



HENRIQUE ARAKEN MARTINS

**Estruturas de avaliação escolar para
mapear habilidades tomando como base
as Taxonomias de Bloom em questões
de múltipla escolha**

Santo André - SP, 2016



Universidade Federal do ABC

PROFMAT - Mestrado Profissional de Matemática

Henrique Araken Martins

Estruturas de avaliação escolar para mapear habilidades tomando como base as Taxonomias de Bloom em questões de múltipla escolha

Orientador: Prof. Dr. Valdecir Marvulle

Trabalho de Conclusão de Curso do Profmat -
Mestrado Profissional em Matemática da
Universidade Federal do ABC, como requisito
parcial para obtenção do título de Mestre em
Matemática

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE
DEFENDIDA PELO ALUNO HENRIQUE ARAKEN MARTINS,
E ORIENTADA PELO PROF. DR. VALDECIR MARVULLE.

Santo André - SP, 2016

Ficha Catalográfica

O sistema de bibliotecas da UFABC está disponibilizando o serviço de elaboração de ficha catalográfica, destinado aos interessados em publicar materiais bibliográficos.

A ficha catalográfica contém informações bibliográficas necessárias para identificação e localização do material disponibilizado em uma coleção. Nela constam o título, o nome dos autores e o assunto abordado no documento.

Para criá-la é necessário o envio de algumas informações para a biblioteca: a folha de rosto, resumo e o sumário da obra, preferencialmente em PDF.

As informações devem ser enviadas para: servicos.biblioteca@ufabc.edu.br

DEDICATÓRIA

Dedico esta dissertação ao meu filho, a razão da minha vida, Diogo de Oliveira Martins, que me inspira e me traz força para nunca desistir. E a minha esposa, Eliene Teixeira de Oliveira, que esteve em todos os momentos ao meu lado, desde o princípio de minha jornada acadêmica.

AGRADECIMENTOS

Agradeço a minha mãe que me ajudou nos momentos mais difíceis; ao meu pai e meus irmãos que sempre me apoiaram; a equipe de professores que compraram a ideia e se esforçaram para o desenvolvimento do projeto; aos meus alunos que são o foco da pesquisa; aos meus amigos de estudo no Profmat, principalmente aqueles que sempre me acompanham; a todos os professores que eu tive o privilégio de conhecer, aprender e se espelhar; ao professor orientador que norteou e aperfeiçou esta pesquisa e principalmente você leitor que está interessado em nosso trabalho.

"Não devemos ter medo das novas ideias! Elas podem significar a diferença entre o triunfo e o fracasso..."
— Napoleon Hill

RESUMO

Na educação, desde sempre existe a necessidade de melhorar os resultados do desempenho de aprendizado de nossos alunos. Sendo assim, com grande frequência surgem novas propostas com intenção de aperfeiçoar o todo, ou parte do processo educativo. *“As estruturas de avaliação escolar para mapear habilidades tomando como base a Taxonomia de Bloom em questões de múltipla escolha”*, tem por objetivo: propor uma metodologia de elaboração de questões seguindo os patamares da hierarquia de aprendizado proposto por Bloom; aplicação on-line de avaliações de múltiplas escolhas; priorizar a devolutiva dos resultados com a finalidade de mapear por meio de um gráfico de cores as habilidades dos alunos, das turmas, das séries, e da escola; obter indicadores do ensino dos professores e obter indicadores dos conteúdos envolvidos.

Na prática, desenvolvemos uma planilha de correção on-line que dinamizou o processo de correção, fornecendo-nos indicadores de processo instantâneos, permitindo-nos refletir, discutir e propor intervenções do nosso ensino, através dos resultados de aprendizado de nossos alunos.

Palavras-chave: Planilhas de autocorreção, Taxonomia de Bloom e Avaliação

ABSTRACT

In education, we always felt the need to improve the outcome of the learning process of our students. Often we hear about new proposals, aimed at improving the whole teaching process, or just part of it. *Models of school evaluation to map skills, based on Bloom's Taxionomy in multiple choice questions* has the following objectives: to propose a methodology for the processing of the questions, based on the different levels of the learning hierarchy suggested by Bloom; the online application of multiple choice evaluations; to prioritize the feedback about results, in order to map the skills of students, classes, and of a given school through a color graph; to obtain indicators about the teaching performance of the teacher, and to the relevant contents that are involved. We developed a worksheet for the online correction, which can dinamize the proofing process and result in instantaneous process indicators, thus allowing us to reflect, discuss and suggest interventions in our education system, based on the performance of our students.

Keywords:Self-Healing Grid, Bloom's Taxonomy and Evaluation

SUMÁRIO

1	INTRODUÇÃO	1
2	AVALIAÇÃO	5
2.1	AVALIAÇÃO ESCOLAR INTERNA	6
2.1.1	Avaliação Diagnóstica	6
2.1.2	Avaliação Discursiva	6
2.1.3	Avaliação Objetiva	6
2.1.4	Avaliação Contínua	7
2.1.5	Avaliação Formativa	7
2.1.6	Avaliação Somativa	7
2.2	AVALIAÇÃO EDUCACIONAL EXTERNA	8
2.2.1	PISA	8
2.2.2	IDEB	9
2.2.3	SAEB e PROVA BRASIL	9
2.2.4	SARESP	11
2.2.5	ENEM	11
2.3	TEORIA CLÁSSICA DE TESTE	13
2.4	TEORIA DE RESPOSTA AO ITEM	14
2.5	CONSTRUÇÃO DE ITENS	15
2.5.1	Questões de Múltipla Escolha - QME em TRI	16
3	VALIDAÇÃO PEDAGÓGICA	19
3.1	TAXONOMIA DE BLOOM	19
3.1.1	Benjamin Samuel Bloom	19
3.1.2	Taxonomia de Bloom nas Estruturas de Avaliação para Mapeamento de Habilidades	19
4	CRONOLOGIA DAS ESTRUTURAS DE AVALIAÇÃO	29
4.1	ETAPA I	29
4.2	ETAPA II	31

4.2.1	Nível I	31
4.2.2	Nível II	32
4.2.3	Nível III	32
4.3	ETAPA III	35
5	RESULTADOS OBTIDOS	37
5.1	Metodologias Estatística	37
5.1.1	Teste de Friedman	37
5.1.2	Teste de Wilcoxon para dados emparelhados	37
5.2	Comparações das questões - Descritivos	38
5.2.1	Questões da disciplina de Matemática	38
5.2.2	Questões da disciplina de Biologia	39
5.2.3	Questões da disciplina de Química	40
5.2.4	Questões da disciplina de Física	41
5.2.5	Questões - Geral	42
6	DISCUSSÃO	45
7	CONCLUSÃO	53
7.1	ALUNOS	53
7.2	PROFESSORES	54
7.3	COORDENADORES	54
7.4	GESTÃO	54

1

INTRODUÇÃO

O trabalho “Estruturas de Avaliação escolar para mapear habilidades tomando como base a Taxonomia de Bloom em questões de múltipla escolha”, foi desenvolvido com o paralelismo de aspectos da Educação Matemática observando o Currículo/Avaliação Bimestral e por métodos Estatísticos, garantindo um alto nível de confiabilidade dos resultados apresentados. Tentamos demonstrar através de testes estatísticos a validade das técnicas de mapeamento sugeridos na formatação, aplicação e devolutiva das avaliações feitas na unidade escolar EE Prefeito Nestor de Camargo - Barueri. O prazo da devolutiva dos resultados foi também priorizado, pois nada adiantaria um resultado com alto nível de confiabilidade sendo entregue após um ou dois meses da realização de Avaliações Bimestrais. De modo concomitante, vimos as técnicas de mapeamento discutindo a dependência de habilidades na formatação das questões dos temas dados utilizando conceitos da Taxionomia de Bloom[B1].

Existem diversos tipos de Avaliações, e nesta pesquisa analisamos apenas avaliações objetivas (Questões de Múltipla Escolha - QME), descrevendo a sua elaboração com as mesmas orientações para se construir ítems em Teoria de Resposta ao Item (TRI) [Ra]. Mantivemos grande foco na aplicação digital, na correção instantânea e principalmente, na devolutiva dos resultados para todos os envolvidos.

A decorrente busca de resultados em avaliações externas nos faz refletir e tentar correlacionar junto às Avaliações Internas meios de influenciar beneficemente o aprendizado de nossos alunos. Hoje as avaliações externas de larga escala como: PISA, SAEB, Prova Brasil e o ENEM em nível Nacional e o SARESP em nível Estadual, tornaram-se referências em devolutiva com um alto padrão de confiabilidade, sendo realizadas com a metodologia da TRI. Porém, essas avaliações são realizadas apenas uma vez por ano (ou a cada dois ou três anos, sendo que as vezes é observado apenas uma amostra), e as devolutivas, devido a suas características de correção, são entregues em longo prazo. Além disso, suas intervenções propostas para melhora desses resultados não atingem 100% da população dos avaliados e não mostram as defasagens individuais. Portanto,

seria de suma importância equiparar as avaliações bimestrais com as avaliações externas, construindo itens para as provas bimestrais com as mesmas normatizações das provas externas.

As habilidades presentes nessas avaliações externas, após serem feitos todos os processos de correção e análise da TRI, são classificadas em níveis de complexidade de acordo com sua régua de proficiência, sendo eles: Abaixo do Básico, Básico, Adequado e Avançado. Para adquirirmos maturidade para construção de nossas avaliações, analisamos os resultados fornecidos no relatório pedagógico de matemática dos SARESPs realizados nos anos de 2008 a 2013. Neles observamos que os exercícios (itens) com características de reconhecimento se inseriam nos níveis Abaixo do Básico/Básico e os exercícios com características de compreensão, em sua maioria, se inseriam nos níveis Adequado/Avançado. Analisamos também antigas avaliações internas da área de Ciências da Natureza e Matemática dos anos de 2012, 2013 e primeiro Bimestre de 2014, e observamos que as questões tinham características de resolução mecânica com muitos termos e aspectos “livrísticos”.

Na escola EE Pref. Nestor de Camargo - Barueri, SP, desde o início de 2012, aplicávamos Avaliação Bimestral contendo quarenta questões divididas igualmente nas quatro disciplinas de Ciências da Natureza e Matemática. Já na Avaliação Bimestral do segundo bimestre de 2014, conseguimos elaborar itens com um pouco mais rigor nas características da Teoria Clássica de Testes (TCT), propondo também um aspecto formativo e observando Habilidades e Competências em torno de nove questões relacionadas aos conteúdos desenvolvidos no Bimestre de cada uma das quatro disciplinas da área, procurando mensurar o que o aluno assimilou no Bimestre.

A avaliação passou a ter trinta e seis questões. Cada disciplina atendeu todo o conteúdo programado no Bimestre agrupado em três temas/assunto trabalhados. De cada tema o professor forneceu três questões com níveis e grau de complexidade distintos, sendo a primeira questão de reconhecimento, a segunda com aspecto mecânico e a última de compreensão. Dessa forma, pudemos comparar os resultados observando um gráfico de cores, construído com a dependência hierárquica de conhecimentos relacionados na Taxonomia de Bloom dentre as três questões.

Os recursos tecnológicos foram essenciais para realização dessa empreitada. No final de 2012 recebemos na Escola 120 Net books, para uso diverso voltado para educação e aprendizado. No início de 2013 pensamos em utilizar os Nets para melhorar o processo de nossa Avaliação Bimestral, que continham quarenta questões das quatro disciplinas.

Num primeiro momento estudamos uma forma de correção automática com o uso da tecnologia. Fizemos uma avaliação on-line no Formulário Google e uma planilha de correção no software Excel do pacote Office da Microsoft. Nossa primeira intenção foi dinamizar o processo de correção, pois, mesmo sendo uma Avaliação Objetiva, demandava muito tempo para devolver os resultados dos mais de trezentos alunos matriculados. A primeira avaliação com correção automática foi realizada no segundo Bimestre de 2013.

Como dito antes, as avaliações até o primeiro bimestre de 2014 tinham aspectos de Avaliação Somativa, tendo como principal objetivo obter rapidamente a nota do aluno. A partir do segundo bimestre de 2014 começamos a nos preocupar com Habilidades e Competências. Ou seja, direcionamos para construção de Avaliação Formativa. Como iremos descrever adiante, a devolutiva obtida foi um gráfico de cores, e sua interpretação forneceu o mapeamento individual e coletivo, permitindo-nos analisar agrupamentos e comparar salas e séries distintas.

No ano de 2015 automatizamos todo o processo com o uso dos aplicativos Google *Formulário e Planilha Google*. No decorrer da leitura deste trabalho veremos cada etapa dessa proposta com mais detalhes.

2 | AVALIAÇÃO

De modo técnico, avaliação é um processo mental de julgamento e classificação de tipos em escalas gradientes, a partir de parâmetros positivos de satisfação, excelência, eficiência e desempenho realizados com base nos valores morais, éticos e estéticos de quem executa a ação (Dicionário informal - on-line). Entretanto, a definição de avaliação vai muito além do que mencionaremos nesta pesquisa. Dessa forma, para obtermos o resultado esperado utilizaremos algumas das diversas formas de se avaliar. Pensando em Avaliação na instituição de ensino, com o conceito acima, para aperfeiçoarmos o ato de avaliar deveremos caracterizar e explicar os parâmetros escolhidos, os avaliados devem saber como se preparar e quais serão os valores estéticos para o seu melhor desempenho. De acordo com SANT'ANNA, (1997, p. 31), *“Avaliação é um processo pelo qual se procura identificar, aferir, investigar e analisar as modificações do comportamento e atendimento do aluno, do educador, do sistema, confirmando se a construção do conhecimento se processou, seja este teórico (mental) ou prático”*. Devemos distinguir com clareza a diferença entre medir e avaliar. Ao tratar-se de pessoas e da definição de seu futuro, temos por obrigação nos responsabilizar pelo processo de sua evolução. Por exemplo, para medir a altura de uma pessoa precisaremos de apenas um único recurso e obteremos um resultado exato. Isto é completamente diferente de medir a capacidade intelectual, onde acreditamos que necessitaríamos de várias ferramentas bem estruturadas, possibilitando-nos aproximar de um valor plausível desta medida.

