 1º ano do EM	106.830
2010	LP e Matemática	5º, 9º Anos do EF e 1º, 3º anos EM	83.471
2011	LP, Matemática e Ciências da Natureza	5º, 9º Anos do EF e 1º, 3º anos do EM	114.254
2012	LP, Matemática e Ciências Humanas	5º, 9º Anos do EF e 2º, 3º anos do EM	106.525
2013	LP, Matemática, Ciências da	5º, 9º Anos do EF e 3º Ano do EM	93.922

	Natureza Ciências (EF)		
2014	LP, Produção de Texto, Matemática e Ciências Humanas	5º, 9º Anos do EF e 3º Ano do EM	99.343
2015	LP, Produção de Texto, Matemática e Ciências da Natureza e Ciências (EF)	5º, 9º Anos do EF e 3º Ano do EM	103.621

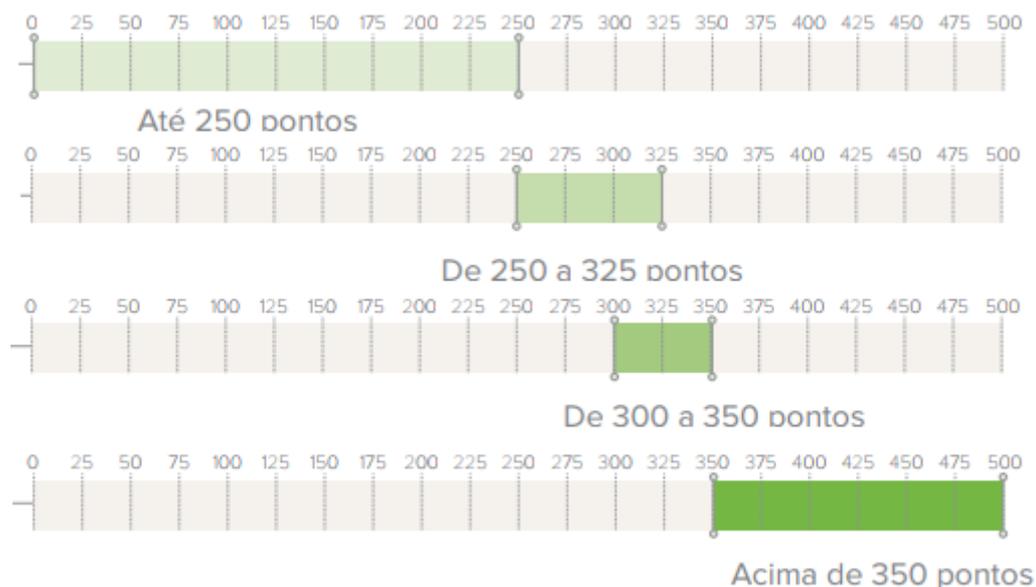
Fonte: Espírito Santo, 2015.

É possível perceber que, desde sua implementação, o PAEBES vem apresentando um aumento das etapas avaliadas e das disciplinas, como a introdução de produção de texto em 2014. O Programa avalia alternadamente as áreas de Ciências da Natureza e Ciências Humanas e vale ressaltar que a aplicação dos testes, em 2012, para o 2º ano do Ensino Médio, teve como objetivo a coleta de dados para a pesquisa interna da SEDU.

Assim como SADEAM, o PAEBES apresenta em seus resultados, os instrumentos para seu desenvolvimento, como a Matriz de Referência e os Padrões de Desempenho e os Níveis de Desempenho, porém, o PAEBES disponibiliza o resultado das médias de proficiência e da distribuição dos estudantes por padrões de desempenho do Estado, por Superintendências (SREs) e por escolas, possibilitando assim uma comparação desses resultados em diferentes escalas.

O Programa organiza os Padrões de Desempenho em Abaixo do básico, básico, proficiente e avançado e seus níveis são apresentados na Figura 10.

Figura 10 - Níveis de Desempenho do PAEBES



Fonte: Espírito Santo, 2015.

A Escala de Proficiência de Biologia do PAEBES é de 250,114, 73,483, ou seja, a média das proficiências dos estudantes é de 250,114 e o desvio padrão de 73,483 pontos que permitem a distribuição dos resultados do desempenho dos alunos, no período de escolaridade avaliado, revelando, assim, o desempenho da avaliação.

1.5.3 AVALIE BA

O sistema de Avaliação Baiano da Educação (SABE) se configura como uma política pública de avaliação que foi implantado em 2007 com o objetivo de avaliar a qualidade do ensino e aprendizagem dos estudantes da Bahia. Além disso, o programa promove um diagnóstico mais preciso do desempenho dos estudantes do estado da Bahia, subsidiando a (re)formulação, o monitoramento de políticas educacionais e a implementação de ações no âmbito escolar (CAEd 2017). Em 2011, o sistema passou a avaliar o Ensino Fundamental e foi dividido em Avalie Alfa e Avalie Ensino Médio.

A parceria com o CAEd, iniciada em 2010, propiciou aplicar e produzir os resultados das avaliações em larga escala no Estado em conjunto com a Secretaria de Educação do Estado da Bahia. A Tabela 4 apresenta a trajetória do Programa de

Avaliação da Bahia ao longo dos anos, o número de estudantes, as disciplinas e as etapas de escolaridade avaliadas.

Tabela 4 - Trajetória do Avalia Ba

Ano	Disciplinas	Etapas	Estudantes Avaliados
2008	Interdisciplinar, com base na Matriz do Enem – 2008.	1º ano do EM	233 unidades escolares
2009	Linguagens, CH, CN e Matemática	2º ano do EM	233 unidades escolares
2010	Linguagens, CH, CN e Matemática.	3º ano do EM	233 unidades escolares
2011	Linguagens, CH, CN e Matemática	1º ano do EM regular e 2º ano da Educação Profissional Integrada ao Ensino Médio	131.316
2012	Linguagens, CH, CN e Matemática	2º ano do EM regular e 3º ano da Educação Profissional	99.070
2013	Linguagens, CH, CN e Matemática	1º e 2º anos do EM e 2º e 3º anos da Educação Profissional e de forma amostral, na 3º ano do EM regular e 4º ano da Educação Profissional Integrada ao Ensino Médio (EPI)	240.624

Fonte: Bahia, 2013.

Assim como os outros Programas de Avaliação já descritos, observa-se que, desde sua implementação, o programa teve um aumento do número de estudantes e das etapas avaliadas, como a inclusão da Educação Profissional Integrada para o Ensino Médio. Esse desenho de avaliação constitui um estudo longitudinal, no qual, a cada ano, é realizada a avaliação de um ano específico do Ensino Médio e da Educação Profissional Integrada ao Ensino Médio, com o objetivo de realizar um acompanhamento completo de uma determinada amostra de estudantes ao longo de uma etapa específica de ensino.

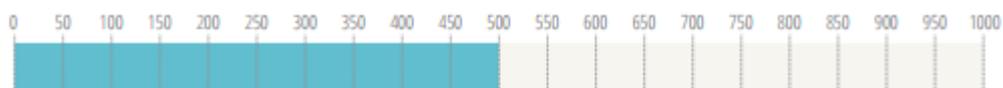
O Avalia Ba utiliza, como instrumentos da Avaliação, a Matriz de Referência, os itens, a escala de proficiência e os Padrões de Desempenho. Nas primeiras edições do Avalie Ensino Médio, os itens foram produzidos de acordo com a Matriz do ENEM e somente a partir de 2010, em parceria com o CAEd, as Matrizes específicas do Programa foram produzidas.

O Avalie Ensino Médio utiliza a TRI para o cálculo da proficiência do estudante, que não depende unicamente do valor absoluto de acertos, já que depende também da dificuldade e da capacidade de discriminação das questões que o estudante acertou e/ou errou. O modelo gera um balanceamento de graus de dificuldade entre as questões que compõem os diferentes cadernos e as habilidades avaliadas em relação ao contexto escolar. Esse balanceamento permite a comparação dos resultados dos estudantes ao longo do tempo e entre diferentes escolas (Revista Pedagógica do Avalia Ba, 2013).

Diferentemente dos outros programas apontados nesse trabalho, o Avalie Ba apresenta níveis de proficiência com intervalos que vão de 0 a 1000 pontos, conforme Figura 11.

Figura 11 - Níveis de Proficiência do Avalia Ba

Até 500 PONTOS



DE 500 ATÉ 550 PONTOS



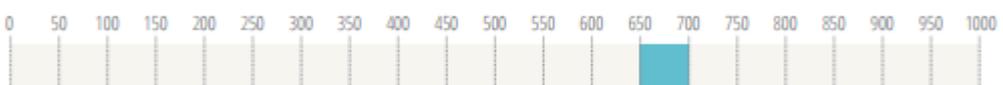
DE 550 ATÉ 600 PONTOS



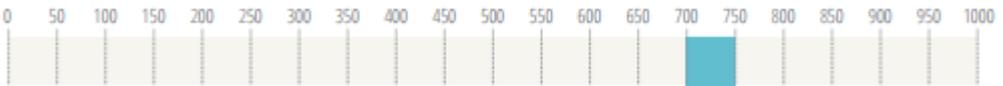
DE 600 ATÉ 650 PONTOS



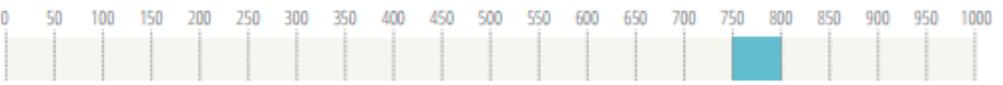
DE 650 ATÉ 700 PONTOS



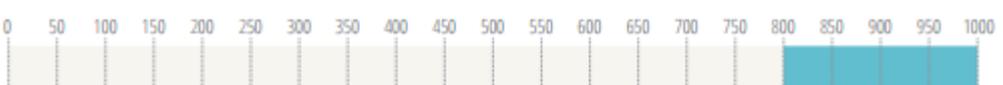
DE 700 ATÉ 750 PONTOS



DE 750 ATÉ 800 PONTOS



ACIMA DE 800 PONTOS



Fonte: Bahia, 2013.

Na escala de proficiência, que apresenta média de 500,124 e desvio padrão de 179,717 pontos, os resultados da avaliação são apresentados em níveis, de modo a conter, na mesma “régua”, a distribuição dos resultados das proficiências dos estudantes em períodos de escolaridade distintos, revelando, assim, o desempenho na avaliação. A média de proficiência obtida deve ser alocada na descrição dos intervalos da Escala de Proficiência no ponto correspondente, permitindo a realização de um diagnóstico pedagógico bastante útil (Revista Pedagógica do Avalia Ba, 2013).

O próximo capítulo tem por objetivo realizar uma análise dos itens dos Programas de Avaliação descritos com o objetivo de identificar o motivo que causa

os comportamentos estatísticos e, a partir disso, realizar uma análise pedagógica desses itens, identificando os fatores que interferem na qualidade dos itens atrelados à elaboração.

2 A ANÁLISE DOS ITENS DE BIOLOGIA

Esse Capítulo tem como objetivo apresentar e discutir os parâmetros dos itens selecionados de três Programas de Avaliação produzidos pelo CAEd e suas respectivas secretarias, com o intuito de identificar os fatores que influenciam no percentual de acerto dos estudantes.

A análise desse processo terá como foco os itens do Sistema de Avaliação Baiano de Educação, do Programa de Avaliação da Educação Básica do Espírito Santo e do Sistema de Avaliação do Desempenho Educacional do Amazonas e será dividido em quatro seções principais. Serão apresentados os itens que serão analisados, bem como seus parâmetros estatísticos.

Esse capítulo apontará também os conceitos da TRI, destacando sua importância, metodologia, aplicação nas avaliações e sua influência na elaboração dos itens. Assim, será necessário levantar os conceitos abordados no Guia de Elaboração de Itens utilizado nos Laboratórios de Instrumentos de Avaliação do CAEd, os relatórios produzidos e suas interpretações. Com isso, espera-se determinar os fatores que influenciam os estudantes durante o teste que refletem nos baixos percentuais de acerto identificados. Assim, será possível realizar intervenções específicas no processo de elaboração de itens com a expectativa de adequá-los à população.

2.1 A APLICAÇÃO DA TEORIA DE RESPOSTA AO ITEM

A Teoria de Resposta ao Item é uma forma de analisar e avaliar os resultados obtidos pelos estudantes nos testes, levando em consideração as habilidades demonstradas e os graus de dificuldade dos itens, permitindo a comparação entre testes realizados em diferentes anos. Esse método sugere formas de representar a probabilidade de um indivíduo dar uma resposta correta a um item em função da sua habilidade e das características do item.

De acordo com Klein (2003), esse método surge da necessidade de superar as limitações da apresentação de resultados somente através de percentuais de acertos ou escores dos testes e ainda da dificuldade de comparar resultados de diferentes testes em diversas situações.

A Teoria de Resposta ao item (TRI) muda o foco de análise da prova como um todo para a análise de cada item. Klein destaca esse método como um conjunto de modelos matemáticos nos quais a probabilidade de resposta a um item é modelada como função da proficiência do aluno e de parâmetros que expressam certas propriedades dos itens.

Segundo Klein,

Uma das propriedades importantes da TRI é o fato dos parâmetros dos itens e a proficiências dos indivíduos serem invariantes. Tanto os parâmetros dos itens obtidos de grupos diferentes de alunos testados quanto os parâmetros de proficiência baseado em grupos diferentes de itens são invariantes, exceto pela escolha de origem e escala (KLEIN, 2003. p. 127).

Essa propriedade é importante para as Avaliações em Larga Escala, pois, associadas a outros métodos estatísticos, permite estimar a distribuição de proficiências da população e monitorar os progressos de um sistema educacional, o que não é possível pela Teoria Clássica dos Testes.

Pasquali & Primi (2003) complementam as ideias de Klein, abordando o contexto histórico e a lógica básica da TRI. Além disso, apresenta seus dois postulados que são:

- O desempenho do sujeito numa tarefa pode ser predito a partir de um conjunto de fatores ou variáveis hipotéticas, ditos aptidões ou traços latentes.
- A relação entre o desempenho e os traços latentes pode ser descrita por uma equação matemática monotônica crescente, chamada de Curva Característica do Item – CCI.

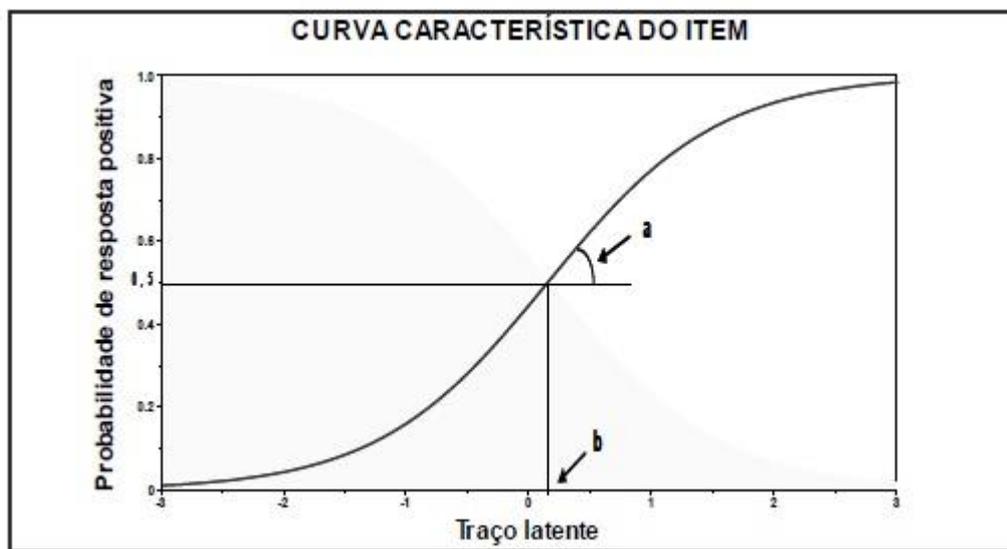
Paralelo a essa ideia, os autores inferem que a TRI não veio para substituir toda a Psicometria Clássica, mas apenas partes dela, particularmente na análise dos itens e no tema da fidedignidade da medida e destaca os avanços dessa Teoria, que são:

- O cálculo do nível de aptidão do sujeito independe da amostra de itens utilizados.
- O cálculo dos parâmetros dos itens (dificuldade e discriminação) independe da amostra de sujeitos utilizada.
- A TRI permite emparelhar itens com a aptidão do sujeito.

- A TRI constitui um modelo que não precisa fazer suposições que aparentam serem improváveis, tais como os erros de medida serem iguais para todos os estudantes;
- A TRI não necessita trabalhar com testes estritamente paralelo.

A representação gráfica do modelo matemático utilizado pela TRI é chamada de curva característica do Item e auxilia no entendimento dos três parâmetros conforme Figura 12.

Figura 12 - Curva Característica do Item



Fonte: SARTES; FORMIGONI, 2013.

Esse gráfico demonstra a relação entre a proficiência e a chance de acerto do item. Sua forma descreve como a mudança no traço latente relaciona-se com a mudança na probabilidade.

De acordo com o modelo de três parâmetros da TRI, as CCIs podem ser especificadas pelo parâmetro de discriminação (a), que possibilita distinguir estudantes de proficiências diferentes, pelo parâmetro de dificuldade (b), que mede a dificuldade de um determinado item e pelo parâmetro do acerto casual (c), que se refere à possibilidade do estudante acertar o item mesmo quando ele tiver um nível de proficiência abaixo do desejado (OLIVEIRA, 2008). Esses conceitos são detalhados na próxima sessão.

2.2 OS PARÂMETROS ESTATÍSTICOS

Para que os testes de proficiência sejam eficazes, são necessários vários procedimentos que viabilizam a avaliação dos itens através de análises estatísticas que buscam levantar dados dos percentuais de acerto, distribuição dos percentuais por opção de resposta, ancoragem do item e parâmetros da Teoria de Resposta ao Item (TRI). A partir desses dados, é possível analisar se aquele item cumpriu seu objetivo de avaliação e qual foi seu comportamento no teste.

Dessa maneira, na TRI, cada item é considerado individualmente, sendo indicados os fatores que afetam a probabilidade de cada item ser respondido de maneira correta. Dessa forma, tem-se o modelo de TRI com três parâmetros, a saber: parâmetro **a**: poder de discriminação, parâmetro **b**: grau de dificuldade e parâmetro **c**: acerto ao acaso (CAEd, 2014).

De acordo com Andrade, Tavares e Valle (2000), o parâmetro **a** se refere à discriminação dos itens, ou seja, a capacidade do item de discriminar os estudantes que desenvolvem habilidades daqueles que não desenvolvem. Espera-se que, para um item com boa qualidade, a porcentagem de acerto seja maior para o grupo com maior proficiência, e quanto maior for a diferença entre as porcentagens de acertos dos dois grupos (com melhor e com pior proficiência), maior será a discriminação do item. Portanto, estudantes que apresentam valores de proficiência baixos terão pequena probabilidade de acertar o item e vice-versa.

O valor parâmetro **a** é dado pela inclinação da Curva Característica do Item em relação ao eixo das abscissas. Portanto, quanto mais inclinada a curva, maior o ângulo formado e, com isso, mais discriminado é o item. Quanto maior o valor desse parâmetro, maior é a separação entre as probabilidades de dois estudantes, com proficiências diferentes, acertarem o item. Ao contrário, se os valores do parâmetro **a** forem muito baixos, estudantes com níveis de proficiência distintos têm a probabilidade próxima de acertar o mesmo item.

De acordo com Rabelo (2013), a classificação apresentada na Tabela 5 pode ser utilizada como referência na determinação de itens discriminativos ou não.

Tabela 5 - Classificação de itens de acordo com a discriminação

VALORES DO PARÂMETRO "a"	DISCRIMINAÇÃO
$a = 0$	Nenhuma
$0 < a \leq 0,0035$	Muito baixa
$0,0035 < a \leq 0,0065$	Baixa
$0,0065 < a \leq 0,0135$	Moderada
$0,0135 < a \leq 0,0170$	Alta
$a > 0,0170$	Muito alta

Fonte: Rabelo (2013, p. 138).

O parâmetro **b** está relacionado ao percentual de estudantes que respondem ao item corretamente. Assim, quanto maior o valor desse parâmetro, mais difícil é o item e maior é o nível de exigência de proficiência para responder corretamente esse item. Esse parâmetro é representado na mesma escala que a proficiência dos estudantes e pode ser considerado o nível de habilidade mínima necessária para um indivíduo ter uma probabilidade alta de acertar o item. Esse parâmetro corresponde ao ponto, na reta, perpendicular ao eixo das abscissas.

Segundo Rabelo (2013), a tabela a seguir mostra a distribuição e a classificação dos itens adotados por grande parte dos autores para testes de avaliação educacional em uma escala de média 500 e desvio padrão 100.

Tabela 6 - Classificação dos itens de acordo com o parâmetro b

CLASSIFICAÇÃO	DISTRIBUIÇÃO ESPERADA	DIFICULDADE DO ITEM
Muito fáceis	10%	Até 372
Fáceis	20%	De 373 a 448
Medianos	40%	De 449 a 551
Díficeis	20%	De 552 a 627
Muito Díficeis	10%	628 em diante

Fonte: Rabelo (2013, p. 134).

O ideal, portanto, é que, nas avaliações em larga escala, a distribuição dos níveis de dificuldade dos itens seja de 10% para níveis das faixas mais extremas, 20% em cada uma das faixas extremas e 40% dos itens na faixa média.

Já o parâmetro **c** leva em consideração a probabilidade do acerto ao acaso, ou seja, a probabilidade de o estudante "chutar" e acertar o item. Nos itens de ensino

médio elaborados, que possuem 5 alternativas, o parâmetro c considerado ideal pode ser, no máximo, 0,20, pois corresponde à probabilidade de acertar uma das alternativas. Se o acerto ao acaso for superior a este valor, pode ser que o gabarito seja atrativo para estudantes de baixa proficiência, essa característica sugere que os itens possam ter sido mal elaborados ou tiveram suas alternativas de resposta mal elaboradas.

A partir desses parâmetros psicométricos estimados dos itens, é possível construir uma escala de proficiência, que é utilizada na interpretação pedagógica das habilidades de um indivíduo que pode ser posicionado nela de acordo com sua proficiência. Segundo Oliveira,

Uma boa análise de uma escala de proficiência deve oferecer informações relevantes sobre o desenvolvimento cognitivo dos alunos, descrevendo, por meio de resultados apresentados através de escores de proficiência, níveis distintos de ações e operações mentais utilizadas pelos alunos ao resolverem os desafios cognitivos expressos nos itens dos testes. A interpretação da escala deve apresentar, portanto, as habilidades desenvolvidas e consolidadas pelos alunos, como características de etapas próprias de seu desenvolvimento cognitivo, dentro de cada área do conhecimento avaliada (OLIVEIRA, 2008, p.19).

O CAEd utiliza os modelos de três parâmetros da TRI (A, B e C) mencionados anteriormente, pois este é o mais utilizado no Brasil nas avaliações educacionais em larga escala, sendo inclusive o modelo adotado pelo SAEB.

2.3 METODOLOGIA DE PESQUISA

Este trabalho é caracterizado por um estudo de caso, pois visa conhecer uma entidade bem definida como uma pessoa, uma instituição, um curso, uma disciplina, um sistema educativo, uma política ou qualquer outra unidade social, essa investigação assume-se um caso particular (PONTE, 2006).

Segundo Schramm (1971), o objetivo do estudo de caso é tentar esclarecer uma decisão ou um conjunto de decisões, além de explicar o motivo pelo qual foram tomadas, como foram implementadas e quais foram seus resultados. Portanto, a utilização do estudo de caso como estratégia de pesquisa compreende um método abrangente que incorpora abordagens específicas à coleta e análise de dados. Além disso, essa metodologia permite ao investigador um aprofundamento em relação ao fenômeno estudado, favorecendo uma visão holística sobre os acontecimentos da

vida real, destacando-se seu caráter de investigação empírica de fenômenos contemporâneos (Yin, 2005).