Segundo as Diretrizes Curriculares Nacional (DCN, 2013), a concepção de avaliação vai além da visão tradicional, que focaliza o controle externo do aluno mediante a notas ou conceitos, para ser compreendida como parte intrínseca ao processo educacional. A avaliação deve ser compreendida como um conjunto de ações que tem a função de alimentar e orientar a intervenção pedagógica.

Verifiquemos um breve resumo de alguns tipos de avaliações existentes para expormos a nossa proposta de avaliação:

2.1 AVALIAÇÃO ESCOLAR INTERNA

Para mensurar o aprendizado do aluno em nossas escolas, a instituição escolar e os professores normalmente utilizam de diversos tipos de avaliação. Superficialmente, comentaremos alguns a seguir.

2.1.1 Avaliação Diagnóstica

É a avaliação de entrada, que tem por objetivo verificar e aferir o nível de conhecimento e de dificuldades que um indivíduo e/ou uma população possui em determinado tema ou fase. Os resultados da avaliação diagnóstica nos permitem traçar estratégias direcionadas ao aprendizado e no aprofundamento dos conhecimentos do indivíduo, de determinado grupo ou da população total [Ho]. A avaliação diagnóstica serve para o professor obter informações necessárias para propor atividades e gerar novos conhecimentos. Em paralelo, o aluno consegue ter consciência do que ele aprendeu e do que poderá aprender de determinado conteúdo.

2.1.2 Avaliação Discursiva

É a avaliação em que o aluno demonstra suas ideias através de textos ou cálculos. Sua correção normalmente não é tão rápida, pois, dificilmente será feita de modo computadorizado [Ho].

2.1.3 Avaliação Objetiva

É a avaliação constituída por questões de múltipla escolha (QME), São úteis para avaliar uma grande extensão de conhecimentos e habilidades. Tem a grande vantagem de poder computadorizar as respostas, obtendo a correção dos testes de modo dinâmico e

rápido. Para construir questões de múltipla escolha (QME) deve-se tomar cuidado com suas normatizações e orientações, principalmente na elaboração dos distratores [Ra]. Existe um ponto de atenção neste tipo de avaliação: o avaliado pode obter pontuação com acerto aleatório, ou seja, acertar a questão sem ao menos ter lido a mesma. O aprofundamento referente à elaboração de questões, está contido no subcapítulo: 2.5 - *Elaboração de itens*.

2.1.4 Avaliação Contínua

É a avaliação de todo o processo. Ocorre diariamente observando principalmente as atitudes do estudante, sua participação oral, escrita, em grupo, etc. Neste tipo de avaliação, o registro do professor é fundamental para o êxito e validade do resultado[Lu].

2.1.5 Avaliação Formativa

Muito parecida com a avaliação diagnóstica, com a diferença de ter todo o seu processo contínuo. Se compromete com o desenvolvimento das capacidades dos alunos, expressando-se na profundidade de saberes constituídos. A avaliação formativa, prioritariamente, deve nos trazer indicadores da aprendizagem do aluno e do ensino oferecido, permitindo conhecer bem os saberes, as atitudes, a capacidade e o estágio de desenvolvimento dos alunos, proporcionando indicações claras do que se deve fazer para prosseguir[DCNs].

2.1.6 Avaliação Somativa

É o tipo de avaliação onde é observado e somado apenas os acertos. Nas Escolas Estaduais do Estado de São Paulo, na **Resolução SE - 61, de 24-9-2007**, no **Art. 1º**, diz: *Nas escolas da rede estadual de ensino, o registro das sínteses bimestrais e finais dos resultados da avaliação do aproveitamento do aluno, em cada componente curricular, será efetuado em escala numérica de notas em números inteiros de 0 (zero) a 10 (dez)(...)*, ficando padronizado. A avaliação somativa, se resume em fragmentar os dez pontos que podem ser obtidos no bimestre em diversas e distintas avaliações contínuas, dentre, qualitativas

e quantitativas. Após a realização dessas avaliações, são somadas todas as notas obtidas em cada avaliação aplicada no Bimestre totalizando 10.

2.2 AVALIAÇÃO EDUCACIONAL EXTERNA

Abaixo comentaremos algumas das avaliações externas que são aplicadas no Brasil e/ou aplicadas no Estado de São Paulo.

2.2.1 PISA

Segundo o INEP, PISA, *Programme for International Student Assessment* - Programa Internacional de Avaliação de Estudantes - é uma iniciativa de avaliação comparada, aplicada a estudantes na faixa dos 15 anos, idade em que se pressupõe o término da escolaridade básica obrigatória na maioria dos países. Esta avaliação, no Estado de São Paulo é aplicada por amostragem.

O programa é desenvolvido e coordenado pela Organização para Cooperação e Desenvolvimento Econômico (OCDE). Em cada país participante há uma coordenação nacional. O Brasil é um dos países que participam dessa avaliação, sendo aqui coordenado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). O objetivo do PISA é produzir indicadores que contribuam para a discussão da qualidade da educação nos países participantes, de modo a subsidiar políticas de melhoria do ensino básico. A avaliação procura verificar até que ponto as escolas de cada país participante estão preparando seus jovens para exercer o papel de cidadãos na sociedade contemporânea.

Além de observar as competências dos estudantes em Leitura, Matemática e Ciências, o PISA coleta informações para a elaboração de indicadores contextuais, os quais possibilitam relacionar o desempenho dos alunos a variáveis demográficas, socioeconômicas e educacionais. Essas informações são coletadas por meio da aplicação de questionários específicos para os alunos, para os professores e para as escolas.

O PISA utiliza a Teoria de Resposta ao Item na análise dos dados da avaliação. Os resultados desse estudo podem ser utilizados pelos governos dos países envolvidos como instrumento de trabalho na definição e refinamento de políticas educativas, procurando tornar mais efetiva a formação dos jovens para a vida futura e para a participação ativa

na sociedade.

PISA é único porque ele desenvolve testes que não estão diretamente relacionadas com o currículo escolar. Normalmente existe uma matriz de referência que atende competências e habilidades da prática do cotidiano. Os testes foram projetados para avaliar em que medida o aluno no final do ensino obrigatório pode aplicar seu conhecimento para situações da vida real e ser equipado para a plena participação na sociedade.

2.2.2 IDEB

O IDEB (Índice de Desenvolvimento da Educação Básica) foi criado em 2007 e para efetuar o seu cálculo reúnem-se dois indicadores:

- I. o fluxo (taxa de aprovação dos alunos por escola);
- II. o nível de proficiência (taxa de desempenho nos testes).

Assim, obtemos o indicador para a verificação do cumprimento das metas fixadas no Termo de Adesão ao Compromisso *Todos pela Educação*, eixo do PDE (Plano de Desenvolvimento da Educação) do Ministério da Educação, que trata da educação básica. É nesse âmbito que se enquadra a ideia das metas intermediárias para o IDEB. A lógica é a de que, para que o Brasil chegue à média 6,0 em 2021, período estipulado tendo como base a simbologia do bicentenário da Independência em 2022, cada sistema deve evoluir segundo pontos de partida distintos, e com esforço maior daqueles que partem em pior situação, com um objetivo implícito de redução da desigualdade educacional.

2.2.3 SAEB e PROVA BRASIL

O SAEB (Sistema da Avaliação do Ensino Básico) teve sua primeira aplicação em 1990. Tinha a finalidade de conhecer o sistema educacional brasileiro, e foi reformulado em 1995 passando a utilizar a Teoria de Respostas ao Item (TRI), que faz uma correção

populacional baseada em testes probabilísticos e traz a oportunidade de comparação tanto no decorrer dos anos como com a avaliação educacional ocorridas em outros países[INEP]. A experiência obtida no SAEB fez com que essas técnicas de correção da TRI fossem difundidas para outras avaliações. Uma dessas avaliações é o ENEM (Exame Nacional do Ensino Médio), considerada uma das maiores avaliações realizadas no mundo. O SAEB é dividido em duas avaliações: o ANEB (Avaliação Nacional do Ensino Básico) e a ANRESC (Avaliação Nacional do Rendimento Escolar), que é mais conhecida como Prova Brasil. São dois exames complementares que compõem o Sistema de Avaliação da Educação Básica. A Prova Brasil foi aplicada pela primeira vez em 2005. O Sistema Nacional de Avaliação da Educação Básica (SAEB) é realizado pelo Inep/MEC e abrange estudantes das redes públicas e privadas do país, localizados em área rural e urbana, matriculados nas séries finais do ensino fundamental I e ensino fundamental II, ou seja, a antiga, 4^a e 8^a séries, sendo hoje, 5^o e 9^o anos. E também na 3^a série do ensino médio. São aplicadas provas de Língua Portuguesa e Matemática. A avaliação é feita por amostragem e os resultados são apresentados para cada unidade da Federação e para o Brasil como um todo. Em 2013, o SAEB teve uma outra mudança: foi acrescentado uma terceira frente, a ANA (Avaliação Nacional de Alfabetização), direcionada à alunos do 3^o ano do Ensino Fundamental. Essa iniciativa faz parte das medidas do Pacto Nacional da Alfabetização na idade certa. Em resumo, a ANA é uma avaliação censitária observando a qualidade, equidade e eficiência no ciclo de alfabetização [RA].

Ressaltamos que a grande diferença entre ANEB e Prova Brasil é que o ANEB tem avaliação amostral e a Prova Brasil é uma avaliação censitária, ou seja, analisam-se todas as escolas, todos os professores e todos os alunos da rede educacional em estudo. segundo o INEP, *a participação no Saeb e na Prova Brasil é voluntária. Para o Saeb, são feitos sorteios das escolas que irão participar da avaliação. Quanto à Prova Brasil, a adesão é feita pelas secretarias estaduais e municipais de educação. Cabe ressaltar, porém, que o comprometimento dos participantes é fundamental para a qualidade dos resultados apurados, e é fundamental para que a escola ou rede participe para que tenha seu Ideb calculado.*

No ANEB e na Prova Brasil é avaliado itens retirados de uma matriz de referência que contemplam habilidades essenciais de cada etapa do ensino básico, com o porém, de poder ser avaliado em testes de múltipla escolha, e tendo descritores e contextos que expressam o ensino-aprendizado adquiridos pelo aluno, representados em ações e operações mentais que possibilitem analisar a interpretação qualitativa dos níveis de

proficiência atingidos após os testes.

2.2.4 SARESP

O SARESP (Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo) é uma avaliação externa aplicada de modo censitário pela Secretaria de Educação do Estado de São Paulo. Pois, no Artigo 2º da RESOLUÇÃO SE Nº 27, DE 29 DE MARÇO DE 1996 - afirma que: *O Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo abrangerá todas as escolas da rede estadual e as redes municipal e particular que aderirem à proposta, contemplando, de forma gradativa e contínua (...). Avaliando o todo, temos melhorias no monitoramento e na elaboração de planos e metas para o ensino das escolas públicas visando à melhoria da qualidade da educação. Essa avaliação é aplicada anualmente desde 1996 e têm por objetivo avaliar o Ensino Básico da rede estadual em larga escala, fazendo uso de procedimentos metodológicos formais e científicos cada vez mais aprimorados para coletar, sistematizar dados e produzir informações sobre o desempenho dos alunos. As avaliações dos SARESP atendem o ensino fundamental e também o ensino médio. Na edição de 2007 o SARESP passou a utilizar a metodologia dos exames nacionais SAEB, Sistema de Avaliação da Educação Básica e Prova Brasil, o que permitiu a comparação de resultados, que são utilizados para calcular o IDESP (Índice de Desenvolvimento da Educação do Estado de São Paulo). O resultado da prova, combinado ao fluxo de alunos de cada escola, ajuda a compor o IDESP, indicador que avalia a qualidade da educação no Estado. As mudanças foram introduzidas ao SARESP nessa edição para permitir um melhor acompanhamento da evolução da qualidade do sistema estadual de ensino ao longo dos anos. No Brasil o desempenho dos alunos na educação básica tem sido medido por meio da métrica do SAEB. A escala de proficiência permite a comparação dos resultados entre o SARESP, SAEB e a Prova Brasil.*

2.2.5 ENEM

O Exame Nacional do Ensino Médio (ENEM) foi criado em 1998 com o objetivo de avaliar o desempenho do estudante ao fim da educação básica, buscando contribuir para a melhoria da qualidade desse nível de escolaridade. Tem como eixos estruturadores

interdisciplinaridade e contextualização, aferindo principalmente as competências e habilidades obtidas no exame.

O ENEM vem cumprindo o seu papel de avaliar o ensino médio brasileiro, desde o seu início em 1998. Nos questionários apontavam que 70% dos participantes que realizavam a prova almejavam ingressar no ensino superior. Então, uma das propostas de unificar o processo de seleção para Universidades Federais de Ensino Superior foi atendida com a reformulação do ENEM a partir de 2009, possibilitando também a certificação de jovens e adultos. Tiveram várias mudanças no exame. A principal em seu processo de correção, passando a ser utilizado a Teoria de Resposta ao Item (TRI), onde a pontuação do examinado é obtido após análises psicométricas, estatísticas e probabilísticas de toda a população que realizou a avaliação, só sendo possível com o uso de altos recursos de informática. A contagem é diferente das avaliações somativas, que contam os acertos. Na TRI, o item (a questão) mede níveis de conhecimentos diferentes. Há perguntas com diferentes níveis de complexidade, com pontuações diferentes. Um dos objetivos desse modelo de correção é conseguir separar quem sabe o conteúdo de quem tenta acertar ao acaso. Tanto é que, diversas pessoas que acertam o mesmo número de questões não têm notas iguais, porque acertaram e erraram questões diferentes. Um outro objetivo é garantir que as provas do Enem em anos distintos tenham o mesmo grau de dificuldade e possam ser comparadas. A partir de 2009, o exame também passou a ser utilizado como mecanismo de seleção para o ingresso no ensino superior.

No ENEM, é avaliado itens presentes na Matriz de referência de determinadas áreas do conhecimento. São quatro áreas no total:

- i. Linguagens, códigos e suas tecnologias.*
- ii. Ciências Humanas e suas Tecnologias.*
- iii. Ciências da Natureza e suas Tecnologias*
- iv. Matemática e suas tecnologias.*

Para cada uma das quatro áreas foi organizado um conjunto de amplas competências a serem avaliados, desdobrados nos cinco eixos cognitivos (Dominar Linguagens; Compreender Fenômenos; Enfrentar situações problemas; Construir Argumentação e Elaborar Propostas) separados em habilidades específicas. O ENEM obedece as mesmas características do SAEB, até mesmo a régua dos níveis de proficiência. Na avaliação do ENEM o valor da proficiência pode chegar até a 1000 pts. As situações problemas dadas em testes de múltipla escolha pretendem avaliar os saberes cognitivos e conceituais, ou seja, as competências que tomam como referência participação social incluindo criativi-

dade, senso crítico, capacidade de solucionar problemas e domínio de conhecimentos.

2.3 TEORIA CLÁSSICA DE TESTE

Os esboços da Teoria Clássica de Testes (TCT) surgiram por volta de 1880, na análise de instrumentos para testes psicométricos, em pesquisas referentes ao fenômeno psicológico em avaliações objetivas. Desses estudos desenvolveram-se diversas fases da TCT, e diversos estudiosos deram sua contribuição a esta Teoria. Entre eles temos Galton (1880), Cattell (1890), Binet (1900) e Spearman. Tivemos algumas "eras" que direcionaram estes estudos, como a era de Binet, após os trabalhos de Spearman relacionados a correlação, a era dos testes de inteligência (1920 - 1930) e a era da sistematização e da análise fatorial (1940 - 1980).

Um dos objetivos da Teoria Clássica de Testes é ver em perspectiva multidisciplinar e em variados contextos indicadores que possam cooperar no direcionamento de intervenções para o indivíduo e sua população após os testes. A TCT analisa dois princípios fundamentais: *i.a validade e ii. a confiabilidade dos testes*. Desde a era de Binet, já havia sido proposto a maioria dos métodos operacionais para sua análise. A validade de um teste pode ser definida como sua capacidade de realmente medir aquilo a que se propõe a medir, A confiabilidade de um teste se refere à reprodutibilidade da medida, ou seja, o grau de concordância entre múltiplas medidas de um mesmo sujeito inter e intra-indivíduos.

Desde 1980, com a evolução tecnológica da informática, entramos na era da psicometria moderna, onde foi adotado a Teoria de Resposta ao Item (TRI), que será discutido posteriormente no próximo subtítulo. Os estudos da Teoria Clássica de Teste, ainda permeiam até os dias atuais. Mesmo sendo constatado a insuficiência de informações fornecidas, vale ressaltar que a TCT ainda tem grande importância nos estudos psicométricos.

2.4 TEORIA DE RESPOSTA AO ITEM

A Teoria de Resposta ao Item (TRI) é um modelo proposto para medidas psicométricas obtidas através de testes e questionários. Foi proposto no início da década de 1950 por Frederick Lord, que apresentou os modelos teóricos para estimar os parâmetros dos itens. Lord utilizou testes dicotômicos e binários do tipo certo e errado. No início da década de 1970, Samejima generalizou o modelo para itens com respostas politômicas. Esse modelo contém muitos cálculos de estimação de parâmetros, e seria impossível realizá-los sem o uso de computadores. Apenas na década de 1980 com o avanço tecnológico da informática, a TRI teve um alcance maior nas suas aplicações. No Brasil, a TRI começou a ser utilizada em 1995 na correção das avaliações do SAEB, posteriormente sendo utilizada também em outras avaliações [Ra].

A Teoria de Resposta ao Item é um conjunto de modelos matemáticos que procuram representar a probabilidade de um indivíduo dar uma resposta certa a um item como função dos parâmetros do item e da habilidade (ou habilidades) do respondente. Os resultados da TRI têm a grande importância de fornecer resultados individualizados e resultados populacionais em torno dos itens propostos.

As principais características que diferenciam a TRI da TCT são:

- i. foco da avaliação ser os itens e não a prova como um todo;*
- ii. a proficiência estimada pelos métodos estatístico e não a pontuação obtida (score);*
- iii. indivíduos e itens são colocados em uma escala comum mesmo que submetidos a provas diferentes.*

Em resumo, uma das grandes vantagens da TRI sobre a TCT é que ela permite a comparação entre populações, desde que submetidas a provas que tenham alguns itens comuns, ou ainda, a comparação entre indivíduos da mesma população que tenham sido submetidos a provas totalmente diferentes. Isto porque uma das principais características da TRI é que ela tem como elementos centrais os itens, e não a prova como um todo.

Para se obter êxito nos resultados da TRI, o momento de construir itens para a avaliação é um ponto crítico, e tem que se seguir uma série de recomendações. Essas recomendações serão descritas na seção seguinte.

Em um teste, a Teoria de Resposta ao Item busca avaliar a aptidão que o avaliado possui para desenvolver o exercício, ou seja, o traço latente do indivíduo através de um conjunto de questões. No Brasil, tomando como exemplo o ENEM e o SAEB, o modelo

matemático mais utilizado é o de três Parâmetros. Este modelo consegue relacionar as variáveis nessas situações desejando saber na individualidade o traço latente de cada habilidade do indivíduo tentando explicar os devidos erros e acertos. Assim, a TRI estuda o comportamento do aluno frente a cada item que ele respondeu. Vejamos o três parâmetros:

- Acerto ao acaso – é a probabilidade de acerto de modo aleatório, conhecido também como "acerto no chute". Normalmente ocorre quando o aluno de baixo traço latente acerta item classificado como difícil. A TRI estima a probabilidade desse fato ter ocorrido.
- Dificuldade do item - Pode ser definido como o nível mínimo de proficiência que um candidato deve obter para ter grande chance de acertar um item. De modo específico, deseja-se saber o valor de aptidão para que a probabilidade de acerto deste ítem seja de no mínimo 50%.
- Discriminação do item - É concebido como a capacidade do item de diferenciar indivíduos com habilidades e proficiências distintas.

2.5 CONSTRUÇÃO DE ITENS

Ainda temos como grande problema a construção de um caderno de prova para avaliação. Os aspectos que mais causam incoerência são voltados para elaboração de questões que atinjam os propósitos de avaliar o aprendizado do aluno estabelecido nas orientações pedagógicas na escola ou instituição que atuam. Em uma avaliação com o formato de devolutiva da TRI, os aspectos precisam ser bem definidos para que não haja discordâncias nos indicadores obtidos, principalmente se a discordância for causada devido a influência de itens elaboradas de modo errôneo. Em TRI, os aspectos devem ser bem definidos tomando como base os princípios de competências e habilidades presentes em documentos legais que regulam os diversos níveis de educação no Brasil. Assim, se torna obrigatório elaborar questões que permitam avaliar

habilidades que desenvolvem uma competência específica, definindo como avaliar e o que avaliar, observando os conhecimentos estudados em sala de aula. Em TRI, os itens podem ser do tipo múltipla escolha, certo ou errado, resposta fechada, resposta curta aberta ou construída e dissertativa (redação). Devido o foco deste projeto comentaremos apenas as sugestões de elaboração dos itens de múltipla escolha.

2.5.1 Questões de Múltipla Escolha - QME em TRI

Um item elaborado no modelo de QME tem a vantagem de suas respostas ter a possibilidade de serem interpretadas por softwares que permitem uma análise e correção computacional. Neste caso também deve atender as especificações da TRI, sendo a sua construção dividida em três partes:

- **Texto-base** - Nesta parte podemos contextualizar uma questão tomando o cuidado para não apenas utilizar o texto-base como um pretexto para solução ou uma textualização para retirada de informações. Então, para contextualizar um item, podemos inserir na questão: figura, tabela, gráfico, texto, etc. A escolha do texto-base é fundamental na elaboração de um item. Preferencialmente é recomendável utilizar textos, gráficos, etc., de fontes primárias, sendo curto, integral e de fácil compreensão. Lembrando que é de suma importância adequar a linguagem ao nível dos estudantes. São permitidos recortes de um texto, retirar imagens, figuras e gráficos de canais de comunicação e fazer adaptações, tendo a obrigatoriedade de manter a idéia central, indicar se houve adaptações e citar fonte de referência no modelo exigido pelas normas.
- **Enunciado** - É o comando, ou seja, a explicitação do desafio para que se evidencie o desenvolvimento da competência avaliada. O comando poderá ser formulado como uma pergunta direta ou uma frase incompleta.
- **Alternativas** - Uma única das alternativa deverá ser o Gabarito ou também conhecido como chave de resposta, sendo inequivocadamente correta. As outras alternativas serão incorretas chamadas de distratores. Obrigatoriamente os distratores serão plausíveis ao enunciado, e de modo algum representar um absurdo extremo ou induzir o aluno ao erro, pois um distrator com essas características destoariam dos outros distratores prejudicando o item.

Em resumo, na construção de um item, o elaborador deve construir as três partes: texto-base, enunciado e alternativas com muita clareza, objetividade, originalidade, precisão e impessoalidade, indo direto ao assunto sem delongas, esclarecendo conceitos, evitando incluir gírias, expressões de uso local, expressão de uso pessoal e fala informal. Evitar também utilizar partes ou trechos de livros e apostilas que possam beneficiar pessoas específicas. O item deve admitir uma única interpretação e uma só resposta [Ra]. Exemplo:

1.jpg

QUESTÃO 168 —————

Um carpinteiro fabrica portas retangulares maciças, feitas de um mesmo material. Por ter recebido de seus clientes pedidos de portas mais altas, aumentou sua altura em $\frac{1}{8}$, preservando suas espessuras. A fim de manter o custo com o material de cada porta, precisou reduzir a largura.

A razão entre a largura da nova porta e a largura da porta anterior é

Alternativas

A $\frac{1}{8}$

B $\frac{7}{8}$

C $\frac{8}{7}$

D $\frac{8}{9}$

E $\frac{9}{8}$

Diagram annotations:

- Arrow pointing to the text: Texto - base
- Arrow pointing to the question: Interrogação

Figura 1: ENEM Amarelo 2014

3

VALIDAÇÃO PEDAGÓGICA

Para dar consistência teórica ao gráfico de mapeamentos, apresentaremos neste capítulo as Taxonomias de Bloom, que trará base pedagógica para as análises das avaliações que serão desenvolvidas. Nossa avaliação terá características: Diagnóstica; Formativa e Contínua.

3.1 TAXONOMIA DE BLOOM

3.1.1 Benjamin Samuel Bloom

Benjamin Samuel Bloom (1913-1999), nascido em Lasfordf, Pennsylvania, foi um psicólogo educacional americano que se preocupou em analisar conceitos psicométricos na classificação de objetivos educacionais. Ele também dirigiu uma equipe de pesquisa que realizou uma grande investigação sobre o desenvolvimento de talentos excepcionais. Em 1956, Bloom editou o primeiro volume de Taxonomia de objetivos educacionais: a classificação dos objetivos educacionais, que delineou uma classificação dos objetivos de aprendizagem que veio a ser conhecido como a Taxonomia de Bloom e continua a ser um elemento fundamental e essencial dentro da comunidade educativa. Algumas de suas obras estão listadas nas referências.

3.1.2 Taxonomia de Bloom nas Estruturas de Avaliação para Mapeamento de Habilidades

A Taxionomia de Bloom (Taxonomy of Educational Objectives: The Classification of Educational Goals) ou *Taxonomia de objetivos educacionais: A Classificação de Metas Educativas*, proposta e direcionada por Benjamin Samuel Bloom foi executada por ele e uma comissão de pesquisadores em universidades americanas. Iniciou-se em

meados do ano de 1950 projetando a classificação hierárquica de domínio psicométricos dependentes entre si, divididas em três grandes grupos:

- O cognitivo, abrangendo a aprendizagem intelectual;
- O afetivo, abrangendo os aspectos de sensibilização e gradação de valores;
- O psicomotor, abrangendo as habilidades de execução de tarefas que envolvem o aparelho motor.

Essa taxonomia será uma das principais ferramentas pedagógicas e educativas contidas neste Trabalho. Assim, nos restringiremos apenas às informações do domínio cognitivo, observando a hierarquia taxonômica de complexidade em aprendizagem intelectual em construção de questões objetivas, e principalmente para validação do gráfico de cores que expõe a devolutiva da assimilação de alunos no período estudado.

Segundo a Taxonomia de Bloom, os objetivos educacionais no domínio cognitivo são classificados em uma hierarquia de seis níveis: Conhecimento (lembrar), Compreensão (entender), Aplicação, Análise, Síntese e Criação. Esses seis níveis são ilustrados nas seguintes etapas de demonstração de aquisição e devolução cognitiva.