A metodologia da pesquisa utilizada pode ser definida como qualitativa que, segundo Gerhardt e Silveira (2009),

[...] não se preocupa com representatividade numérica, mas, sim, com o aprofundamento da compreensão de um grupo social, de uma organização, etc. [...] Os pesquisadores que utilizam os métodos qualitativos buscam explicar o porquê das coisas, exprimindo o que convém ser feito, mas não quantificam os valores e as trocas simbólicas nem se submetem à prova de fatos, pois os dados analisados são não-métricos (suscitados e de interação) e se valem de diferentes abordagens. Na pesquisa qualitativa, o cientista é ao mesmo tempo o sujeito e o objeto de suas pesquisas. O desenvolvimento da pesquisa é imprevisível. O conhecimento do pesquisador é parcial e limitado (GERHARDT; SILVEIRA, 2009. p 32).

Nesse sentido, a abordagem qualitativa foi escolhida visando compreender melhor o sistema de elaboração de itens do CAEd e identificar os problemas relacionados a essa elaboração.

Quanto ao tipo de pesquisa, trata-se de uma análise documental baseada na metodologia de Yin (2005) no qual o material coletado e analisado é utilizado para corroborar e valorizar as evidências de outras fontes e/ou acrescentar informações.

Essa pesquisa foi pautada nos dados estatísticos e os respectivos itens das aplicações dos testes de proficiências produzidos pelo CAEd. A busca desses dados teve início na Coordenação de Instrumentos de Avaliação (CIA) que realizou o levantamento da documentação com a colaboração da equipe de estatística. Porém, para a liberação e acesso a esses dados, foi necessário realizar uma solicitação junto ao Departamento Jurídico do CAEd, além de assinar termos de confidencialidade, pois as imagens dos itens presentes nos relatórios são sigilosas e não podem ser publicizadas.

Nesse sentido, para conseguir expor a análise dos itens, foi realizado uma adequação na abordagem desses, criando itens semelhantes em aspectos de elaboração, ou seja, que abordassem os mesmos descritores, temas, tipos de suportes, comandos e tipos de alternativas. Essa adequação preserva a imagem dos itens aplicados e torna possível a visualização dos elementos para a realização de uma análise pedagógica detalhada do item.

2.3.1 Instrumentos Metodológicos

Os itens serão analisados por meio dos parâmetros estatísticos presentes nos relatórios produzidos pelo CAEd dos itens de Biologia do SADEAM, PAEBES e AVALIA BA.

Meu primeiro foco de pesquisa será voltado para a correlação bisserial que se trata de uma medida estatística que mede a correlação do resultado de um item em particular do teste com o resultado desse, sendo, portanto, uma medida de discriminação do item em relação ao teste. Através dessa análise, será possível identificar se os estudantes proficientes na habilidade avaliada no item estão de fato acertando, sendo atraídos por um ou outro distrator.

Posteriormente, serão analisados os parâmetros **a** que se refere à discriminação do item, o **b** relacionado ao grau de dificuldade e o **c** baseado no acerto ao acaso dos itens. Através dessas análises, será possível identificar os fatores que interferem na qualidade dos itens atrelados à elaboração, que podem ser responsáveis pela produção desses dados, ou caso não sejam identificados os fatores, as possíveis relações com os fatores externos e internos que influenciam o desempenho escolar.

A Teoria Clássica dos Testes (TCT) gera resultados que dependem do conjunto particular de questões que compõem o teste e dos indivíduos que o fizeram, ou seja, as análises e interpretações estão sempre associadas ao teste como um todo e ao grupo de indivíduos (KLEIN, 2003).

A análise clássica dos itens é focada no escore obtido no teste. Assim, um estudante que responde a uma série de itens e recebe um ponto por cada item respondido corretamente, obtém, ao final, um escore total (que é a soma destes pontos). Essa análise apresenta índices de percentuais de acerto que mostram a porcentagem de estudantes que acertaram aquele item.

O Coeficiente bisserial é uma medida que associa o desempenho do estudante em um item e o desempenho desse mesmo estudante na prova. Para um item de boa qualidade é necessário que apenas a alternativa correta apresente valores positivos para a bisserial. Caso alguma alternativa, além do gabarito, apresente valores positivos, significa que alunos com escores elevados estão sendo atraídos para essa alternativa, o que pode indicar um problema na formulação dos itens.

Os relatórios gerados pelo CAEd também disponibilizam a ancoragem que indica o posicionamento do item na escala de proficiência. Essa ancoragem corresponde à proficiência necessária para uma probabilidade de acerto de 65%, além dos percentuais de marcação por opção de resposta e das bisseriais de cada alternativa.

2.3.2 Os Relatórios estatísticos

A Coordenação de Instrumentos de Avaliação gera relatórios que apresentam o comportamento estatístico de cada item dos testes de proficiência aplicados. A Figura 13 apresenta um modelo do relatório de um dos itens de um teste.

Figura 13 - Modelo do relatório de um dos itens na aplicação de um teste

Item:		Descritor: D49		Gabarito: A		Suporte: NENHUM		Bloco: 1		Pos.: 5		
Parametro da TRI			Análise Clássica			Percentuais por opção de resposta					Curva	
A	B	C	Perc. de Acerto	Bisserial	Ancoragem	B/N	A	B	C	D	E	
0,0116	345,9	0,10723	0,317	0,505	368,149	0,004	0,317	0,174	0,171	0,116	0,219	
OBS:						Biss A	Biss B	Biss C	Biss D	Biss E		
						0,505	-0,036	-0,189	-0,257	-0,231		

Fonte: Acervo pessoal da autora.

Esses relatórios apresentam vários dados que permitem analisar o comportamento dos itens nos testes de proficiência. Na primeira linha, é possível identificar o descritor utilizado na elaboração do item, o gabarito, que indica a letra correspondente a alternativa correta, o número do bloco de itens e a posição do item nesse bloco. A terceira e a quarta linhas apresentam os dados referentes aos três parâmetros da TRI, o percentual de acerto daquele item e a correlação bisserial que caracteriza uma medida de discriminação do item em relação ao teste. Além disso, mostra o ponto da escala que aquele item foi alocado, representando pelo ponto de ancoragem e a distribuição dos percentuais de acerto por alternativa. Por fim, a quinta linha apresenta os valores das bisseriais por alternativa e no lado direito estão disponíveis a representação gráfica da Curva Característica do Item e a Curva de Informação do Teste.

2.3.3 Análise dos itens

Para atingir o objetivo do trabalho, da compreensão dos fatores que interferem na qualidade dos itens de biologia nas avaliações do SADEAM, PAEBES e AVALIA BA, inicialmente foi necessário identificar o comportamento dos itens em relação ao percentual de acerto e valores do parâmetro *b*. Para isso, os itens de todas as aplicações foram classificados de acordo com o percentual de acerto, cuja metodologia é citada por Rabelo (2013). A Tabela 7 apresenta a classificação desses itens.

Tabela 7 - Percentual de acerto

	Total válidos	Total efetivo	Até 20%	20 a 50%	50 a 70%	70 a 100%
SADEAM	496	393	101 (26,1%)	241 (61,2%)	41 (10,4%)	9 (2,3%)
PAEBES	279	260	45 (17,3%)	162 (62,3%)	45 (17,3%)	8 (3,1%)
AVALIE BA	316	264	36 (13,7%)	191 (72,3%)	32 (12,1%)	5 (1,9%)

CAEd 2017.

Os valores apresentados por total válidos se referem ao número total de itens aplicados nos testes de proficiências e o total efetivo corresponde ao número total de itens que não foram eliminados nos testes.

Através desses dados é possível observar que a maior parte dos itens levantados apresentou um percentual de acerto menor que 50%, e o maior número em todos os Programas enquadram-se na faixa de 20 a 50%.

A partir desse panorama, os 182 itens que apresentaram os menores índices de acerto (até 20%) foram classificados de acordo com o tema geral utilizado em sua elaboração sendo, assim, possível determinar um conteúdo específico de biologia, conforme apresentado na Tabela 8.

Tabela 8 - Temas gerais dos itens com níveis de acerto de até 20%

Tema Geral	Número de Itens	Percentual
Biotecnologia	11	6,1%
Citologia	61	33,5%
Corpo Humano	4	2,2%
Ecologia	27	14,8

Embriologia	9	4,9%
Energia	8	4,4%
Evolução	20	11%
Genética	11	6,1%
Seres vivos	15	8,2%
Nutrientes	16	8,8%
Total	182	

Fonte: Acervo pessoal da autora.

Os itens que avaliam os conhecimentos de citologia correspondem a 33,5% do total de itens com até 20% de acerto, enquanto os itens de ecologia 14,8% e evolução 11%. Esses dados sugerem que os estudantes apresentam mais dificuldade nos itens de citologia ou que esses itens apresentam falhas em sua elaboração ocasionando, assim, o comportamento estatístico observado.

Nesse sentido, os itens, correspondentes aos descritores desse tema, foram analisados em todos os intervalos dos percentuais de acerto na busca de características na elaboração que justificassem esses dados. Para isso, foi necessário, inicialmente, realizar um levantamento dos descritores utilizados na produção dos itens de citologia conforme Tabela 9.

Tabela 9 - Descritores utilizados na produção dos itens analisados

SADEAM	
D23	Analisar os modelos das estruturas do DNA e RNA e a sua participação na síntese proteica.
D31	Comparar a organização e o funcionamento dos diferentes tipos celulares.
D32	Reconhecer as estruturas e organelas celulares e suas funções.
D33	Identificar a natureza química do DNA e do RNA.
D34	Identificar características das etapas do processo de síntese proteica.
D35	Reconhecer os processos de divisão celular a partir de gráficos, desenhos e textos.
D48	Interpretar, em diferentes formas de linguagem, os processos de síntese proteica e divisão celular.
PAEBES	
D63	Reconhecer as estruturas e as organelas que compõem as células e sua importância.
D55	Identificar a importância das organelas e do núcleo para o metabolismo celular.

D57	Reconhecer os processos de divisão celular.
D54	Identificar os componentes bioquímicos da célula e suas principais funções.
AVALIA BA	
D84	Identificar a natureza química do DNA e do RNA.
D81	Comparar a organização e o funcionamento dos diferentes tipos celulares.
D36	Reconhecer os processos de divisão celular a partir de gráficos, desenhos e textos.
D32	Identificar características das etapas do processo de síntese proteica.

Fonte: BAHIA; AMAZONAS; ESPÍRITO SANTO, 2017.

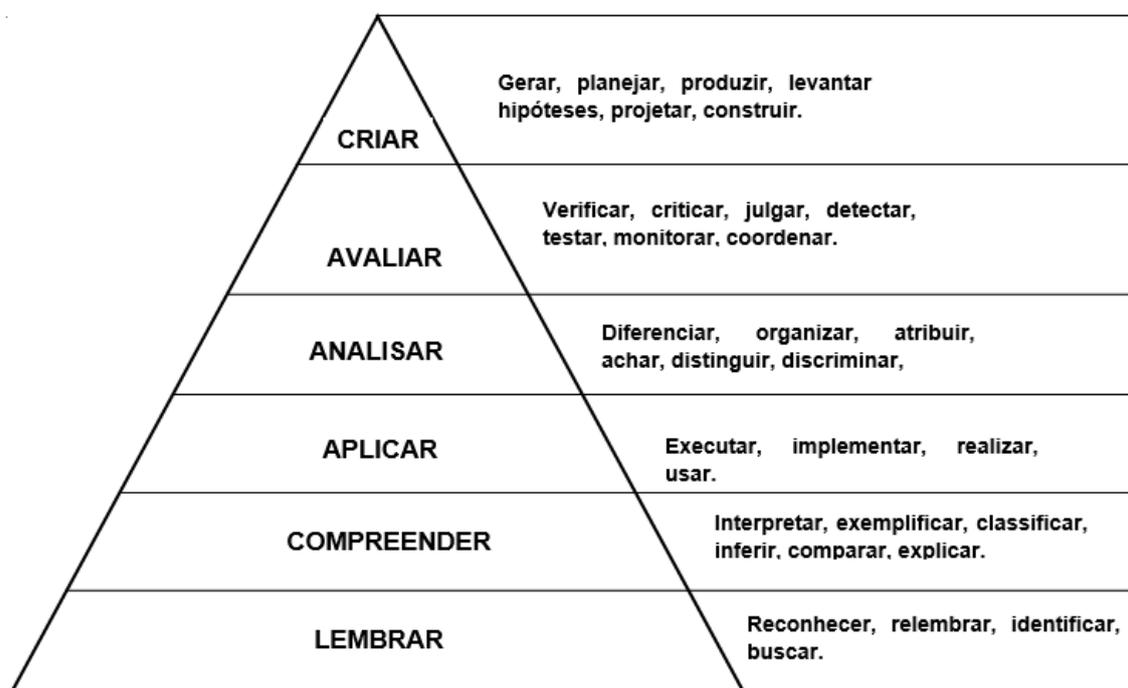
Para analisar os descritores foi utilizada a taxonomia de Bloom, cuja metodologia nos permite identificar o verbo mais adequado na elaboração dos itens em função do comportamento esperado, organizando os objetivos de aprendizagem em níveis crescentes.