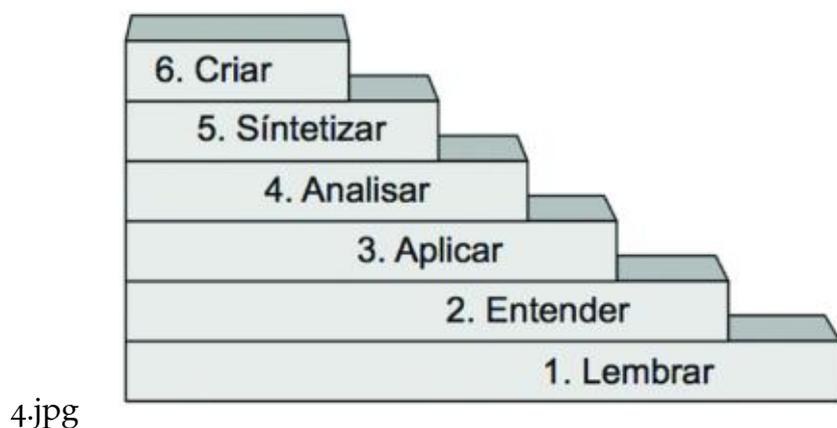


Figura 2: Tabela Hierarquica na escala de aprendizado

Fonte: Imagem retirada de: http://www.scielo.br/scielo.php?script=sci_arttextpid=S0104-530X2010000200015

Apresentaremos cada patamar da Taxonomia de Bloom, exemplificando com um modelo de exercício aplicado em nossas avaliações. Classificaremos, o nível de comple-

xidade de cada exercício, porém, explicaremos os níveis de complexidade no capítulo posterior: *Cronologia das Estruturas de Avaliação*

1. Lembrar / Conhecer: Habilidade de lembrar informações e conteúdos previamente abordados como fatos, datas, palavras, teorias, métodos, classificações, lugares, regras, critérios, procedimentos etc. Esta habilidade pode envolver lembrar uma significativa quantidade de informação ou fatos específicos. O objetivo principal desta categoria é trazer à consciência esses conhecimentos.

Na construção de nossa avaliação tomamos muito cuidado na elaboração de questões de múltipla escolha para não contrastar com as normatizações de construção de itens e privilegiar as habilidades contidas no conhecimento. As questões que denominamos de Reconhecer enfatizam as subcategorias do conhecer, abrangendo: Conhecimento específico; Conhecimento de terminologia; Conhecimento de formas e significados relacionados às especificidades do conteúdo: Conhecimento de convenção; Conhecimento de tendência e sequência; Conhecimento de classificação e categoria; Conhecimento de critério; Conhecimento de metodologia; Conhecimento universal e abstração, relacionado a um determinado campo de conhecimento; Conhecimento de princípios e generalizações; Conhecimento de teorias e estruturas.

Na construção de itens, nos preocuparemos em utilizar os seguintes verbos para caracterizar as questões de reconhecer. São eles: enumerar, definir, descrever, identificar, denominar, listar, nomear, combinar, realçar, apontar, relembrar, recordar, relacionar, reproduzir, solucionar, declarar, distinguir, rotular, memorizar, ordenar e reconhecer.

Acreditamos que o aluno que não reconhece (lembra de) um tema, não tenha condições de acertar exercícios mais complexos, e no caso de acerto em QME haveria grande possibilidade de ser um acerto ao acaso. Isto resultará na configuração das cores em nossos gráficos seguindo sua ordem hierárquica.

Exemplo 1:

(Av. Diagnóstica - Nestor de Camargo) *Um triângulo, devido suas características pode ser acutângulo (todos os ângulos $< 90^\circ$), retângulo (um ângulo $= 90^\circ$) ou obtusângulo (um ângulo $> 90^\circ$). Podemos determinar sua característica utilizando o critério do lado maior (Lm) e seus outros dois lados (a e b). Vejamos: acutângulo: $(Lm)^2 < a^2 + b^2$; retângulo: $(Lm)^2 = a^2 + b^2$ e obtusângulo: $(Lm)^2 > a^2 + b^2$. Pergunta-se: Os lados de um triângulo ABC medem 5cm, 12cm e 13cm. Reconhece-se que:*

- a. *Esse triângulo é retângulo*
- b. *Esse triângulo é acutângulo*
- c. *Esse triângulo é obtusângulo*
- d. *NÃO é possível responder!*

Este exercício de *Nível 1 - Reconhecer/Identificar*, foi aplicado em uma das avaliações na escola. Esta questão, foi utilizada em uma avaliação diagnóstica aplicada para todas as séries da escola. Lembrando que, a competência *Teorema de Pitágoras* é habilidade estruturante de conteúdos pertinentes à 1^a série - Trigonometria; à 2^a série Geometria Espacial e à 3^a série Geometria Analítica.

Neste exemplo, vimos um exercício informativo com a prerrogativa de analisar a habilidade de leitura e interpretação matemática verificando se o avaliado relembra e reconhece a relação existente em *triângulo retângulo* e o *teorema de Pitágoras*

2. Entender: Habilidade de compreender e dar significado ao conteúdo. Essa habilidade pode ser demonstrada por meio da tradução do conteúdo compreendido para uma nova forma (oral, escrita, diagramas etc.) ou contexto. Nessa categoria, encontra-se a capacidade de entender a informação ou fato, de captar seu significado e de utilizá-la em contextos diferentes. Normalmente os exercícios de compreensão em QME vêm contextualizados explorando uma ou mais subcategorias, como: Translação; Interpretação ou Extrapolação.

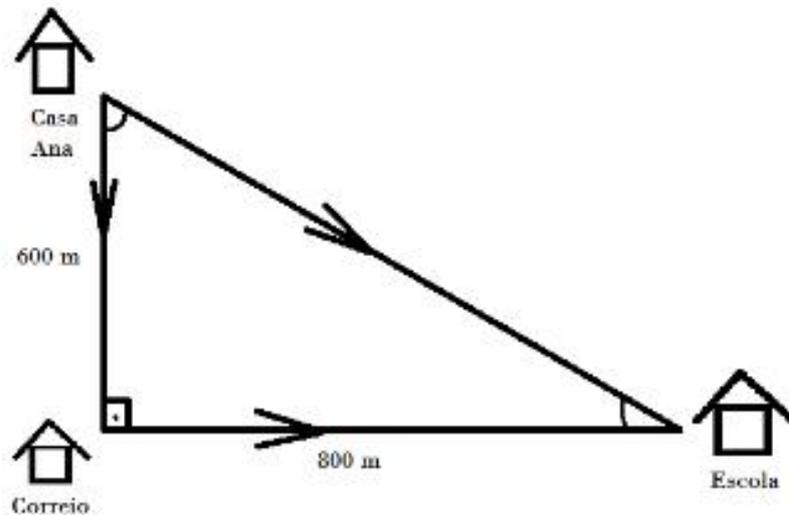
Os verbos utilizados para esta habilidade são: alterar, construir, converter, decodificar, defender, definir, descrever, comparar, distinguir, discriminar, estimar, explicar, generalizar, dar exemplos, ilustrar, inferir, reformular, prever, reescrever, resolver, resumir, classificar, discutir, identificar, interpretar, reconhecer, redefinir, selecionar, situar e traduzir.

Na construção das avaliações notamos que quando utilizamos os verbos acima em questões das disciplinas com aprendizado e foco na leitura e interpretação (Biologia e Química), esta habilidade têm aspectos de processo direto (mecânico) de resolução. Por outro lado, em disciplinas que tem foco maior em cálculo (Matemática e Física), os exercícios tornam-se mais complexos.

Exemplo 2:

(SAEB 2009) *Hélio e Ana partiram da casa dela com destino à escola. Ele foi direto de casa*

para a escola e ela passou pelo correio e depois seguiu para a escola, como mostra a figura a seguir.



2.jpg

Figura 3: Nível 2

De acordo com os dados apresentados, a distância percorrida por Ana foi maior que a percorrida por Hélio em:

- a. 200m
- b. 400m
- c. 600m
- d. 1400m

Este exercício de *Nível 2 - Resolver/Desenvolver*, também foi aplicado na mesma avaliação diagnóstica referida no exemplo anterior.

Observem que foi fornecido os valores dos catetos. Assim, a prerrogativa deste teste foi verificar o desenvolvimento mecânico para determinar o valor da hipotenusa e conferir a distância.

3. Aplicar: Habilidade de usar informações, métodos e conteúdos aprendidos em novas situações concretas. Isso pode incluir aplicações de regras, métodos, mo-

delos, conceitos, princípios, leis e teorias. Como dito antes, deparamo-nos com exercícios de duas características: Nas disciplinas que tem foco maior em cálculo (Matemática e Física), os exercícios ficam com as características de resolução com processo mecânico, pois, em muitos casos apenas substituímos valores em fórmulas e desenvolvemos o cálculo, por outro lado, nas disciplinas com foco em leitura e interpretação, a aplicação seria transladar todo o conhecimento obtido em situações de contextos diversos.

Verbos: aplicar, alterar, programar, demonstrar, desenvolver, descobrir, dramatizar, empregar, ilustrar, interpretar, manipular, modificar, operacionalizar, organizar, prever, preparar, produzir, relatar, resolver, transferir, usar, construir, esboçar, escolher, escrever, operar e praticar. Vimos também que vários verbos repetem-se em duas ou mais escalas hierárquicas, até o estudo atual concluímos que o contexto do exercício e a disciplina influencia em seu grau de complexidade.

Exemplo 3:

(Nestor de Camargo - Av Bim) *Desenvolvendo a expressão $x^2 - 7x + 10 = 0$. obteremos como resultados satisfatórios:*

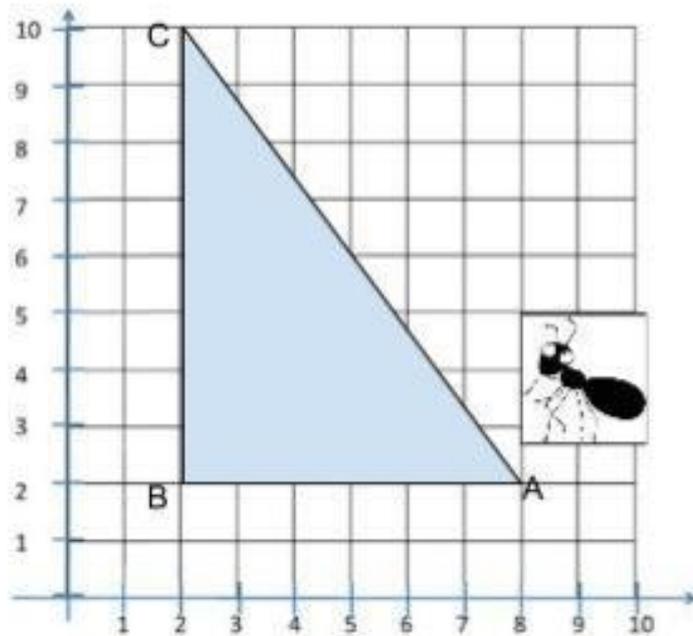
- a. 7 e -10
- b. -7 e 10
- c. 2 e 5
- d. 3 e -3

Este exercício de *Nível 2 - Resolver/Desenvolver* foi aplicado em uma avaliação bimestral. A prerrogativa desta questão é saber se o avaliado consegue utilizar métodos e modelos na aplicação do processo mecânico de desenvolver equação do 2º grau.

4. Analisar: Habilidade de subdividir o conteúdo em partes menores com a finalidade de entender a estrutura final. Essa habilidade pode incluir a identificação das partes, análise de relacionamento entre as partes e reconhecimento dos princípios organizacionais envolvidos. Identificar partes e suas inter-relações. Nesse ponto é necessário não apenas ter compreendido o conteúdo, mas também a estrutura do objeto de estudo.

Poderíamos aqui falar de Subcategorias: Análise de elementos; Análise de relacionamentos; e Análise de princípios organizacionais. Os verbos relacionados a esta habilidade são: analisar, reduzir, classificar, comparar, contrastar, determinar, deduzir, diagramar, distinguir, diferenciar, identificar, ilustrar, apontar, inferir, relacionar, selecionar, separar, subdividir, calcular, discriminar, examinar, experimentar, testar, esquematizar e questionar.

Exemplo 4: (Av. Diagnóstica - Nestor de Camargo) *Na figura abaixo, cada unidade equivale à 1cm. Sendo assim, para a formiga se locomover de A até C em linha reta, percorrerá:*



3.jpg

Figura 4: Nível 3

- a. 8 cm
- b. 10 cm
- c. 12 cm

d. 14 cm

Este exercício de *Nível 3 - Compreender/Aplicar*, Fechou o grupo de três níveis de complexidade para o fechamento da análise do *tema* em estudo.

A prerrogativa deste exercício, é analisar as partes, o avaliado precisa retirar informações do plano cartesiano, para posteriormente efetuar os cálculos da estrutura final, ou seja, o triângulo retângulo.

5. Sintetizar: Habilidade de agregar e juntar partes com a finalidade de criar um novo todo. Essa habilidade envolve a produção de uma comunicação única (tema ou discurso), um plano de operações (propostas de pesquisas) ou um conjunto de relações abstratas (esquema para classificar informações). Combinar partes não organizadas para formar um “todo”.

Poderíamos aqui falar das Subcategorias: Produção de uma comunicação original; Produção de um plano ou propostas de um conjunto de operações; Derivação de um conjunto de relacionamentos abstratos.

Os verbos relacionados a esta habilidade são: categorizar, combinar, compilar, compor, conceber, construir, criar, desenhar, elaborar, estabelecer, explicar, formular, generalizar, inventar, modificar, organizar, originar, planejar, propor, reorganizar, relacionar, revisar, reescrever, resumir, sistematizar, escrever, desenvolver, estruturar, montar e projetar.

Em nossa avaliação, as questões mais complexas podem chegar aos níveis cognitivos quatro e/ou cinco, sendo que esse tipo de questão se aproxima muito das questões do Enem.

Exemplo: (ENEM - 2014) *Uma loja acompanhou o número de compradores de dois produtos, A e B, durante os meses de janeiro, fevereiro e março de 2012. Com isso, obteve este gráfico:*

A loja sorteará um brinde entre os compradores do produto A e outro brinde entre os compradores do produto B.

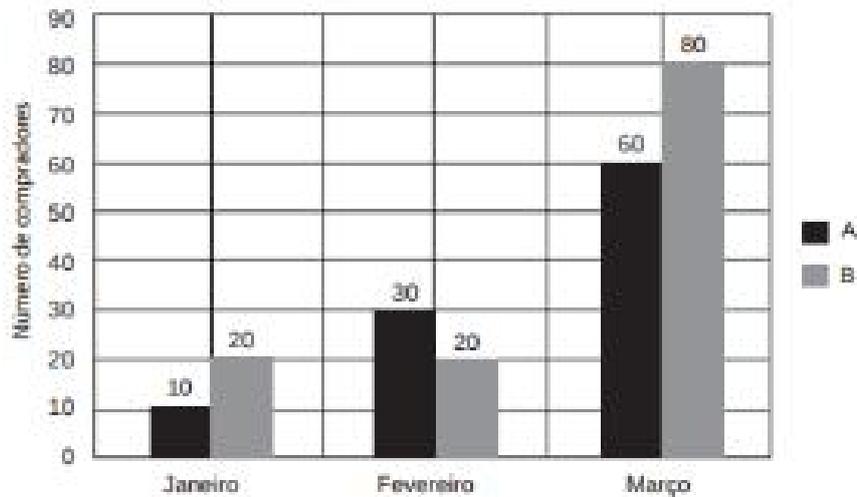


Figura 5: Nível 3

Qual a probabilidade de que os dois sorteados tenham feito suas compras em fevereiro de 2012?

- $\frac{1}{20}$
- $\frac{3}{242}$
- $\frac{5}{22}$
- $\frac{6}{25}$
- $\frac{7}{15}$

Este exercício de *Nível 3 - Compreender/Aplicar*, se insere no nível 5 da Taxonomia de Bloom, devido a união de habilidades para retirar as informações do gráfico e calcular a probabilidade.

- Criar: Habilidade de julgar o valor do material (proposta, pesquisa, projeto) para um propósito específico. O julgamento é baseado em critérios bem definidos que podem ser externos (relevância) ou internos (organização) e podem ser fornecidos ou conjuntamente identificados. Julgar o valor do conhecimento.
Subcategorias: Avaliação em termos de evidências internas; Julgamento em termos

de critérios externos. Verbos: Avaliar, averiguar, escolher, comparar, concluir, contrastar, criticar, decidir, defender, discriminar, explicar, interpretar, justificar, relatar, resolver, resumir, apoiar, validar, escrever um resumo sobre, detectar, estimar, julgar e selecionar.

O nível seis, raramente é encontrado neste tipo de avaliação (QME), porém, é orientado que os professores apliquem em sala de aula em torno de questões discursivas, avaliação contínua e seminários.

4

CRONOLOGIA DAS ESTRUTURAS DE AVALIAÇÃO

Apresentaremos neste capítulo o desenvolvimento das Estruturas de Avaliação escolar para mapear habilidades tomando como base a Taxonomia de Bloom em questões de múltipla escolha, na ordem de seus acontecimentos. O trabalho se iniciou como uma boa prática escolar, e evoluiu com o decorrer do tempo. Dividimos esta evolução em três etapas da pesquisa, que são apresentadas a seguir.

4.1 ETAPA I

No final de 2012, recebemos na Escola 120 Net books, para usos diversos, desde que voltados para a educação e aprendizado. No início de 2013, pensamos em utilizar os Nets para auxiliar em nossa Avaliação Bimestral. Avaliação, essa que continha quarenta questões de quatro disciplinas, a qual anteriormente era executada de modo tradicional, ou seja, construíamos a avaliação e imprimíamos uma para cada aluno. Depois eram corrigidas uma a uma, sendo a nota transportada para uma planilha e compartilhada com os colegas. Todo o processo demorava em torno de sete dias para ser finalizado. Assim, a primeira intenção foi diminuir o tempo de finalização da avaliação, dinamizando o processo de correção. Pois, mesmo sendo uma Avaliação Objetiva, demandava muito tempo para devolver os resultados dos mais de trezentos alunos matriculados. Esse tipo de avaliação perdurou na escola até o primeiro bimestre de 2013.

No segundo bimestre de 2013, construimos a Avaliação completamente on-line utilizando como ferramenta os formulários do Google. Após o aluno enviar a avaliação, as informações se direcionam para uma planilha Google. Então, copiávamos as informações da turma e colávamos em um gabarito preparado no Excel, que fornecia o resultado geral. Devido à quantidade de computadores disponíveis, realizávamos a avaliação em três turmas. A correção eletrônica, tinha o objetivo de mostrar a quantidade de acertos da Avaliação Somatória de modo instantâneo. Conseguimos, efetuar todo o processo

de correção, com a média de cinco minutos. Portanto, as nossas expectativas estavam satisfeitas momentaneamente.

Na realização das provas tivemos alguns inconvenientes, sendo o mais grave a questão da rede, que não suportava a conexão, online, de todos os 120 nets. Resolvemos o problema deixando a avaliação no papel, pedindo para que os alunos digitassem o gabarito. Assim, não precisávamos mais de um computador por aluno, e sim, cinco ou seis computadores por sala, o suficiente para atender a demanda, com a vantagem de uma única aplicação. O resultado final da avaliação era dado na forma apresentada abaixo:

1.jpg

E.E. PREFEITO NESTOR DE CAMARGO - ENSINO MÉDIO INTEGRAL								
RESULTADOS DA AVALIAÇÃO - CNT - 1ª SÉRIE A								
NOME:	Nº:	MATEMÁTICA	FÍSICA	QUÍMICA	BIOLOGIA	ACERTOS	NOTA	ASSINATURA
gabarito	0	10	10	10	10	40	10,0	
0	0	0	0	0	0	0	0,0	
Alessandro de souza lima	1	4	4	7	5	20	6,3	
Aryane conceição monteiro	2	5	2	1	3	11	3,4	
Barbara dos Santos	3	7	7	5	4	23	7,2	
Bruno de souza de	4	4	5	3	4	16	5,0	
Erika De Souza Santos	6	6	5	4	5	20	6,3	
Fabricio da Silva Santana	7	3	3	3	3	12	3,8	
Gabriel Felipe Pereira	8	1	3	2	4	10	3,1	
Gabriella da silva araujo	9	2	5	5	8	20	6,3	
Gabrielle gonçalves alves	10	2	2	1	1	6	1,9	
Geovanna da Silva Santos	11	4	5	2	2	13	4,1	
Guilherme Domingues	12	6	4	6	6	22	6,9	
Gustavo Nascimento	14	2	3	2	2	9	2,8	
Hiago de Jesus Almeida	15	1	5	7	3	16	5,0	
Ivanor Lucas dos Santos	16	6	6	7	7	26	8,1	
Izadora Souza Luna	17	2	4	1	5	12	3,8	
Jefferson Geraldo Prado	19	2	3	3	4	12	3,8	
joyce barbosa correa	20	4	3	2	3	12	3,8	
Julia da silva barbosa	21	7	9	10	9	35	10,0	
Karina Nunes Valentim	22	2	3	4	4	13	4,1	
lucas alves de souza	24	4	3	4	6	17	5,3	
Lucas da silva lima	25	0	0	1	5	14	4,4	
Mariana Ribeiro Gonçalves	27	2	3	7	6	18	5,6	
Nayara Alves Dias	30	7	6	6	6	25	7,8	
paloma (tuti) de oliveira	31	10	9	7	7	33	10,0	
Priscilla Santos Pereira	32	5	4	4	4	17	5,3	
Stefany Pauxão Almeida	34	4	3	6	4	17	5,3	
tatiana cunha da silva	36	5	6	6	6	25	7,8	
victor otavio de oliveira	37	4	1	4	5	14	4,4	
0	0	0	0	0	0	0	0,0	
0	0	0	0	0	0	0	0,0	
0	0	0	0	0	0	0	0,0	

Figura 6: Correção automatizada dividida por disciplina

* Especificamente nesta avaliação, combinamos que, o aluno teria nota máxima (10,0) obtendo 32 ou mais acertos, e nota proporcional as quantidades inferiores à 32

Como podemos ver os resultados não nos traz argumentos para efetuar análises mais aprofundadas, os resultados se restringiam apenas a contagem.

4.2 ETAPA II

A aplicação de Avaliações com resultados quase que instantâneos nos trouxe um dinamismo no tempo de correção. Porém, percebemos que poderíamos explorar ainda mais os recursos que já manuseávamos.

Começamos a analisar as alternativas, observando o percentual de distribuição entre as respostas por questão. Com isso obtivemos ganho no momento da correção, e na devolutiva da Avaliação aos alunos. Conseguimos identificar rapidamente a alternativa que a maioria optou. Caso ela fosse a errada, tínhamos discussões de grande valia com os alunos e principalmente com a equipe de elaboração das questões.

Neste momento nosso objetivo não se restringia em ter a nota do aluno, e quantas questões ele acertou por disciplina. Tínhamos a ambição de retirar mais informações da avaliação. Assim, modificamos completamente o modelo da prova; de Avaliação Somatória passamos a elaborar Avaliação Formativa, pensando na formação do aluno e na mensuração do que ele assimilou no Bimestre. Nossa avaliação passou a ter trinta e seis questões, sendo nove por disciplina. Cada disciplina fechava todo conteúdo em três temas/assunto trabalhado no Bimestre. De cada tema o professor fornecia três questões com níveis de interpretação e resolução distintas, sendo uma questão de cada nível. Exemplificamos modelos do tipo de questão que pode ser elaborada, no capítulo anterior, Taxonomias de Blom (3.1). As informações de complexidade contidas em cada nível são descritas a seguir.

4.2.1 Nível I

Identificação / Observação – Uma questão mais simples possível, contextualizada de modo teórico, ou de um fato cotidiano. O aluno teria somente a necessidade de ter prestado atenção na explicação do professor, ou simplesmente conhecer brevemente sobre do que o assunto tratava. Em linguagem técnica, essa questão se insere no primeiro patamar da Taxonomia de Bloom: Lembrar.

4.2.2 Nível II

Resolução / Desenvolvimento – Uma questão que aborda principalmente o processo mecânico. O aluno deve saber como desenvolver o cálculo ou desenvolver as ideias presentes em seu contexto, mesmo que seja de modo direto. Seria o segundo e o terceiro patamar da taxonomia de Bloom: Entender e Aplicar.

4.2.3 Nível III

Compreensão / Aplicação – Uma questão teoricamente mais complexa, podendo ser uma situação problema ou uma atividade retirada de Avaliação Externa, ou de Vestibulares. O aluno precisa reconhecer a parte teórica e desenvolver conhecimentos para resolver a atividade. Ou seja, atinge até o patamar 5 da Taxonomia de Bloom.

Com essas normalizações conseguimos comparar alunos de turmas, e séries diferentes. Para facilitar na comparação, relacionamos essas três questões, construindo um gráfico de cores, que nos traz várias informações que nos direcionam a comentar e intervir no processo.

Apresentando a configuração completa da planilha de respostas, temos a tabulação por pergunta. Observemos a imagem na página a seguir:

3	Acertou as três questões - Possui a habilidade plenamente
2	Acertou duas questões (incluindo a Q1) - Possui a habilidade
1	Acertou ou errou apenas a Q1 - Análise pedagógica
0	Errou as três questões ou acertou apenas à Q2 ou Q3 - Não possui a habilidade

Figura 8: Definição das cores no gráfico

A ilustração abaixo é um exemplo do mapeamento da turma:

Matemática		Biologia			Química			Física			Índice de Competência	
C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11		C12
3	3	3	3	3	3	3	3	3	3	3	3	12
1	0	0	1	3	0	2	2	0	1	3	2	5
1	0	1	2	1	3	2	2	0	1	3	2	6
0	0	0	2	3	3	2	3	1	0	3	2	7
0	1	0	2	3	1	2	2	2	0	3	2	7
2	0	0	0	2	1	2	2	0	2	3	3	7
0	1	1	0	2	0	3	0	1	0	0	0	2
1	0	0	2	3	3	2	0	2	0	2	2	7
1	0	0	3	1	3	2	2	2	0	3	3	7
1	0	0	2	3	3	2	2	0	0	3	2	7
0	0	0	2	3	0	2	1	0	1	1	1	3
3	0	0	2	0	2	0	0	0	0	3	0	4
1	2	0	2	3	3	3	2	3	2	3	2	10
2	1	0	2	1	2	1	1	2	0	2	3	6
0	1	1	0	0	1	0	0	0	0	1	0	0
3	1	3	2	3	3	3	3	3	0	3	2	10
1	1	1	3	3	3	2	2	0	1	1	2	6
2	2	0	3	3	3	3	0	1	0	3	2	8
2	0	0	2	3	1	2	1	2	0	3	2	7
0	0	0	0	3	3	2	2	1	0	3	2	6
2	0	0	2	0	3	0	0	0	1	0	1	3
1	1	0	0	1	0	2	0	2	1	3	0	3
2	0	1	2	2	3	3	3	1	0	3	2	8
1	1	1	2	3	2	2	2	1	0	3	3	7
1	0	2	3	2	3	2	2	2	0	3	2	9
2	1	2	3	3	3	2	1	2	1	1	2	8
0	0	0	1	2	0	2	2	2	2	1	1	5
0	0	1	0	0	2	1	2	2	0	0	2	4
2	0	2	3	3	1	2	1	0	0	3	2	7
1	2	2	2	3	3	2	2	3	2	3	2	11
0	0	0	0	0	0	1	0	0	0	0	2	1
2	0	2	3	3	3	3	2	3	0	3	2	10
2	1	2	2	1	3	3	3	2	1	3	2	9

C1	Estudo dos pontos: Plano Ortogonal ; Coordenadas Cartesianas; Distância de dois pontos; Ponto médio
C2	Matrizes e G.A.:Equação da Reta; Condição de alinhamento de três pontos; Área de uma região circular.
C3	Inclinação da Reta: Coeficiente angular, Posição relativa entre duas retas, Equação da Reta
C4	Categorias Taxonomicas: Reino, Filo, Classe,Ordem, Família, Gênero e Espécies
C5	Nomenclatura: Escrita científica e suas aplicações
C6	Definição de Espécie: Animais híbridos
C7	Gases Atmosféricos
C8	Equilíbrio químico
C9	Rapidez na Transformação Química
C10	Consumo de Energia
C11	Circuitos elétricos
C12	Campos e Forças Eletromagnéticas.