A Taxonomia de Bloom foi criada em 1948 por um grupo de psicólogos que buscavam facilitar a comunicação entre examinadores e estimular a pesquisa sobre avaliação através da articulação entre os envolvidos nesse processo. Assim, realizaram uma classificação de objetivos educacionais como ponto de partida para o planejamento educacional divididos em três domínios: cognitivo, afetivo e psicomotor.

Segundo Bloom *et. al* (1983), os objetivos educacionais do domínio cognitivo focam a aprendizagem de conhecimentos, desde a recordação e compreensão de algo estudado até a capacidade de aplicar, analisar e reorganizar a aprendizagem de um modo singular e criativo, reordenando o material ou combinando-o com ideias ou métodos anteriormente aprendidos. Já no domínio afetivo, os objetivos dão ênfase aos sentimentos, emoções, aceitação ou rejeição de algo. No domínio psicomotor, os objetivos educacionais são ligados à habilidade motora, manipulação de objetos ou ações que requerem coordenação neuromuscular; são, geralmente, relacionados à caligrafia, à arte mecânica, à educação física e a cursos técnicos.

Esta pesquisa aborda a Taxonomia de Bloom do domínio Cognitivo, que é estruturada em níveis de complexidade crescente, mostrando assim que, para adquirir uma nova habilidade pertencente ao próximo nível, um indivíduo precisa ter dominado e adquirido a habilidade do nível anterior conforme mostra a Figura 14.

Figura 14 - Níveis de Complexidade da Taxonomia de Bloom



Fonte: Bloom, 1983.

Os verbos utilizados, na maioria dos descritores selecionados nessa pesquisa, como “reconhecer”, “identificar” e “comparar”, ancoram-se nas categorias mais elementares, enquanto apenas um descritor apresenta o verbo “analisar” que perpassa por um nível maior de complexidade. Partindo desse princípio, para a resolução dos itens elaborados a partir desses descritores, os estudantes devem reconhecer, recordar e reproduzir informações, o que significa buscar informações memorizadas.

A Figura 15 apresenta um item semelhante¹ ao aplicado no teste do SADEAM do 1º ano do Ensino Médio que registrou um percentual de acerto de 16%.

¹ Para analisar os itens de acordo com os critérios da TRI e à luz da Taxonomia de Bloom, produzi itens semelhantes aos aplicados no teste, uma vez que esses são sigilosos e não podem ser publicados. Os itens apresentados são “espelhos” dos itens do teste, ou seja, apresentem temas, suportes, alternativas semelhantes aos itens originais e que não afetam sua análise nesta pesquisa.

Figura 15 – Item espelho do SADEAM

Exemplo 1 - A imagem abaixo apresenta uma organela celular.



Qual a função dessa organela ?

- A) Absorver energia luminosa
- B) Armazenar substâncias.
- C) Realizar a respiração celular.
- D) Sintetizar cílios e flagelos.
- E) Transportar proteínas.

Fonte: Acervo pessoal da autora.

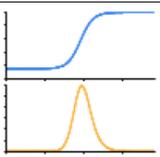
Para responder corretamente a esse item, de acordo com seu descritor, os estudantes deveriam reconhecer as estruturas e organelas celulares e suas funções. O verbo reconhecer, segundo a taxonomia de Bloom, ancora a categoria mais elementar da pirâmide e busca do estudante a habilidade de localizar o conhecimento na memória de longa duração que seja coerente com o material apresentado.

No item acima, os estudantes deveriam associar a organela apresentada na imagem à sua função, que corresponde a alternativa “C”. Porém, o suporte desse item (imagem) se torna meramente ilustrativo, pois, junto a organela, está escrito o seu nome. De acordo com o Guia de Elaboração de Itens (CAEd s/d), o suporte de um item deve ter a função de motivar ou compor uma situação-problema formulada, fazendo com que o estudante mobilize um recurso cognitivo para chegar a resposta, o que não acontece nesse caso, pois exige do estudante apenas a memorização de uma característica.

Quando isso acontece, o item perde sua qualidade em avaliar o conhecimento, pois não é possível identificar quais habilidades foram consolidadas ou quais processos foram utilizados pelos estudantes para responder.

Os dados estatísticos² mostrados na Figura 16 sugerem que os estudantes apresentaram dificuldades na resolução desse item, pois o parâmetro **b**, com valor de 954,638 está elevado. Além disso, esse item não discrimina satisfatoriamente os estudantes, ou seja, não diferencia os estudantes que são proficientes nessa habilidade dos que não são, o que pode ser comprovado pelo parâmetro **a** levantado.

Figura 16 - Relatório estatístico do Item

Item: Descritor: D32 Gabarito: C Suporte: NENHUM Bloco: 7 Pos.: 9												
Parametro da TRI			Análise Clássica			Percentuais por opção de resposta					Curva	
A	B	C	Perc. de Acerto	Bisserial	Ancoragem	B/N	A	B	C	D	E	
0,00739	954,638 3	0,14684	0,16	0,254	983,531	0,003	0,188	0,295	0,16	0,071	0,282	
OBS:							Biss A	Biss B	Biss C	Biss D	Biss E	
							-0,145	0,027	0,254	-0,097	-0,046	

Fonte: Acervo pessoal da autora.

É possível perceber também, que apesar do gabarito ser a alternativa “C” o maior percentual de marcações foi no distrator “B” o que pode indicar que os estudantes não encontram meios de desenvolver o item e acabam buscando, nas alternativas, termos mais comuns no seu cotidiano.

Para evitar esse comportamento estatístico, é possível, através do mesmo descritor, elaborar um item a partir de uma situação problema como mostrado no exemplo 2 (elaborado pela autora).

Exemplo 2 - As Olimpíadas são marcadas pelos sucessivos recordes alcançados em todas as áreas. O aumento gradativo do rendimento dos atletas mostra claramente um maior preparo físico que está ligado à ciência e à tecnologia. Esses recursos têm sido importantes aliados na melhoria do desempenho que está atrelado ao papel das mitocôndrias no organismo.

Essa organela tem como função

² Os dados estatísticos apresentados se referem ao item original. Nos itens espelhos foi mantido mesmo gabarito e ordem das alternativas para não comprometer as análises.

- A) Absorver energia luminosa
- B) Armazenar substâncias.
- C) Realizar a respiração celular.
- D) Sintetizar cílios e flagelos.
- E) Transportar proteínas.

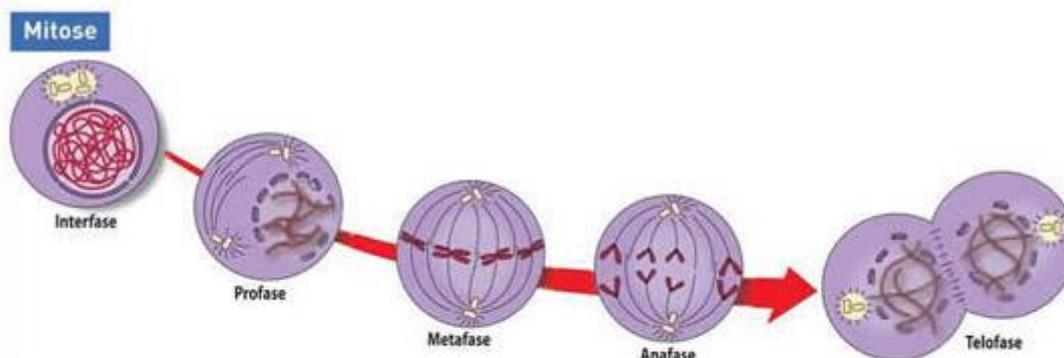
Esse exemplo de item utiliza como suporte uma situação-problema contextualizada sobre um tema bastante conhecido que são as Olimpíadas. A contextualização se propõe a situar e relacionar os conteúdos escolares a diferentes contextos de sua produção, apropriação e utilização (KATO; KAWASAKI, 2011). É através dela que se criam situações, condições, problematizações e enredos capazes de mensurar competências.

Para responder o item é necessário que o estudante consulte e faça uma interpretação dos elementos contidos no suporte. A utilização de fatos e acontecimentos reais propõem ao estudante se colocar na situação-problema, refletir sobre ela e tomar uma decisão. Além disso, para chegar à resposta, é necessário que haja uma mobilização do recurso cognitivo a partir das informações contidas no suporte, fazendo com que o estudante pense numa resposta antes mesmo de ler as alternativas, evitando assim o chute.

Um descritor sobre citologia, comum aos três programas de avaliação, aborda os conceitos de divisão celular. A Figura 17 ilustra um exemplo de um item desse descritor em uma aplicação do PAEBES.

Figura 17 – Item espelho do PAEBES

EXEMPLO 3 – Observe a imagem abaixo.



Nessa imagem a formação do fuso ocorre em qual fase?

- A) Anáfase.
- B) Interfase.
- C) Metáfase.
- D) Prófase.
- E) Telófase.

Fonte: Acervo pessoal da autora.

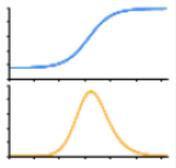
Para responder corretamente a esse item, de acordo com seu descritor, os estudantes deveriam reconhecer os processos de divisão celular a partir de gráficos, desenhos e textos. Como o exemplo do item anterior, o verbo reconhecer ancora uma categoria elementar da pirâmide.

Nesse item, os estudantes deveriam associar as etapas de divisão celular na mitose, identificar a etapa na qual ocorre a formação de fuso para marcar a alternativa correta representada pela letra “D”.

O suporte desse item também pode ser considerado um suporte dispensável, uma vez que não realiza a função motivadora e de levar o estudante a mobilizar o recurso cognitivo para chegar à resposta. Para responder esse item, os estudantes deveriam memorizar as etapas das divisões celulares com seus principais eventos. Deve-se considerar também que os termos apresentados no item são complexos aos estudantes o que pode acarretar em “múltiplos chutes”.

Os dados estatísticos, mostrados na Figura 18, revelam o percentual de acerto do item de 17,4%, indicando dificuldades na resolução do item pelos estudantes.

Figura 18 - Estatísticas do Item do PAEBES

Item:		Descritor: D57		Gabarito: D		Suporte: NENHUM		Bloco: 5		Pos.: 4		
Parametro da TRI			Análise Clássica			Percentuais por opção de resposta					Curva	
A	B	C	Perc. de Acerto	Bisserial	Ancoragem	B/N	A	B	C	D	E	
0,00656	624	0,15586	0,174	0,263	654,926	0,007	0,137	0,174	0,26	0,174	0,248	
OBS:						Biss A	Biss B	Biss C	Biss D	Biss E		
						0	0	0	0	0		

Fonte: Acervo pessoal da autora.

O parâmetro **a**, segundo Rabelo (2013), está baixo e, com isso, o item não discrimina os estudantes proficientes na habilidade daqueles que não alcançam a habilidade. Além disso, é possível perceber que o gabarito não apresenta o maior percentual de marcações de resposta, indicando uma ineficiência na avaliação dessa habilidade por esse item.

Uma forma de avaliar essa habilidade é apresentar um suporte ou situação problema que relacione a uma temática conhecida dos estudantes e que ofereça informações capazes de instruir e criar caminhos para a resolução do problema apresentado, como o exemplo 4 (elaborado pela autora).