Figura 9: Correção automatizada com o gráfico de mapeamento de habilidades

A nova configuração foi representada com tabulações por questões e o gráfico de cores.

Observando o gráfico podemos, num primeiro momento, tirar algumas conclusões. Por exemplo, se fizermos a leitura verticalmente, estaremos analisando a assimilação de conteúdos, conseqüentemente, o desempenho do professor:

- Vemos que a sala teve baixo desempenho nos conteúdos indicados por C₂, C₃ e C₁₀, respectivamente nas disciplinas de Matemática e Física. A devolutiva das questões envolvidas devem ser mais acentuadas.

Se analisarmos Horizontalmente, veremos o desempenho individual dos alunos:

- Os alunos que atingiram "0" e "1" competência, demonstraram grande possibilidade de ter colocado as alternativas de modo aleatório.
- Os alunos que não obtiveram nenhuma cor vermelha em seus resultados, Identificaram todos os conteúdos ministrados no referido Bimestre. Neste caso, precisamos verificar como é sua atenção em sala, pois, eles teriam grande chance de gabaritar provas desse nível.

Muitas outras hipóteses poderiam ser formadas se olharmos este gráfico com mais atenção.

4.3 ETAPA III

No final de 2014 percebemos que essa proposta poderia se tornar um projeto de pesquisa. Tínhamos ainda as seguintes questões: Esse gráfico de mapeamento (colorido) é confiável ou não? Ele realmente representa uma devolutiva de aprendizagem efetiva do aluno? Assim, decidimos analisar com um pouco mais de detalhes o gráfico de cores, utilizando ferramentas estatísticas. Os resultados serão apresentados no capítulo a seguir.

5

RESULTADOS OBTIDOS

Apresentamos aqui os resultados das avaliações com trinta e seis questões respondidas por setenta e oito alunos. Como estes dados não possuem uma distribuição normal, utilizaremos Estatística não paramétrica para analisá-los. Escolhemos uma única amostra para analisarmos as variações das características na trajetória desses elementos no período de 2º Bim de 2014 à 2º Bim de 2015. Nossa intenção é compreender a movimentação das variáveis nas respostas após várias aplicações desse tipo de avaliação, observando tanto como é a distribuição percentual dos resultados entre os níveis de dificuldade nos temas das disciplinas como também no desenvolvimento do aluno ao longo deste ano.

5.1 METODOLOGIAS ESTATÍSTICA

5.1.1 Teste de Friedman

O teste de Friedman é o teste não-paramétrico utilizado para comparar dados amostrais vinculados, ou seja, quando o mesmo indivíduo é avaliado mais de uma vez. O teste de Friedman não utiliza os dados numéricos diretamente, mas sim os postos ocupados por eles após a ordenação feita para cada grupo separadamente. Após a ordenação é testada a hipótese de igualdade da soma dos postos de cada grupo.

5.1.2 Teste de Wilcoxon para dados emparelhados

É um teste não paramétrico (assim como o de Friedman) aplicável quando as medidas são feitas apenas duas vezes sobre o mesmo indivíduo. É utilizado como pós teste de Friedman, para se detectar os pares de observações que diferem significativamente

entre si. Neste caso, precisamos utilizar a correção de Bonferroni para não alterarmos o nível de significância do teste.

5.2 COMPARAÇÕES DAS QUESTÕES - DESCRITIVOS

Analisamos a seguir o percentual de acertos de cada tipo de questão, por cada disciplina, ao longo de 3 bimestres. Organizamos uma única planilha, com todos os dados referentes às três avaliações bimestrais realizadas por setenta e oito indivíduos, dentre os anos de 2014 e 2015. Será observado, a ausência do provão bimestral do 4º Bimestre de 2014, decorrido pelo motivo de ter sido uma avaliação anual com quantidade de questões diferentes de nosso provão bimestral, não possibilitando sua comparação com as demais avaliações. Utilizamos como recurso tecnológico, o SPSS, um software aplicativo, com uso estatístico. para descrever as tabelas abaixo.

5.2.1 Questões da disciplina de Matemática

Tabela 1: Matemática

Bimestre	Questão	Média (DP)	IC 95%	Mediana (Mín-Máx)	p*
2º/ 2014	1	0,399 (0,291)	(0,332; 0,466)	0,330 (0,000 – 1,000)	0,009
	2	0,520 (0,317)	(0,447; 0,593)	0,670 (0,000 – 1,000)	
	3	0,400 (0,306)	(0,330; 0,470)	0,330 (0,000 - 1,000)	
3º/ 2014	1	0,503 (0,255)	(0,444; 0,562)	0,670 (0,000 – 1,000)	0,225
	2	0,5488 (0,187)	(0,445; 0,532)	0,330 (0,000 – 1,000)	
	3	0,449 (0,289)	(0,383; 0,516)	0,330 (0,000 - 1,000)	
1º/2015	1	0,480 (0,298)	(0,411; 0,548)	0,330 (0,000 – 1,000)	0,001
	2	0,341 (0,310)	(0,269; 0,412)	0,330 (0,000 – 1,000)	
	3	0,266 (0,275)	(0,203; 0,329)	0,330 (0,000 - 1,000)	

* *Teste de Friedman*

No 2º Bimestre de 2014, encontramos que a proporção de acertos de Q2 difere significativamente tanto de Q1 (com $p = 0,006$) como de Q3 (com $p = 0,045$). A proporção de acertos de Q1 e Q3 não diferem significativamente (teste de Wilcoxon com correção de

Bonferroni). Neste caso os alunos tiveram maior rendimento na questão de processo mecânico.

Já no 3º Bimestre de 2014, não encontramos diferenças significativas nas porcentagens de acerto das 3 questões ($p = 0,255$). Ou seja, a amostra estudada não diferenciou o grau de complexidade entre as três questões.

E no 1º Bimestre de 2015, encontramos que a proporção de acertos de Q1 difere significativamente tanto de Q2 (com $p = 0,003$) como de Q3 (com $p < 0,001$). A proporção de acertos de Q2 e Q3 não diferem significativamente (teste de Wilcoxon com correção de Bonferroni). Nesse caso, a movimentação teve o aspecto esperado $Q1 > Q2 > Q3$.

5.2.2 Questões da disciplina de Biologia

Tabela 2: Biologia

Bimestre	Questão	Média (DP)	IC 95%	Mediana (Mín-Máx)	p*
2º/2014	1	0,663 (0,288)	(0,597; 0,729)	0,670 (0,000 – 1,000)	0,001
	2	0,800 (0,275)	(0,737; 0,863)	1,000 (0,000 – 1,000)	
	3	0,676 (0,301)	(0,606; 0,745)	0,670 (0,000 – 1,000)	
3º/2014	1	0,752 (0,275)	(0,688; 0,815)	0,670 (0,000 – 1,000)	0,001
	2	0,556 (0,225)	(0,504; 0,607)	0,670 (0,330 – 1,000)	
	3	0,431 (0,324)	(0,357; 0,506)	0,330 (0,000 – 1,000)	
1º/2015	1	0,587 (0,305)	(0,517; 0,658)	0,670 (0,000 – 1,000)	0,001
	2	0,774 (0,270)	(0,712; 0,836)	1,000 (0,000 – 1,000)	
	3	0,458 (0,280)	(0,393; 0,522)	0,330 (0,000 – 1,000)	

* *Teste de Friedman*

No 2º Bimestre de 2014, encontramos que a proporção de acertos de Q2 difere significativamente tanto de Q1 (com $p = 0,009$) como de Q3 (com $p = 0,006$). A proporção de acertos de Q1 e Q3 não diferem significativamente (teste de Wilcoxon com correção de Bonferroni). Neste caso os alunos tiveram maior rendimento na questão de processo mecânico.

Já no 3º Bimestre de 2014, encontramos que a proporção de acertos de Q2 difere significativamente tanto de Q1 (com $p = 0,001$) como de Q3 (com $p = 0,045$). A proporção de acertos de Q1 e Q3 diferem significativamente (com $p = 0,001$) (teste de Wilcoxon com correção de Bonferroni). Nesse caso, a movimentação teve o aspecto esperado $Q1 > Q2 > Q3$.

E no 1º Bimestre de 2015, encontramos que a proporção de acertos de Q1 difere significativamente tanto de Q2 (com $p < 0,001$) como de Q3 (com $p = 0,009$). A proporção de acertos de Q2 e Q3 diferem significativamente (com $p < 0,001$) (teste de Wilcoxon com correção de Bonferroni). Neste caso, devido $Q2 > Q1$, os alunos tiveram maior rendimento na questão de processo mecânico.

5.2.3 Questões da disciplina de Química

Tabela 3: Química

Bimestre	Questão	Média (DP)	IC 95%	Mediana (Mín-Máx)	p*
2º/2014	1	0,570 (0,280)	(0,505; 0,634)	0,670 (0,000 – 1,000)	0,001
	2	0,502 (0,267)	(0,441; 0,563)	0,330 (0,000 – 1,000)	
	3	0,351 (0,301)	(0,281; 0,420)	0,330 (0,000 - 1,000)	
3º/2014	1	0,757 (0,259)	(0,697; 0,816)	0,670 (0,000 – 1,000)	0,001
	2	0,333 (0,280)	(0,268; 0,397)	0,330 (0,000 – 1,000)	
	3	0,551 (0,256)	(0,492; 0,610)	0,670 (0,000 - 1,000)	
1º/2015	1	0,493 (0,327)	(0,418; 0,568)	0,330 (0,000 – 1,000)	0,135
	2	0,480 (0,271)	(0,417; 0,542)	0,330 (0,000 – 1,000)	
	3	0,417 (0,259)	(0,357; 0,476)	0,670 (0,000 - 1,000)	

* *Teste de Friedman*

No 2º Bimestre de 2014, encontramos que a proporção de acertos de Q3 difere significativamente tanto de Q1 (com $p < 0,001$) como de Q2 (com $p = 0,003$). A proporção de acertos de Q1 e Q2 não diferem significativamente (teste de Wilcoxon com correção de Bonferroni). Nesse caso, a movimentação teve o aspecto esperado $Q1 > Q2 > Q3$. Porém percebemos que a amostra estudada não diferenciou o grau de complexidade dentre as

questões Q1 e Q2.

Já no 3º Bimestre de 2014, encontramos que a proporção de acertos de Q1 difere significativamente tanto de Q2 (com $p < 0,001$) como de Q3 (com $p = 0,009$). A proporção de acertos de Q2 e Q3 diferem significativamente (com $p < 0,001$) (teste de Wilcoxon com correção de Bonferroni). Neste caso, devido $Q2 < Q3 < Q1$, os alunos tiveram menor rendimento na questão de processo mecânico.

E no 1º Bimestre de 2015, não encontramos diferenças significativas nas proporções de acerto das três questões ($p = 0,135$) (teste de Wilcoxon com correção de Bonferroni). Nesse caso, a movimentação teve o aspecto esperado $Q1 > Q2 > Q3$. Porém percebemos que a amostra estudada não diferenciou o grau de complexidade dentre as três questões.

5.2.4 Questões da disciplina de Física

Tabela 4: Física

Bimestre	Questão	Média (DP)	IC 95%	Mediana (Mín-Máx)	p*
2º/2014	1	0,623 (0,336)	(0,546; 0,700)	0,670 (0,000 – 1,000)	0,001
	2	0,604 (0,319)	(0,531; 0,678)	0,670 (0,000 – 1,000)	
	3	0,413 (0,324)	(0,338; 0,488)	0,330 (0,000 - 1,000)	
3º/2014	1	0,770 (0,257)	(0,711; 0,829)	0,670 (0,000 – 1,000)	0,001
	2	0,391 (0,331)	(0,314; 0,467)	0,330 (0,000 – 1,000)	
	3	0,663 (0,283)	(0,598; 0,728)	0,670 (0,000 - 1,000)	
1º/2015	1	0,712 (0,271)	(0,650; 0,774)	0,670 (0,000 – 1,000)	0,001
	2	0,449 (0,345)	(0,369; 0,528)	0,330 (0,000 – 1,000)	
	3	0,569 (0,295)	(0,501; 0,637)	0,670 (0,000 - 1,000)	

* *Teste de Friedman*

No 2º Bimestre de 2014, encontramos que a proporção de acertos de Q3 difere significativamente tanto de Q1 (com $p < 0,001$) como de Q2 (com $p = 0,001$). A proporção de acertos de Q1 e Q2 não diferem significativamente (com $p < 0,979$) (teste de Wilcoxon com correção de Bonferroni). Nesse caso, a movimentação teve o aspecto esperado $Q1 > Q2 > Q3$. Porém percebemos que a amostra estudada não diferenciou o grau de

complexidade dentre as questões Q1 e Q2.

Já no 3º Bimestre de 2014, encontramos que a proporção de acertos de Q1 difere significativamente tanto de Q2 (com $p < 0,001$) como de Q3 (com $p = 0,021$). A proporção de acertos de Q2 e Q3 diferem significativamente (com $p < 0,001$)(teste de Wilcoxon com correção de Bonferroni). Neste caso, devido $Q2 < Q3 < Q1$, os alunos tiveram menor rendimento na questão de processo mecânico.

E no 1º Bimestre de 2015, encontramos que a proporção de acertos de Q1 difere significativamente tanto de Q2 (com $p < 0,001$) como de Q3 (com $p = 0,006$). A proporção de acertos de Q2 e Q3 diferem significativamente (com $p < 0,018$)(teste de Wilcoxon com correção de Bonferroni). Neste caso, devido $Q2 < Q3 < Q1$, os alunos tiveram menor rendimento na questão de processo mecânico.

5.2.5 Questões - Geral

Tabela 5: Geral

Bimestre	Questão	Média (DP)	IC 95%	Mediana (Mín-Máx)	p*
2º/2014	1	0,563 (0,193)	(0,519; 0,608)	0,580 (0,170 – 0,920)	0,001
	2	0,607 (0,199)	(0,561; 0,653)	0,580 (0,170 – 1,000)	
	3	0,459 (0,205)	(0,412; 0,506)	0,420 (0,080 - 1,000)	
3º/2014	1	0,695 (0,218)	(0,645; 0,746)	0,750 (0,000 – 1,000)	0,001
	2	0,444 (0,203)	(0,397; 0,490)	0,420 (0,170 – 0,920)	
	3	0,524 (0,243)	(0,468; 0,580)	0,500 (0,000 - 1,000)	
1º/2015	1	0,567 (0,198)	(0,522; 0,613)	0,580 (0,080 – 0,920)	0,001
	2	0,511 (0,201)	(0,465; 0,557)	0,500 (0,000 - 1,000)	
	3	0,427 (0,155)	(0,392; 0,463)	0,420 (0,080 - 0,830)	

* *Teste de Friedman*

No 2º Bimestre de 2014, encontramos que a proporção de acertos de Q2 difere significativamente tanto de Q1 (com $p = 0,048$) como de Q3 (com $p < 0,001$). A proporção de acertos de Q1 e Q3 não diferem significativamente (com $p < 0,001$)(teste de Wilcoxon com correção de Bonferroni). Neste caso os alunos tiveram maior rendimento na questão

de processo mecânico.

Já no 3º Bimestre de 2014, encontramos que a proporção de acertos de Q1 difere significativamente tanto de Q2 (com $p < 0,001$) como de Q3 (com $p = 0,001$). A proporção de acertos de Q2 e Q3 diferem significativamente (com $p < 0,001$)(teste de Wilcoxon com correção de Bonferroni).Neste caso ocorreu uma anormalidade, os alunos tiveram menor rendimento na questão de processo mecânico, percebemos também as questões Q2 e Q3 tiveram o mesmo grau de dificuldade.

E no 1º Bimestre de 2015, encontramos que a proporção de acertos de Q1 difere significativamente tanto de Q2 (com $p < 0,015$) como de Q3 (com $p = 0,001$). A proporção de acertos de Q2 e Q3 diferem significativamente (com $p < 0,001$)(teste de Wilcoxon com correção de Bonferroni).Neste caso, a movimentação teve o aspecto esperado $Q1 > Q2 > Q3$. Porém percebemos que a amostra estudada não diferenciou o grau de complexidade dentre as três questões.

6

DISCUSSÃO

Neste capítulo discutiremos os resultados obtidos.

No capítulo 5, apresentamos a análise estatística das avaliações realizadas na Escola. As avaliações neste modelo tiveram seu início no 2º Bimestre de 2014 para todas as três séries do Ensino Médio da escola, divididas em dez turmas. Essa quantidade de classes continuou até o fim de 2015. Tínhamos a possibilidade de analisar todas as avaliações já realizadas obtendo uma série de informações. Porém, para não delongar a pesquisa, optamos por escolher uma única amostra que participou de todas as avaliações. Escolhemos os alunos matriculados na 3ª série em 2015 que respectivamente estudaram na 2ª série de 2014, totalizando setenta e oito alunos.

Foram analisados, conjuntos de temas que abrangeram todo o conteúdo Bimestral, nas disciplinas de: Matemática, Química, Física e Biologia em três avaliações (2º Bim de 2014, 3º Bim 2014 e 1º Bim 2015) totalizando doze conjuntos de temas. Denominamos cada conjunto, como *competência*. Tendo em vista, o fato, que o avaliado possui a habilidade de identificar, desenvolver e compreender um tema, é constatado sua competência no assunto.

Fizemos também, mais três análises do geral em cada uma das três aplicações de Avaliação Bimestral. Para análise, separamos os três tipos de questões, dentro de cada disciplina, obtendo um conjunto de três questões em cada grau de complexidade.

O foco de nossa análise, não é comprovar a evolução desses alunos na realização deste tipo de avaliação, e sim, discutir os fatos ocorridos com essa amostra. Quanto maior a discussão, mais informações teremos sobre o aluno e sobre o processo, consequentemente construiremos mais proposta de intervenção que otimizariam futuras avaliações, beneficiando o aprendizado de nossos alunos. Sabemos que em novas teorias e metodologias propostas, não bastam apenas análises e discussões, deve-se haver progresso, caso contrário, não se justifica a sua inserção na rede educacional. Nossa intenção é constatar evolução no produto final, ou seja, nos resultados das avaliações externas.

De acordo com a nossa metodologia de predeterminar o grau de complexidade de cada habilidade para analisar uma competência, tendo as Taxonomias de Bloom como alicerces da avaliação, esperávamos certo comportamento na análise dos resultados. Lembrando, temos três habilidades que compõem o conjunto de uma competência, dadas em respectivamente três questões. Essas três questões chamaremos de: Q₁ a questão de identificação, Q₂ a questão de processo mecânico e Q₃ a questão de compreensão.

Teóricamente o grau de complexidade dentre as três questões seriam: $Q_1 < Q_2 < Q_3$. Logo, é esperado que na análise do comportamento de resposta obteríamos percentual de acertos: $Q_1 > Q_2 > Q_3$. Estatisticamente, não basta apenas obter o percentual de acertos de cada questão. Para termos grau de confiabilidade nos resultados se fez necessário utilizar de testes estatísticos para constatar se existia diferenças significativas nas porcentagens de acertos. Nesta amostra, de modo geral, percebemos que 53% dos conjuntos analisados obedeceram o comportamento esperado $Q_1 > Q_2 > Q_3$, 20% $Q_1 > Q_3 > Q_2$ os outros 27% tiveram o comportamento de $Q_2 > Q_1 > Q_3$. Observamos que a Q₃ realmente, sempre ficou com maior grau de complexidade em relação à Q₁. E em apenas três casos (20%) ficou menor que Q₂. Esse fato será discutido abaixo. Outro fato que será discutido é a relação entre Q₁ e Q₂ que quase não diferenciou o grau de complexidade na análise mais técnica.

Os caso em que Q₃ teve mais acertos que Q₂ foram nas avaliações de Química 3º Bim 2014, Física 3º Bim 2014 e Física 1º Bim 2015.

Química 3º Bimestre de 2014

Q2. *Considerando o valor da temperatura de ebulição do HCl (-85°C), quais devem ser mais fortes: as ligações covalentes existentes entre os átomos de H e de Cl na molécula de HCl ou as interações chamadas dipolo-dipolo e, sabendo que a mudança de estado físico implica o rompimento de interações intermoleculares e que, em condições ambientes o HCl é um gás e sua temperatura de ebulição é de - 85°C), pode-se concluir que é preciso:*

- a. fornecer pouca energia para que passe do estado líquido para gasoso*
- b. fornecer muita energia para romper as ligações*
- c. no estado líquido as forças de interações são bem fracas*

d. é necessário fornecer muita energia para quebrar as ligações no estado gasoso.

Q3. *Admite-se que a ligação entre os átomos de Hidrogênio e de Cloro Cl na molécula de cloreto de hidrogênio (HCl), envolve uma distribuição assimétrica de elétrons, favorecendo a formação de um dipolo. Levando isso em conta, que tipo de interação ocorre entre as moléculas de HCl.*

- a. as moléculas são apolares*
- b. as moléculas são polares*
- c. as moléculas são tetraédricas*
- d. as moléculas são angulares*

Lendo os exercícios do nível 2 e do nível 3 de química, vimos que as duas questões aparentemente possuem o mesmo grau de complexidade. Portanto, qualquer uma delas poderia ter maior porcentagem de acertos. Ambas se inserem em compreender o tema estudado. Este fato fez a diferença no resultado. Nesse caso, obtemos um exemplo de falha na elaboração do conjunto de habilidades, o que alterou as cores no gráfico, fato que resultará em estratégias de intervenção voltada á formação do professor.

Física no 3º Bimestre de 2014.

Q2. *Um violinista deseja aumentar a frequência do som emitido por uma das cordas do seu instrumento. Isto poderá ser conseguido:*

- a. aumentando-se o comprimento vibratório e tracionando-se mais intensamente a corda;*
- b. diminuindo-se o comprimento vibratório e tracionando-se menos intensamente a corda;*
- c. diminuindo-se o comprimento vibratório e tracionando-se mais intensamente a corda;*
- d. aumentando-se o comprimento vibratório e tracionando-se menos intensamente a corda;*

Q3. *Um estudante, fazendo um experimento no laboratório de sua escola, acoplou um gerador de audiodfrequência a um alto-falante. Aumentando, então, a frequência do aparelho de 200Hz para 2800Hz, ele notou que o som produzido pelo sistema ficou:*

- a. menos intenso ou mais fraco;*
- b. mais alto ou agudo;*

- c. *mais baixo ou grave;*
- d. *mais rico em harmônicos;*

Vemos que em física 3º Bimestre 2014, aconteceu o mesmo caso ocorrido em química. Nota-se, aparentemente que ambos são exercícios de compreensão, assim qualquer um deles poderia ter maior porcentagem de acertos.

Física no 1º Bimestre de 2015. Conjunto 1 (C1)

Q2. *Que símbolos representam as unidades de corrente, tensão, potência e frequência de cada aparelho?*

- a. *Ampere (A); Volt (V); Watt (W); hertz (Hz).*
- b. *Volt (V); Watt (W); hertz (Hz); Ampere (A).*
- c. *Watt (W); hertz (Hz); Ampere (A); Volt (V).*
- d. *hertz (Hz); Ampere (A); Volt (V); Watt (W).*

Q3. *Um resistor de resistência equivalente a $10\ \Omega$ é percorrido por uma intensidade de corrente elétrica igual a 6 A. Qual a ddp (U) entre os extremos do resistor?*

- a. *40V*
- b. *50V*
- c. *60V*
- d. *70V*

No primeiro conjunto de questões, em Física 1º Bimestre 2015, temos que a Q2 e a Q3 são de processo mecânico. Porém, Q2 trabalha com a identificação de unidades de medidas, e Q3 exige que o aluno apenas calcule o produto entre 10 e 6, fato que torna o exercício bem menos complexo.

Física no 1º Bimestre de 2015. Conjunto 2 (C2)

Q2. *Qual grandeza pode ajudar você na avaliação do consumo de energia elétrica?*

- a. *Corrente elétrica*

- b. *Voltagem*
- c. *Frequência*
- d. *Potência*

Q3. *Um chuveiro de 5400 W – 220 V ligado por 30 minutos, a avaliação do consumo de energia elétrica será de:*

- a. *1350 wh*
- b. *1500 wh*
- c. *2600 wh*
- d. *27000 wh*

No segundo conjunto de questões, da avaliação de Física 1º Bimestre 2015, temos que Q2 é um exercício de compreensão e Q3 é de processo mecânico, exigindo que o aluno apenas calcule o produto entre 5400 e 0,5. Esta inversão acentuou maior acerto de Q3. Estas competências cruzadas foram cruciais para sacramentar a anormalidade ocorrida, devido erro na ordem das questões.

Vamos discutir o caso em que o comportamento ficou $Q_2 > Q_1 > Q_3$, em Biologia, 1º Bimestre de 2015.

O conjunto de questões elaborados em Biologia obedeceram o critério de Q_1 - *Identificação*, Q_2 - *Desenvolvimento* e Q_3 - *Aplicação*. Em geral os alunos obtiveram ótimo rendimento nesta disciplina com média de 64% por exercício, porém, como na análise percebemos que o resultado não obedeceu o padrão esperado, discuremos as possíveis causas.

Os exercícios de desenvolvimento tiveram rendimento melhor que os exercícios de identificação. Sobre esse fato, levantamos algumas hipóteses sobre a falha e elaboramos algumas propostas de intervenção. Uma das hipóteses levantadas seria que os alunos apenas conseguiram assimilar o processo mecânico sem entender toda a completude do assunto estudado.