Exemplo 4 – A Síndrome de Down é uma alteração genética caracterizada pela presença de um cromossomo extra nas células de um indivíduo. Porém, alguns indivíduos podem apresentar características específicas dessa alteração sem ter um comprometimento do sistema nervoso. Isso ocorre devido à presença de tecidos que apresentam células com o número normal de cromossomos e outras células com um cromossomo a mais em um dos pares (trissomia). Esse fato se dá devido a uma falha no mecanismo de divisão celular.

Essa falha ocorre devido

- A) à ausência de condensação na intérfase.
- B) à migração dos centrômeros para um só lado.
- C) à não disjunção dos homólogos na anáfase.
- D) ao fuso acromático na prófase.
- E) ao pareamento incorreto na metáfase.

Esse exemplo de item utiliza uma situação problema sobre uma alteração genética, possivelmente conhecida pelos estudantes, sem aprofundar em conceitos de genética, o que não aumenta a dificuldade do item. Assim, é possível associar a presença de um cromossomo a mais nas células pela não separação deles durante a divisão celular. Nesse sentido, o item leva o estudante a associar a etapa da divisão celular em que ocorre a separação dos cromossomos com a alteração causada, criando caminhos para a resolução do problema.

Nesse sentido, a contextualização, através de uma situação problematizadora, é importante para possibilitar que os estudantes também busquem soluções em seus próprios conhecimentos, como defende Teixeira (2012) que

[...] não interessa que os alunos apenas memorizem os mais variados conhecimentos nas mais variadas áreas, mas também, que os saibam utilizar nas mais diversas situações. Para isso, os conteúdos trabalhados deverão englobar questões sociais e problemas do cotidiano dos alunos. A contextualização tem, aqui, um papel indispensável. Os alunos só serão capazes de assimilar os conhecimentos trabalhados nas aulas corretamente se lhes atribuírem significado, se posteriormente forem capazes de reutilizar os conteúdos. Só assim estaremos a preparar os alunos para um futuro ativo e participativo pois ao realizar uma recontextualização do saber, o professor estará a procurar situações que permitam atribuir sentido aos conhecimentos a ser ensinados, possibilitando uma reconstrução do conhecimento que esteja ao alcance dos seus destinatários (TEIXEIRA, 2012).

Um outro fator que pode estar relacionado ao aumento da dificuldade desse item e conseqüente diminuição no percentual de acerto é a terminologia utilizada na elaboração. Segundo o Guia de Elaboração de Itens do CAEd, é importante adequar os itens, durante sua produção, aos descritores e apresentar uma linguagem apropriada aos estudantes do período de escolarização. A terminologia técnica e rebuscada pode influenciar nos resultados, uma vez que interfere no entendimento dos estudantes e, conseqüentemente, no acerto ao item.

As características apresentadas dos itens não são restritas aos que apresentam imagens como suporte, em uma aplicação para o 1º ano do Ensino Médio do Avalia Ba um dos itens apresentava apenas um enunciado, como o item mostrado pela figura 19.

Figura 19 – Item espelho do Avalia Ba

Exemplo 5 - O DNA é o ácido nucleico que contém as informações genéticas dos indivíduos. Suas moléculas são polinucleotídeos formados por duas cadeias dispostas em forma de hélice. As duas cadeias estão unidas entre si pelas bases nitrogenadas.

Qual o açúcar presente nessa molécula?

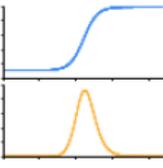
- A) Desoxirribose.
- B) Frutose.
- C) Glicose.
- D) Ribose.
- E) Sacarose.

Fonte: Acervo pessoal da autora.

Essa Figura retrata um exemplo de um item aplicado que teve como objetivo, avaliar a habilidade dos estudantes em identificar a natureza química do DNA. O descritor que ampara essa habilidade é comum aos três Programas de Avaliação. Para acertar o item, os estudantes devem associar a molécula citada no enunciado com o açúcar correspondente. Apesar de apresentar um pequeno texto sobre a molécula, as informações fornecidas pelo elaborador não auxiliam os estudantes na resolução do item, pois essas informações não são necessárias para que eles respondam corretamente. Portanto, mesmo apresentando um texto, esse tipo de item possui as mesmas características dos apresentados anteriormente, pois exigem dos estudantes a memorização de informações sem estimular os processos para a construção de conhecimentos.

O descritor, utilizado na construção desses itens, utiliza o verbo “identificar” que abarca uma habilidade elementar ancorada na base da pirâmide. Apesar disso, a análise estatística do item original (Figura 20) aponta claramente para a dificuldade na sua resolução, pois o parâmetro **b** está alto (1103,72) e o parâmetro **a** muito baixo. Além disso, o ponto de ancoragem, que corresponde à proficiência necessária para uma probabilidade de 65% dos estudantes acertarem o item, vai além da escala de proficiência do Programa de Avaliação, atingindo um índice de 1144.675.

Figura 20 - Estatísticas do Item do Avalie Ba

Item:		Descritor: D84		Gabarito: A		Suporte: NENHUM		Bloco: 19		Pos.: 2		Curva
Parametro da TRI			Análise Clássica			Percentuais por opção de resposta						
A	B	C	Perc. de Acerto	Bisserial	Ancoragem	B/N	A	B	C	D	E	
0,0063	1103,72	0,1073	0,116	0,267	1144,675	0,011	0,116	0,091	0,568	0,081	0,133	
OBS: BIOLOGIA						Biss A	Biss B	Biss C	Biss D	Biss E		
						0,267	-0,174	-0,031	-0,018	-0,008		

Fonte: Acervo pessoal da autora.

A maioria dos livros didáticos aborda esse tema de uma maneira conteudista, o que dificulta a contextualização e a didática do professor em sala de aula. De acordo com Marcon e Stange (2008), o aprendizado de biologia no ensino médio se dá de modo conteudista, comportamentalista, empirista-indutivista no qual os conceitos estudados são transmitidos e retransmitidos tão somente em forma de definição.

Nessa mesma perspectiva, Pontes e Corazza (2010) afirmam que

[...] apesar de a Biologia fazer parte do dia-a-dia da população, o ensino dessa disciplina encontra-se tão distanciado da realidade que não permite à população, perceber o vínculo estreito existente entre o que é estudado na disciplina Biologia e o cotidiano. Essa visão dicotômica impossibilita ao aluno estabelecer relações entre a produção científica e o seu contexto, prejudicando a necessária visão holística que deve pautar o aprendizado sobre esta ciência (PONTES; CORAZZA, 2010).

Nesse sentido, uma forma de evitar itens como o apresentado é apontar textos ou imagens nos suportes que auxiliem os estudantes alcançarem a habilidade avaliada. Observe o item mostrado no exemplo 6, produzido pela autora.

Exemplo 6: Leia o texto abaixo.

Nos dias de hoje, podemos dizer que praticamente todos os seres humanos já ouviram em algum momento falar sobre o DNA e seu papel na hereditariedade da maioria dos organismos. Porém, foi apenas em 1952, um ano antes da descrição do modelo do DNA em dupla hélice por Watson e Crick, que foi confirmado sem sombra de dúvidas que o DNA é material genético. A estratégia empregada por Watson e Crick foi construir um modelo molecular que levasse em conta o tamanho e a configuração espacial dos nucleotídeos. Assim, através de estudos de difração de raios X, Watson e Crick revelaram que a molécula de DNA é um composto formado por duas longas cadeias paralelas, constituídas por nucleotídeos dispostos em sequência. Os nucleotídeos são moléculas compostas por uma pentose, uma molécula de ácido fosfórico e uma base nitrogenada.

Disponível em: GRIFFITHS, A. J. F. et al. Introdução à Genética. Rio de Janeiro: Guanabara Koogan, 2002.

Qual o açúcar compõe a molécula citada no texto?

- A) Desoxirribose.
- B) Frutose
- C) Glicose.
- D) Ribose.
- E) Sacarose.

Dessa forma, os estudantes têm disponível, meios de traçar um raciocínio acerca das informações fornecidas para associar o tipo de açúcar à molécula do ácido nucleico.

Seguindo a mesma temática, outro descritor, comum entre as matrizes dos três projetos analisados, busca avaliar a habilidade dos estudantes de identificar características das etapas do processo de síntese proteica. Um exemplo de item, aplicado em uma edição do SADEAM para estudantes do primeiro ano do Ensino Médio avalia essa habilidade conforme mostrado na Figura 21.

Figura 21- Item espelho do SADEAM

Exemplo 7: A transcrição é a primeira etapa da expressão do gene.

Essa etapa é responsável pela formação de

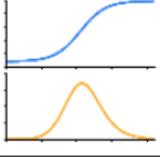
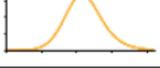
- A) Cadeias de polipeptídeos.
- B) Moléculas de DNA.
- C) Polirribossomos.
- D) RNA mensageiro.
- E) RNA transportador.

Fonte: Acervo pessoal da autora.

Para responder esse item, os estudantes devem compreender os processos que ocorrem com os ácidos nucleicos e que são essenciais para o funcionamento das nossas células. Com isso, a transcrição que se refere a um desses processos é abordada sem nenhuma contextualização ou suporte que ofereça condições do

estudante criar um caminho cognitivo para achar a resposta correta que, conforme Figura 22 é a alternativa D. Além disso, as estatísticas, geradas para o item original, mostram que há duas bisseriais positivas que indicam que boa parte dos estudantes proficientes marcaram o distrator E ao invés do gabarito. Outros índices também indicam a ineficiência do item, como o percentual de acerto baixo (13,1%), parâmetro **b** (1050,063), ponto de ancoragem elevados (1137,485) e parâmetro **a** muito baixo (0,00331).

Figura 22 - Estatísticas do Item do SADEAM

Item:		Descritor: D34		Gabarito: D		Suporte: NENHUM		Bloco: 13		Pos.: 6		Curva
Parametro da TRI			Análise Clássica			Percentuais por opção de resposta						
A	B	C	Perc. de Acerto	Bisserial	Ancoragem	B/N	A	B	C	D	E	
0,00331	1050,06 3	0,07759	0,131	0,272	1137,485	0,003	0,121	0,474	0,107	0,131	0,164	
OBS: PT SADEAM 10 2014							Biss A	Biss B	Biss C	Biss D	Biss E	
							-0,066	-0,137	-0,062	0,272	0,101	

Fonte: Acervo pessoal da autora.

A proposta para evitar os comportamentos estatísticos apresentados é utilizar a contextualização através de situações cotidianas dos estudantes, como mostrado no exemplo 7 (elaborado por mim).

Exemplo 7: Os antibióticos são utilizados no combate à doenças bacterianas, porém o uso indiscriminado desses medicamentos podem gerar graves problemas de saúde pública. Um exemplo foi o surgimento das KPC (*Klebsiella Pneumoniae Carbapenemase*), bactérias resistentes à vários antibióticos. No entanto, uma classe de antibióticos obteve êxito por agir no ribossomo das células bacterianas, impedindo assim o funcionamento desse componente celular.

Com isso, essa classe de antibiótico é eficaz pois,

- A) destrói a membrana plasmática.
- B) evita a transcrição genica.
- C) impede a síntese de proteínas.
- D) modifica o código genético.
- E) promove mutações genicas.

Dessa forma, mesmo se tratando de um tema complexo, a abordagem da ação de antibióticos aproxima os conteúdos escolares do universo dos estudantes auxiliando-os na resolução do problema proposto.

As propostas dos itens apresentados convergem com as reformas do ensino previstas pela Lei de Diretrizes e Bases de 1996, que sugere que o ensino evite a transmissão e a memorização de conteúdos de modo mecânico, objetivo e inflexível, propondo que possibilite uma articulação bem estabelecida entre os conteúdos a serem ensinados e o mundo vivencial dos estudantes por meio da contextualização (BRASIL, 1996, p.46).

As análises realizadas mostram a necessidade de uma reflexão sobre a forma como os itens são elaborados. Sabe-se que o Guia de Elaboração de Itens utilizado pelo CAEd foi produzido inicialmente para criação de itens de Matemática e Língua Portuguesa e atualmente é utilizado em todas as áreas de conhecimento. Diante disso, a elaboração dos itens de Ciências da Natureza segue os moldes de outras áreas de conhecimento desconsiderando suas especificidades.

Os descritores de citologia, utilizados como base para análise desta pesquisa, são objetivos e abordam temas que, para itens de múltipla escolha, perpassam por habilidades de memorização e dificultam a elaboração de itens contextualizados. Um exemplo comum aos três projetos é o descritor que avalia a habilidade em identificar a natureza química dos ácidos nucleicos (DNA e RNA). Os itens elaborados para esse descritor, na maioria das vezes, apresentam como suporte a estrutura dos ácidos nucleicos e o comando aborda uma característica específica ou diferencial de um desses ácidos. Essa utilização pode ser entendida como uma limitação às regras propostas pelo Guia de Elaboração de Itens e ao próprio descritor que não abre margem para a contextualização, pois a identificação da natureza dos ácidos nucleicos é um tema específico, sem muita relação com o cotidiano dos estudantes.

Para confirmar essa hipótese, foi realizado um levantamento dos itens de citologia com baixos índices de acerto (até 50%) nos três Programas de Avaliação e posteriormente classificados de acordo com suas características conforme Tabela 10.

Tabela 10 – Características de Elaboração dos Itens de Citologia com baixos percentuais de Acerto

Características dos itens com baixos percentuais de acerto	Quantidade de itens		
	SADEAM	PAEBES	AVALIE Ba
Suporte dispensável	19	5	4
Linguagem técnica	13	2	16
Memorização	38	19	1
Sem problemas técnicos	20	6	9
TOTAL	90	32	30

Fonte: Acervo pessoal da autora.

Com um total de 152 itens de citologia com percentuais de acerto de até 50%, foi possível perceber que grande parte desses itens tem o suporte dispensável, ou seja, o suporte não oferece recursos cognitivos para os estudantes responderem ao comando. É o caso de itens com imagem de células, organelas e estruturas celulares, cujo comando pede a função dessas estruturas ou a diferença entre elas.

A memorização dos conteúdos de modo mecânico também foi encontrada em boa parte dos itens que exigiam dos estudantes conhecimentos do tema sem qualquer associação com uma situação ou contexto. O exemplo de item apresentado na Figura 19 ilustra esse processo.

Deve-se ressaltar que a linguagem técnica e rebuscada foi expressiva nos itens analisados e afeta diretamente na avaliação da habilidade, pois a dificuldade do entendimento do item por parte dos estudantes impede que consigam alcançar a proficiência esperada.

Diante das análises apresentadas, é necessário considerar a elaboração dos itens que compõem os testes de proficiência do CAEd, a fim de verificar sua consistência e eficácia, além de englobar questões sociais e do cotidiano dos estudantes que auxiliem na formação do processo cognitivo e no alcance das habilidades avaliadas.

Nesse sentido, a partir dos pontos analisados nesta seção, será proposto no próximo capítulo um plano de ação. Esse plano poderá contribuir para o aperfeiçoamento do sistema de elaboração de itens do CAEd para que assim, possa oferecer dados e informações capazes de subsidiar as ações de melhoria da qualidade da educação e equidade nas oportunidades educacionais.

3 PLANO DE AÇÃO

Nesse capítulo, será apresentado um plano de ação que visa a adequação dos instrumentos da Avaliação em Larga escala, para aperfeiçoar a elaboração de itens realizada pela Coordenação de Instrumentos de Avaliação do CAEd.

O objetivo desta dissertação foi compreender quais são os fatores que interferem na qualidade dos itens de biologia nas avaliações dos projetos SADEAM, PAEBES e AVALIA BA no período de 2010 a 2015. Para isso, foi importante apontar no capítulo 1 a metodologia da elaboração dos itens pelo CAEd e o sistema organizacional no processo dessa elaboração.

Além disso, foi necessário apresentar os Programas de Avaliação do Amazonas (SADEAM), Espírito Santo (PAEBES) e da Bahia (AVALIA Ba) que serviram de base de análise da pesquisa e, atrelado a isso, foi apresentada a elaboração de itens de acordo com o Guia de Elaboração do Itens do CAEd e a montagem dos testes de proficiência.

Com as análises psicométricas e pedagógicas realizadas no capítulo 2, foi possível perceber que dos descritores de Biologia, dos três Programas de Avaliação analisados, aqueles que abordam o tema de citologia são os que possuem menores índices de acerto. Essa característica pode estar relacionada à forma como esses itens foram formulados, sem contexto, suporte funcional e linguagem técnica e rebuscada para as etapas de escolaridade avaliadas.

Porém, para possibilitar a melhoria da qualidade dos itens elaborados é necessário rever os descritores das matrizes de referência utilizados como base de elaboração dos itens.

Em função disso, neste capítulo, apresentaremos propostas de reformulação dos descritores das Matrizes de Referência para adequar as especificidades das Ciências da Natureza.

3.1 Reformulação dos descritores das Matrizes de Referência

Para adequar os descritores de modo que se tornem uma base de conhecimentos contextualizados para a produção de itens de qualidade, serão consideradas as competências específicas de Ciências da Natureza e suas Tecnologias para o Ensino Médio da Base Nacional Comum Curricular. A utilização

da BNCC nessa pesquisa é necessária devido ao debate no cenário educacional atual sobre a democratização e aproximação das avaliações em larga escala com a realidade escolar.

A proposta da BNCC para o Ensino Médio é proporcionar aos estudantes uma continuidade ao tratamento dado no Ensino Fundamental, através de um aprofundamento e ampliação das reflexões a respeito das tecnologias em relação ao seu papel na sociedade e às perspectivas futuras de desenvolvimento tecnológico (BRASIL, 2017).

Nesse sentido, o documento apresenta três competências específicas que estão relacionadas às habilidades a serem alcançadas nessa etapa, são elas

1. Analisar fenômenos naturais e processos tecnológicos, com base nas relações entre matéria e energia, para propor ações individuais e coletivas que aperfeiçoem processos produtivos, minimizem impactos socioambientais e melhorem as condições de vida em âmbito local, regional e/ou global.
2. Construir e utilizar interpretações sobre a dinâmica da Vida, da Terra e do Cosmos para elaborar argumentos, realizar previsões sobre o funcionamento e a evolução dos seres vivos e do Universo, e fundamentar decisões éticas e responsáveis.
3. Analisar situações-problema e avaliar aplicações do conhecimento científico e tecnológico e suas implicações no mundo, utilizando procedimentos e linguagens próprios das Ciências da Natureza, para propor soluções que considerem demandas locais, regionais e/ou globais, e comunicar suas descobertas e conclusões a públicos variados, em diversos contextos e por meio de diferentes mídias e tecnologias digitais de informação e comunicação (TDIC) (BRASIL, 2016).

As habilidades dos descritores analisadas nessa pesquisa estão ancoradas na competência específica 2 apresentadas no Quadro 2.

Quadro 2 – Habilidades da Competência 2

HABILIDADES
(EM13CNT201) Analisar e utilizar modelos científicos, propostos em diferentes épocas e culturas para avaliar distintas explicações sobre o surgimento e a evolução da Vida, da Terra e do Universo.
(EM13CNT202) Interpretar formas de manifestação da vida, considerando seus diferentes níveis de organização (da composição molecular à biosfera), bem como as condições ambientais favoráveis e os fatores limitantes a elas, tanto na Terra quanto em outros planetas.
(EM13CNT203) Avaliar e prever efeitos de intervenções nos ecossistemas, nos seres vivos e no corpo humano, interpretando os mecanismos de manutenção da vida com base nos ciclos da matéria e nas transformações e transferências de energia.
(EM13CNT204) Elaborar explicações e previsões a respeito dos movimentos de objetos na Terra, no Sistema Solar e no Universo com base na análise das interações gravitacionais.
(EM13CNT205) Utilizar noções de probabilidade e incerteza para interpretar previsões sobre atividades experimentais, fenômenos naturais e processos tecnológicos, reconhecendo os limites explicativos das ciências.
(EM13CNT206) Justificar a importância da preservação e conservação da biodiversidade, considerando parâmetros qualitativos e quantitativos, e avaliar os efeitos da ação humana e das políticas ambientais para a garantia da sustentabilidade do planeta.
(EM13CNT207) Identificar e analisar vulnerabilidades vinculadas aos desafios contemporâneos aos quais as juventudes estão expostas, considerando as dimensões física, psicoemocional e social, a fim de desenvolver e divulgar ações de prevenção e de promoção da saúde e do bem-estar.

Fonte: BRASIL, 2016.

Os descritores de citologia estão enquadrados na habilidade EM13CNT202 que permite a exploração dos avanços tecnológicos relacionados às aplicações do conhecimento sobre células e suas estruturas que perpassa conhecimentos de células-tronco e bioquímica.

Portanto, nessa seção, apresentarei a proposta da reformulação do Domínio Vida e Ambiente que enquadra os descritores discutidos nessa dissertação. Para a reformulação desses descritores, foram consideradas as propostas da BNCC e PCNs.

No Quadro 3, apresenta-se os descritores elaborados para o 3º ano do Ensino Médio na disciplina de biologia.

Quadro 3 – Domínio Matéria e Ambiente da Matriz de Referência do 3º ano do Ensino Médio de Biologia

Descrição da Habilidade
Vida e Ambiente
Interpretar a função das organelas celulares para explicar os processos biológicos em seus níveis de organização.
Analisar a estrutura celular e o funcionamento de seus componentes afim de explicar os processos de síntese, armazenamento, transporte e metabolismo energético.
Associar a síntese proteica aos processos vitais dos organismos.
Compreender a divisão celular como um processo essencial para o desenvolvimento da vida.
Reconhecer os principais componentes das células prevendo ou explicando a manifestação de características dos seres vivos.

Fonte: Elaborado pela autora.

O domínio, como componente da Matriz de Referência é composto por cinco descritores e contempla a temática de citologia do eixo Vida e Ambiente. As matrizes dos três projetos analisados apresentam uma média de nove descritores nesse domínio que englobam outros temas. Portanto, o número de descritores sugeridos é satisfatório para um diagnóstico de qualidade. A reformulação proposta teve a intenção de auxiliar na elaboração dos itens promovendo a melhoria da qualidade dos testes.

Os três Programas de Avaliação apresentam o descritor comum: “*Identificar a importância das organelas e do núcleo para o metabolismo celular*”. Como já discutido, esse descritor limita a atuação do elaborador na produção de itens mais próximos da realidade dos estudantes, pois o verbo identificar remete ao processo de memorização que pode ser comprovado pela observação dos itens presentes nas avaliações desses Programas.

A reformulação desse descritor propõe a interpretação da função das organelas celulares para explicar os processos biológicos em seus níveis de organização. Diante disso, os itens para esse descritor exigirão dos estudantes a

conversão da informação apresentada para outra, ou seja, construir ou selecionar a mesma informação em uma forma diferente.

Nesse sentido, esse descritor vai além dos processos de memorização e para isso, a informação apresentada deve ser inédita a qual os estudantes não a encontraram durante a instrução. Assim, eles não irão responder o item confiando apenas na memória, o que garante a elaboração dos processos cognitivos e sua avaliação.

Complementando essa ideia, o segundo descritor, apresentado no Quadro 3, propõe a análise da estrutura celular e o funcionamento de seus componentes, a fim de explicar os processos de síntese, armazenamento, transporte e metabolismo energético. Esse descritor se refere à reformulação do descritor: “*Identificar os componentes bioquímicos da célula e suas principais funções*”, presente também nos Programas de Avaliação analisados.

Essa reformulação abarca um tema mais amplo que o descritor original, o que facilita a elaboração de itens pelo fato de possibilitar estratégias distintas sem englobar outro descritor.

Propor a análise de uma habilidade engloba os processos cognitivos *diferenciar, organizar e atribuir*. Os objetivos desse descritor baseiam-se na determinação das partes relevantes ou importantes de uma mensagem (*diferenciar*), os modos nos quais os fragmentos da mensagem são (*organizar*) e o propósito subjacente da mensagem (*atribuir*).

Nesse sentido, os itens produzidos a partir do descritor proposto pode levar os estudantes a diferenciar os tipos celulares, comparar células vegetais e animais, entre outros. Quanto à estrutura do item, os elaboradores dispõem de diversas formas para apresentar um suporte, seja através de fragmentos de textos que expõe o assunto, imagens e até mesmo situações problemas que englobam uma temática relevante na vida dos estudantes.

O terceiro descritor sugere a reformulação do descritor: “*Compreender o processo de síntese proteica*”. Apesar da categoria “compreender”, segundo a taxonomia de Bloom (1983), ser mais ampla que a categoria “lembrar”, os itens elaborados a partir desse descritor apresentam, em sua maioria, imagens das etapas de síntese proteica para que o estudante identifique uma etapa ou o nome do

processo. A proposta de reformulação para esse descritor é a associação da síntese proteica aos processos vitais dos organismos.

O termo “associar” remete à habilidade de ordenar, classificar, agrupar, etc. Nesse sentido, para responder os itens elaborados a partir do descritor proposto, os estudantes deverão compreender as etapas da síntese proteica associando a um dos processos vitais dos organismos. A inclusão dessa associação possibilita a elaboração de itens mais contextualizados, através da utilização dos vários processos vitais dos organismos, como a respiração, fotossíntese, atividades físicas entre outros.

Os itens que abordam os temas de divisão celular do descritor: “*Reconhecer os processos de divisão celular*”, como os itens do descritor anterior, apresentam uma imagem de uma divisão celular e exigem dos estudantes a identificação de uma das etapas apresentadas. A categoria “reconhecer”, segundo a taxonomia de Bloom (1983), está relacionada à localização de um conhecimento na memória de longa duração que seja coerente com o material apresentado, o que justifica o padrão de itens produzidos. Porém, esses itens pouco avaliam os estudantes, que muitas vezes ficam confusos com os termos e as etapas apresentadas, como demonstradas no capítulo 2.

Portanto, a proposta de reformulação desse descritor é a de compreender a divisão celular como um processo essencial para o desenvolvimento da vida. Visto que a categoria “compreender” está além dos processos de memorização, essa proposta busca o entendimento dos estudantes da importância das divisões celulares para o desenvolvimento da vida e não só a memorização das suas etapas.

O processo de formação de gametas nos indivíduos com reprodução sexuada, crescimento e cicatrização dos tecidos entre outros são situações que podem ser abordadas nos itens para avaliar esse descritor. Além disso, esse descritor permite avaliar a habilidade dos estudantes em entender o processo de formação de novas células através de mitose e meiose, células tronco e as fases do desenvolvimento embrionário até o parto.

O último descritor traduz a reformulação do descritor: “*Identificar a natureza química do DNA e do RNA*”. Os itens dessa natureza apresentam, em sua maioria, a imagem da estrutura do DNA e/ou RNA e exigem a diferenciação ou identificação de tais moléculas, buscando mais uma vez os processos de memorização dos estudantes. Ao propor a reformulação, o descritor passa a avaliar a habilidade de

reconhecer os principais componentes das células prevendo ou explicando a manifestação de características dos seres vivos.

Nesse sentido, os estudantes devem associar a estrutura componentes celulares0020 com a manifestação das características dos seres vivos e não apenas a identificação de sua estrutura. Os itens produzidos a partir desse descritor podem exigir desde os processos de replicação do DNA e transcrição do RNA até a explicação da expressão das características dos indivíduos.

Após a construção do domínio que compõe a Matriz de Referência, o próximo passo é a apresentação desse material aos coordenadores e analistas da Coordenação de Instrumentos de Avaliação do CAEd para a avaliação e validação dos descritores.

Com a proposta aprovada, os descritores serão incluídos na matriz geral do CAEd e passarão a servir de base para elaboração de outras matrizes conforme a especificidade dos Programas e da elaboração dos itens que compõem o banco de dados da instituição.

CONSIDERAÇÕES FINAIS

A presente dissertação teve como objetivo compreender quais são os fatores que interferem na qualidade dos itens de biologia nas avaliações dos projetos SADEAM, PAEBES e AVALIA BA no período de 2010 a 2015.

As análises bibliográficas permitiram compreender a trajetória da avaliação em larga escala nas áreas de Ciências da Natureza e sua importância no cenário educacional brasileiro. Além disso, foi mostrado que as avaliações externas são instrumentos úteis para o diagnóstico do ensino, sendo capazes de nortear as intervenções pedagógicas para superar as dificuldades dos estudantes.

No entanto, é importante questionar e aprofundar análises acerca do seu desenho, pois sua qualidade está atrelada à eficácia do sistema no qual os beneficiários da política são diretamente impactados.

O estudo parte do pressuposto de que boa parte dos itens de biologia, presentes nos testes de proficiência desses Programas de Avaliação, apresentam baixos percentuais de acerto. Essa hipótese pode ser confirmada através da análise do comportamento estatístico dos itens em relação ao percentual de acerto e valores do parâmetro *b* obtido nos relatórios fornecidos pelo CAEd.

A análise mostrou que a maior parte dos itens levantados apresenta um percentual de acerto menor que 50% e identificou a citologia como a temática mais representativa entre os itens com baixos percentuais de acerto.

A metodologia qualitativa, aplicada nesse estudo, consistiu na análise pedagógica dos itens de citologia, alcançando assim objetivo inicial do estudo. Como resultado, a análise identificou que um grande número de itens com baixos percentuais de acerto apresenta um padrão de elaboração simples e que não corresponde aos preceitos do Guia de Elaboração de Itens explanados no capítulo 2.

Através dessas análises é possível constatar que a elaboração dos itens de citologia é limitada por descritores que não possibilitam a criação de situações problemas ou contextualização. Itens elaborados dessa forma não avaliam as habilidades esperadas, mas sim a capacidade de memorização dos estudantes. Além disso, uma vez que a maioria dos itens apresenta caráter disciplinar, infere-se que pode ocorrer desmotivação para a prática interdisciplinar e contextualizada descrita nos PCNs. (BRASIL, 2002).

Outro ponto a ser considerado e discutido nessa dissertação é a característica ampla das habilidades presentes na Base Nacional Comum Curricular do Ensino Médio que tem por objetivo qualificar os jovens para o uso das tecnologias, reconhecer e interpretar fenômenos e situações práticas, fazer julgamentos, entre outros (BRASIL, 2017).

Após a leitura dos diversos autores sobre a temática dessa dissertação, compreende-se que a Teoria de Resposta ao Item é primordial nas avaliações em larga escala, uma vez que apresenta uma melhor compreensão do instrumento de mensuração. Além disso, a TRI permite a construção de instrumentos menores e igualmente confiáveis, melhorando a avaliação dos estudantes.

Porém, para atingir seus objetivos, é necessário que a elaboração dos itens esteja adequada à TRI para que possa garantir a qualidade dos testes de proficiência.

Nesse sentido, foi proposto, como plano de ação, a reformulação dos descritores de citologia para possibilitar formas variadas de elaboração dos itens, permitindo aos elaboradores construir itens contextualizados e interdisciplinares de acordo com as orientações das competências específicas e habilidades das Ciências da Natureza e suas tecnologias da BNCC. Essas competências inter-relacionam-se e desdobram-se no tratamento didático proposto para as três etapas da Educação Básica (Educação Infantil, Ensino Fundamental e Ensino Médio), articulando-se na construção de conhecimentos, no desenvolvimento de habilidades e na formação de atitudes e valores, nos termos da LDB (BRASIL, 2017).

A reformulação dos descritores, proposta nesse estudo, permite ao elaborador empregar outros tipos de suporte, como textos ou imagens sobre processos biológicos, relações ambientais e aplicação do conhecimento em diferentes disciplinas através da interdisciplinaridade. Portanto, essa abordagem evita a problematização dos conceitos que eram utilizados de maneira isolada nos itens antes elaborados.

Para os demais descritores que compõem as Matrizes de Referência utilizadas pelo CAEd, é necessário, de acordo com o observado, aprofundar os estudos sobre a redação dos descritores e a forma como os itens são elaborados. Além disso, é importante identificar as congruências e distanciamentos entre o que é previsto oficialmente, o que é abordado em sala de aula e o que é apresentado na avaliação em larga escala, para que as avaliações estejam de acordo com a

proposta da BNCC. Com isso, possam, de fato, promover melhorias significativas no processo de ensino e aprendizagem em Ciências da Natureza ao oferecer um teste coerente com o currículo que possibilite ações possíveis de ser implementadas e desenvolvidas pelos professores.

REFERÊNCIAS

AMAZONAS. Secretaria de Estado de Educação e Qualidade do Ensino - SEDUC. SADEAM – 2015/ Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd. Juiz de Fora, 2015.

ANDRADE, D. F; TAVARES, H. R.; VALLE, R. C. **Teoria da Resposta ao Item: Conceitos e Aplicações**. SINAPE 2000.

_____. Secretaria de Estado de Educação e Qualidade do Ensino - SEDUC. Revista Pedagógica. Ciências da Natureza – Ensino Médio e EJA Ensino Médio. SADEAM – 2015/ Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd. Juiz de Fora, 2015.

_____. Secretaria de Estado de Educação e Qualidade do Ensino - SEDUC. Revista do Sistema de Avaliação - 2015/ Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd. Juiz de Fora, 2015.

_____. Secretaria de Educação do Estado. Sistema de Avaliação do Desempenho Educacional do Amazonas – SADEAM. Disponível em: <<http://www.sadeam.caedufjf.net/>>. Acesso em: 18 setembro 2017.

ANDRADE, D. F; TAVARES, H. R; VALLE, R. C. **Teoria da Resposta ao Item: Conceitos e Aplicações**. Projeto Temático da FAPESP no. 96/01741-7 e PRONEX no. 76.97.1081.00. 2000.

BAHIA. Secretaria da Educação do Estado. SABE (Sistema de Avaliação Baiano de Educação). AVALIE ENSINO MÉDIO – 2013 / Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd. Juiz de Fora, 2013.

_____. Secretaria da Educação. Revista Pedagógica – Ciências da Natureza e suas tecnologias – Ensino Médio. Avalie Ensino Médio – 2013/ Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd. Juiz de Fora, 2013.

_____. Secretaria da Educação. Revista do Sistema de Avaliação. Avalie Ensino Médio – 2013/ Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd. Juiz de Fora, 2013.

_____. Secretaria de Educação do Estado. SISTEMA DE AVALIAÇÃO BAIANO DA EDUCAÇÃO – SABE. Disponível em: <<http://www.avalieba.caedufjf.net/>>. Acesso em: 18 setembro 2017.

BEKMAN, R. M. **Aplicação dos blocos incompletos balanceados na teoria de resposta ao item**. Estudos em Avaliação Educacional, nº 24, jul-dez/2001.

BECKER, F. R. **Avaliação educacional em larga escala: a experiência brasileira**. Revista Iberoamericana de Educación, 53, p. 1-25, 2010.

BRASIL. Constituição da República Federativa do Brasil. Promulgada em 5 de outubro de 1988. Disponível em: <[http://www.planalto.gov.br/ccivil_03/constituicao/constituicao .htm](http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm)>. Acesso em: 18 setembro 2017.

_____. Lei de Diretrizes e Bases da Educação Nacional, Lei nº 9.394, de 24 de dezembro de 1996.

_____. Ministério da Educação. **Brasil no PISA 2015: Análises e reflexões sobre o desempenho dos estudantes brasileiros**. OCDE-Organização para a Cooperação e Desenvolvimento Econômico. — São Paulo: Fundação Santillana, 2016.

_____. Ministério da Educação. Governo Federal. **Base Nacional Curricular Comum**: Ensino Médio. Brasília, 2017. Disponível em: <<https://goo.gl/Ruxwpw>>. Acesso em: 16 de março. 2018.

_____. Ministério da Educação. **PDE: Plano de Desenvolvimento da Educação**: SAEB: ensino médio: matrizes de referência, tópicos e descritores. Brasília: MEC, SEB; Inep, 2008. 127 p.: il.

_____. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Histórico do Saeb. 2015. Disponível em: <<http://portal.inep.gov.br/educacao-basica/saeb/historico-do-saeb>>. Acesso em: 17 set. 2017.

_____. Ministério da Educação e do Desporto. Sistema Nacional de Avaliação da Educação Básica – SAEB. Secretaria de Desenvolvimento, Inovação e Avaliação Educacional. Instituto Nacional de Estudos e Pesquisas Educacionais. Brasília. 1995.

BECKER, R. F. **Avaliação educacional em larga escala: a experiência brasileira.** Revista Ibero-americana de educação. ISSN: 1681-5653. Nº53/1, 26 jun. 2010.

BEKMAN, R. **Aplicação dos Blocos Incompletos Balanceados na Teoria de Resposta ao Item.** Estudos em Avaliação Educacional, nº 24, jul/dez. São Paulo, FCC: 2001. Disponível em: <<http://www.fcc.org.br/pesquisa/publicacoes/eae/arquivos/990/990.pdf>>. Acesso em 17 set. 2017.

BLOOM, B. et al. **Taxonomia dos objetivos educacionais: domínio cognitivo.** Porto Alegre: Globo, 1983.

BROOKE, N. **Marcos históricos na reforma da educação.** 1. Ed. – Belo Horizonte, MG: Fino traço, 2012. p. 55-111.

BROOKE, N; CUNHA, M. A. A.; FALEIROS, Matheus. **A Avaliação Externa como Instrumento da Gestão Educacional nos Estados: Relatório Final.** Disponível em: <<http://www.institutounibanco.org.br/>>. Acesso em 17/09/2017.

CÁRIA, N. P.; OLIVEIRA, S. M. S. S. **Avaliação Em Larga Escala e a Gestão da Qualidade Da Educação.** Revista de Ciências Humanas – Educação, v.16, n. 26, p. 22-40, 2015.

CAEd/UFJF. **Guia de Elaboração de Itens.** Juiz de Fora: 2014.

_____. **Programa de Avaliação da Educação Básica do Espírito Santo.** 2017. Disponível em: <<http://www.paebes.caedufjf.net/?pg=paebes>>. Acesso em: 23 mai. 2017.

_____. **Sistema de Avaliação Educacional do Desempenho Educacional do Amazonas.** 2017. Disponível em: <<http://www.sadeam.caedufjf.net/s>>. Acesso em: 23 mai. 2017.

_____. **Sistema de Avaliação Baiano da Educação - Sabe.** 2017. Disponível em: <<http://www.avalieba.caedufjf.net/>>. Acesso em: 23 mai. 2017.

_____. Guia de Estudos elaborado para Curso para Professores do Estado do Amazonas (CAEd, 2012).

COELHO. M. I. M. **Vinte anos de avaliação da educação básica no Brasil: aprendizagens e desafios.** Ensaio: aval. pol. públ. Educ., Rio de Janeiro, v. 16, n. 59, p. 229-258, abr./jun. 2008.

ESPÍRITO SANTO. Secretaria de Estado de Educação. PAEBES – 2015/ Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd. Juiz de Fora, 2015.

_____. Secretaria de Estado de Educação. Revista Pedagógica. Ciências da Natureza – 3º ano do Ensino Médio. PAEBES – 2015/ Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd. Juiz de Fora, 2015.

_____. Secretaria de Estado de Educação. Revista do Sistema de Avaliação - 2015/ Universidade Federal de Juiz de Fora, Faculdade de Educação, CAEd. Juiz de Fora, 2015.

_____. Secretaria de Estado de Educação. Programa de Avaliação da Educação Básica do Espírito Santo – PAEBES. Disponível em: <<http://www.paebes.caedufjf.net/>>. Acesso em: 18 setembro 2017.

ESQUINSANI, R. S. S; DAMETTO, J. **Avaliações educacionais em larga escala: ponderações críticas alicerçadas na obra de Paulo Freire.** v. 23, n. 1, Passo Fundo, p. 205-217, jan./jun. 2016| Disponível em: <www.upf.br/seer/index.php/rep>. Acesso em: 17 set. 2017.

_____. Entre o caráter regulatório e o potencial emancipatório do sistema de avaliação da Educação Básica. In: WERLE, F. O. C. (org.). **Avaliação em larga escala: questões polêmicas.** Brasília: Liber Livro, 2012.

ESTEBAN, M. T; FETZNER A. R. **A redução da escola:** a avaliação externa e o aprisionamento curricular. Educar em Revista, Curitiba, Brasil, Edição Especial n. 1/2015, p. 75-92

FERNANDES, N. S.; SOARES, T. M.; PENA, A. C.; CUNHA, I. C. **O conhecimento do professor em avaliação educacional e a proficiência do aluno.** Estudos em Avaliação Educacional, São Paulo, v. 21, n. 47, p. 569-590, set./dez. 2010.

FREIRE, **Pedagogia da autonomia:** saberes necessários à prática educativa. 34. ed. São Paulo: Paz e Terra, 2006.

FREITAS, L. C. **Eliminação adiada:** o ocaso das classes populares no interior da escola e a ocultação da (má) qualidade do ensino. Educação e Sociedade, Campinas, vol.28, n.100, Especial, p.956-987, out. de 2007.

GATTI, B. A. Avaliação educacional no brasil: pontuando uma história de ações EccoS Revista Científica, vol. 4, núm. 1, junho, 2002, pp. 17-41 Universidade Nove de Julho São Paulo, Brasil. Disponível em: <http://www.redalyc.org/articulo.oa?id=71540102>.

GERHARDT, T. E; SILVEIRA, D. T. **Métodos de pesquisa.** Universidade Aberta do Brasil – UAB/UFRGS. Graduação Tecnológica – Planejamento e Gestão para o Desenvolvimento Rural da SEAD/UFRGS. – Porto Alegre: Editora da UFRGS, 2009. 120 p.

GOMES NETO, J. B.; RUSEMBERG, L. **Indicadores de qualidade de ensino e seu papel no Sistema Nacional de Avaliação.** Em Aberto, v.15, n.66, p.13-28, abr./jun.1995.

KATO, D. S; KAWASAKI, C. S. **As concepções de contextualização do ensino em documentos curriculares oficiais e de professores de ciências.** Ciênc. educ. (Bauru) vol.17 nº.1 Bauru. 2011.

KLEIN, R. **Utilização da Teoria de Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB)**. Meta: Avaliação | Rio de Janeiro, v. 1, n. 2, p.125-140, mai./ago. 2003.

LUCENA, C. A. A Teoria do Capital Humano: História, trabalho e Capacitação dos trabalhadores. Universidade Federal de Uberlândia. Sd.

MARCON, M. R; STANGE, C. E. B. **A Contextualização de Novas Tecnologias em Citologia (Células-Tronco) No Ensino Da Biologia**. Plano de Desenvolvimento Educacional. Dia a Dia Educação, 2008. Disponível em: <<https://goo.gl/bLG4xR>>. Acesso em 28 mai. 2018.

MARTINS, E. A. **A influência da “matematização” na aprendizagem de Ciências Naturais: um estudo sobre a aprendizagem da cinemática no 9º ano do Ensino Fundamental**. Dissertação (Mestrado em Educação) – Universidade Federal de Mato Grosso, Cuiabá, 2014.

MENEZES, E. T; SANTOS, T. H. Verbete Conferência de Jomtien. **Dicionário Interativo da Educação Brasileira - Educabrazil**. São Paulo: Midiamix, 2001. Disponível em: <<http://www.educabrazil.com.br/conferencia-de-jomtien/>>. Acesso em: 31 de mai. 2018.

OCDE (2016). PISA 2015 Results: Excellence and Equity in Education. Volume I. Paris: OECD Publishing.

OLIVEIRA, L. K. M. **Três Investigações sobre escalas de proficiência e suas interpretações**. Tese (Doutorado em Educação) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008, 216 f.

OLIVEIRA, L. K. M; FRANCO, C; SOARES, T. M. S. Projeto Geres/2005: **Novos Indicadores para Construção e Interpretação da Escala de Proficiência**. REICE – Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio em Educación. 2007, vol. 5, nº 2e.

OLIVEIRA, M. A. M. O; ROCHA, G. **Avaliação em Larga Escala no Brasil nos Primeiros Anos do Ensino Fundamental.** S/d. Disponível em: <http://www.anpae.org.br/congressos_antigos/simposio2007/274.pdf>. Acesso em: 20 set. 2017.

PALERMO, G. A. SILVA; D. B. N; NOVELINO, M. S. F. **Fatores associados ao desempenho escolar: Uma análise da proficiência em matemática dos alunos do 5º ano do ensino fundamental da rede municipal do Rio de Janeiro.** XVIII Encontro Nacional de Estudos Populacionais, ABEP, Águas de Lindóia/SP – Brasil, de 19 a 23 de novembro de 2012.

PASQUALI, L; PRIMI, R. **Fundamentos da Teoria da Resposta ao Item –TRI.** Avaliação Psicológica, 2(2), 2003, pp. 99-110.

PONTES, L. C; CORAZZA, M. J. **Células-Tronco: Contextualizando o Ensino de Biologia.** Plano de Desenvolvimento Educacional. Dia a Dia Educação, 2008. Disponível em: <<https://goo.gl/x6z3gS>>. Acesso em 28 mai. 2018.

RABELO, M. **Avaliação educacional: fundamentos, metodologia e aplicações no contexto brasileiro.** Rio de Janeiro: SBM, 2013.

RAVITCH, D. **Vida e morte do grande sistema escolar americano:** como os testes padronizados e o modelo de mercado ameaçam a educação. Trad. de Marcelo Duarte. Porto Alegre: Sulina, 2011.

RODRIGUES, M. M.M. **Proposta de Análise de Itens das Provas do Saeb sob a Perspectiva Pedagógica e a Psicométrica.** Estudos em Avaliação Educacional, v. 17, n. 34, maio/ago. 2006.

SARTES, L. M. A; FORMIGONI, M. L. O. S. **Avanços na Psicometria: Da Teoria Clássica dos Testes à Teoria de Resposta ao Item.** Psicologia: Reflexão e Crítica, 26(2), 241-250. 2013.

SILVA, M. M. S; REINHORN, C; SOARES, A; SOARES, T. M. **A abordagem da Avaliação Educacional em Larga Escala nos cursos de graduação em pedagogia**. Rev. bras. Estud. pedagog. (online), Brasília, v. 97, n. 245, p. 46-67, jan./abr. 2016. Disponível em <<http://dx.doi.org/10.1590/S2176-6681/367313975>>.

SILVA, F. G. **A inclusão de Ciências Naturais nos Sistemas de Avaliação da Educação: O caso do Avalia BH**. Dissertação (Mestrado em Gestão e Avaliação da Educação Pública – Universidade Federal de Juiz de Fora, Juiz de Fora – MG, 2014, 108f.

SOUZA, S. Z. & OLIVEIRA, R. P. **Sistemas estaduais de avaliação: uso dos resultados, implicações e tendências**. Cadernos de Pesquisa, v.40, n.141, p.793-822, set./dez. 2010.

SCHWARTZMAN, S; CHRISTOPH, M. A Educação em Ciências no Brasil. Academia Brasileira de Ciências. s.d.

SCHRAMM, W. Notes on case studies of instructional media projects. Working paper, the Academy for Educational Development, Washington, DC, 1971.

TEIXEIRA, A. S. F. **A Contextualização do Saber do Ensino da Biologia e Geologia: Papel da Formação Inicial**. Universidade de Lisboa. 2012.

UFJF. Carta de Serviço ao Cidadão. Centro de Políticas Públicas da Educação (CAEd). Disponível em: <<http://www.ufjf.br/servicos/2014/01/24/centro-de-politicas-publicas-e-avaliacao-da-educacao-caed/>>. Acesso em: 19 mai. 2017.

VIEIRA, M. F. **Escola estadual André Vidal de Araújo: Desequilíbrio entre rendimento escolar interno e desempenho nas avaliações do SADEAM**. 2015. 143f. Dissertação - Programa de Pós-Graduação Profissional em Gestão e Avaliação da Educação Pública da Universidade Federal de Juiz de Fora. Juiz de Fora, 2015.

WERLE, F. O. C. **Políticas de avaliação em larga escala na educação básica: do controle de resultados à intervenção nos processos de operacionalização do**

ensino. **Revista Ensaio: Avaliação e Políticas Públicas em Educação**, v. 19, n. 73, p. 769-792, 2011.

YIN, R. K **Estudo de Caso: Planejamento e Métodos**. Porto Alegre: Bookmann, 2010. Wsis - the world summit on the information society. Geneva 2003 /Tunis 2005. Disponível em: <<http://www.itu.int/wsis/index.html>>. Acesso em: 03, abril. 2018.