Este problema poderia ser resultado do método de transmissão da aula. Devido aos resultados, percebemos que seria necessário uma ênfase maior à identificação dos conteúdos aplicados, ou seja, se preocupar um pouco mais com quais os motivos que

levaram ao aparecimento destes estudos em Biologia.

Abaixo, as questões aplicadas.

Biologia no 1º Bimestre de 2015. Conjunto 1 (C1)

Q1. (PUC-SP) O ramo da Biologia que desenvolve a classificação dos seres vivos e estabelece critérios para classificar todos os animais e plantas sobre a Terra em grupos de acordo com as características fisiológicas, evolutivas, anatômicas e ecológicas de cada ser vivo. Identifique esse ramo da Biologia que trata dessa descrição:

- a. Citologia
- b. Bioética
- c. Ecologia
- d. Taxonomia

Q2. (SARESP) A hierarquia correta atualmente desenvolvida e aceita sobre a classificação dos seres vivos:

- a. Gênero-Reino - Espécie - Família - Ordem - Filo – Classe
- b. Classe - Reino - Gênero - Ordem - Espécie - Filo – Família
- c. Reino - Filo - Classe - Ordem - Família - Gênero - Espécie
- d. Reino - Classe –Filo- Ordem - Família - Espécie- Gênero

Q3. (PUC-RJ) O lobo-guará e a onça são dois exemplares da nossa fauna ameaçados de extinção. O diagrama a seguir mostra as principais categorias taxonômicas a que pertencem estes animais: Lobo-guará: Cordado – mamífero – carnívoro – canídeo – Chrysocyon - Chrysocyon brachyurus Onça parda: Cordado – mamífero – carnívoro – felídeo – Puma – Puma concolor A análise do diagrama permite dizer que os dois animais estão próximos na mesma categoria até:

- a. Classe
- b. Gênero
- c. Filo
- d. Ordem

Biologia no 1º Bimestre de 2015. Conjunto 2 (C2)

Q1. (FUVEST) A nomenclatura binomial cria um padrão na escrita de nomes científicos, evitando assim a confusão causada por nomes populares. A respeito das regras de nomenclatura binomial, identifique a alternativa INCORRETA:

- a. Todos os nomes científicos devem vir destacados no texto em itálico ou sublinhados.
- b. O gênero deve ser escrito com inicial maiúscula.
- c. A espécie deve ser escrita com letra minúscula.
- d. A nomenclatura binomial não apresenta nome científico em sua escrita

Q2. (UNISA) Com base nas regras de nomenclatura, indique a alternativa que apresenta seu desenvolvimento correto:

- a. homo sapiens
- b. Canis Familiaris
- c. PANTHERA ONÇA
- d. Puma concolor

Q3. (Unirio - RJ) Carlos Lineu, em 1735, publicou um trabalho, no qual apresentava um plano para classificação de seres vivos. Nele estavam propostos o emprego de palavras latinas e o uso de categorias de classificação hierarquizadas. Deve-se também a Lineu a regra de nomenclatura binominal para identificar cada organismo. Nesta regra, entre outras recomendações, fica estabelecido que devemos escrever:

- a. em primeiro lugar o gênero, depois a família.
- b. em primeiro lugar o gênero, depois a espécie.
- c. em primeiro lugar a espécie, depois o gênero.
- d. em primeiro lugar a espécie, depois o filo

Biologia no 1º Bimestre de 2015. Conjunto 3 (C3)

Q1. (Modelo SARESP) A espécie é a unidade taxonômica fundamental que agrupam seres vivos que possuem a mesma característica cromossômica em um conjunto de organismos semelhantes entre si, capazes de se cruzar e gerar descendentes férteis. Com base na informação identifique onde a espécie é aplicada corretamente:

I- O cavalo e a égua eles podem acasalar-se e dar origem a um descendente fértil.

II- O jumento (macho) e jumento (fêmea) eles podem acasalar-se e dar origem a um descendente fértil.

III- O leão e a tigresa eles podem acasalar-se e dar origem a um descendente infértil.

- a. Apenas a I
- b. Apenas a II e a III
- c. Apenas a I e a II
- d. Todas estão corretas

Q2. (MACK) São considerados animais híbridos os descendentes cujos pais pertencem a espécies diferentes. Encontre nas alternativas o desenvolvimento de hibridação entre os animais:

- a. O jumento e a égua
- b. O boi e a vaca
- c. O bode e a cabra
- d. O carneiro e a ovelha

Q3. (Unifesp) Em uma área de transição entre a mata atlântica e o cerrado, são encontrados o pau-d'arco (*Tabebuia serratifolia*), a caixeta (*Tabebuia cassinoides*) e alguns ipês (*Tabebuia aurea*, *Tabebuia alba*, *Cybistax antisyphilitica*). O cipó-de-são-joão (*Pyrostegia venusta*) é também freqüente naquela região. Considerando os critérios da classificação biológica, no texto são citados

- a. 3 gêneros e 3 espécies
- b. 3 gêneros e 4 espécies
- c. 3 gêneros e 6 espécies
- d. 4 gêneros e 4 espécies

7 | CONCLUSÃO

O gráfico de cores realmente nos trás informações pertinentes que se aproximam muito da realidade dos conteúdos assimilados pelo aluno, em união com nosso objetivo de atingir metas externas. Em relação à construção da avaliação, sua correção computadorizada e as regras de formação das cores, nossos estudos mostram que toda a estrutura pedagógica tem bases em estudos psicométricos; os itens devem ser construídos com todas as normas especificadas para sua formatação abrangendo os conteúdos do bimestre. Vimos também o comportamento dos alunos nos resultados das avaliações, onde percebemos que no decorrer das aplicações, estatisticamente, o resultado converge para o padrão esperado. Desse modo, temos mutuamente as duas frentes pedagógica e estatística se validando.

Vou comentar os fatos estudados iniciando pelos alunos e finalizando com os objetivos da gestão escolar.

7.1 ALUNOS

Na devolutiva de cada avaliação, a maioria dos alunos, demonstra muito interesse e empolgação para escutar os comentários pertinentes ao seu resultado discriminados no gráfico, principalmente o aluno de melhor rendimento. Ele consegue identificar suas dificuldades e / ou os eventuais desvios ocorridos que o impossibilitaram de obter um rendimento melhor.

7.2 PROFESSORES

Após as provas, os professores têm o mapeamento da sala, sabendo até quem colocou as respostas de modo aleatório. O mapeamento da sala é importante também para o professor saber o grau de complexidade que os alunos atingiram e qual conteúdo deve ser repassado. Os professores, também começam a perceber as diferenças dos três níveis de complexidade das questões ao longo do processo, evitando falhas na elaboração das mesmas como foi verificado em algumas avaliações.

7.3 COORDENADORES

O coordenador pedagógico tem o mapeamento de todo o conteúdo transmitido; observando o gráfico verticalmente é possível verificar o desempenho do professor, e monitorar o programa bimestral.

7.4 GESTÃO

A equipe de gestão se preocupa com o todo, avaliações internas e avaliações externas. Essa estrutura de avaliação interna influenciou diretamente nos resultados de avaliações externas. Temos resultados de avaliação externa (SARESP) dos últimos quatro anos (2011 à 2014) nas disciplinas de Português e Matemática:

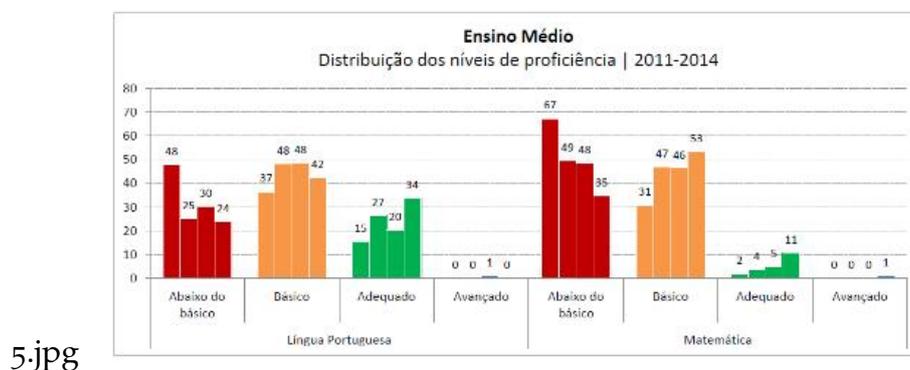


Figura 10:

O interessante é que a estrutura de avaliação foi aplicada apenas em Matemática. Como a mesma população realizou as duas avaliações, pudemos observar o comportamento de rendimento dos alunos nesses anos.

Observando o gráfico das duas disciplinas notamos que o gráfico de Matemática, mesmo sendo inferior dos resultados de Língua Portuguesa, demonstra mais consistência em seu crescimento de maturidade, pois, de modo contínuo decresceu no Abaixo do Básico, e cresceu nos níveis Básico e Adequado.

Nesse período (2011 à 2014) tivemos também avaliações bianuais do SARESP em Ciências da Natureza (2012 e 2014) e Ciências Humanas (2011 e 2013). As disciplinas de Ciências da Natureza são compreendidas por Biologia, Química e Física, foram as únicas disciplinas que utilizaram o processo de Estrutura de Avaliações. As disciplinas de Ciências Humanas, são compreendidas por: História, Filosofia, Sociologia e Geografia. Devido suas datas de aplicação, não comparamos seus resultados.

A melhoria no desempenho em Matemática e Ciências da Natureza (que participaram das Estruturas de Avaliação) com Língua Portuguesa, juntamente com as notas do Estado de São Paulo em Matemática e Língua Portuguesa (que não utilizaram esse recurso) nos anos de 2012 e 2014 são:

	2012	2014	Crescimento
Ciên. Natureza	1,85	2,99	62%
Matemática	1,81	2,61	44%
Língua Portuguesa	3,37	3,66	9%
Mat. Estado	1,64	1,67	2%
Port. Estado	3,1	2,94	-5%

7.jpg

Figura 11: Tabela de ganhos

Com esses dados notamos que as estruturas de Avaliação influenciaram de modo benéfico nas avaliações externas. Ou seja, concluímos que todo o esforço desenvolvido para modificarmos os paradigmas de avaliação utilizando a Taxonomia de Bloom e os gráficos de cores foram decisivos para a melhoria do desempenho da escola estudada nas avaliações externas.

BIBLIOGRAFIA

- [Ho] HOFFMANN, Jussara Maria Lerch; *Avaliação: mito e desafio: uma perspectiva construtivista*, Porto Alegre: Editora Mediação, 1991.
- [Lu] Luckesi, Cipriano Carlos *Avaliação da Aprendizagem escolar: estudos e proposições*, São Paulo: Cortez, 2005.
- [Im] IMBERNÓN, Francisco *Formação Docente e profissional: formar-se para a mudança e a incerteza. - 5. ed.*, São Paulo: Cortez, 2005.
- [MS] MEIER, Marcos; GARCIA, Sandra *Mediação da aprendizagem: contribuições de Fierstein e de Vygotsky*, , Curitiba: Edição do Autor, 2011.
- [BC] BONJORNO, Regina Azenha; CLINTON, Marcio *Física fundamental - Novo: volume único, 2º grau* São Paulo: FTD, 2009.
- [UL] ULTIMURA, TeruKo Yamamoto; LINGUANOTO, Maria; *Química: livro único*, São Paulo: FTD, 1998.
- [1] Ciências da natureza, matemática e suas tecnologias / Secretaria da Educação Básica Brasília: Ministério da Educação, Secretaria da Educação Básica, 2008.
- [Pe] PERRENOUD, Phelippe. *Avaliação da excelência à regulação das aprendizagens entre duas lógicas*, Porto Alegre: Artes Médicas, 1999
- [Sa] SANT'ANNA, Ilza Martins. *Por que Avaliar? : como avaliar? : critérios e instrumentos*, Petrópolis, RJ : Vozes, 1995
- [Ra] RABELO, Mauro. *Avaliação educacional: fundamentos, metodologias e aplicações no contexto brasileiro*, Rio de Janeiro, RJ : Sociedade Brasileira de Matemática, 2013
- [2] Relatório pedagógico: Saesp: Matemática / Secretaria da Educação do Estado de São Paulo São paulo: Secretaria da Educação do Estado de São paulo, 2009.
- [3] Relatório pedagógico: Saesp: Matemática / Secretaria da Educação do Estado de São Paulo São paulo: Secretaria da Educação do Estado de São paulo, 2010.

- [4] Relatório pedagógico:Saresp:Matemática / Secretaria da Educação do Estado de São Paulo São paulo: Secretaria da Educação do Estado de São paulo,2011.
- [5] Relatório pedagógico:Saresp:Matemática / Secretaria da Educação do Estado de São Paulo São paulo: Secretaria da Educação do Estado de São paulo,2012.
- [6] Relatório pedagógico:Saresp:Matemática / Secretaria da Educação do Estado de São Paulo São paulo: Secretaria da Educação do Estado de São paulo,2013.
- [7] Fonte: Portal INEP. Disponível em: <<http://portal.inep.gov.br/web/saeb/aneb-e-anresc>> acesso em: 12/06/2015
- [8] Fonte: Portal INEP. Disponível em: <<http://portal.inep.gov.br/pisa-programa-internacional-de-avaliacao-de-alunos>> acesso em: 12/06/2015
- [9] Fonte: Portal INEP. Disponível em: <http://download.inep.gov.br/educacao_basica/enem/n> acesso em: 12/06/2015
- [10] Fonte: Portal INEP. Disponível em: <<http://mapaitensenem.inep.gov.br/mapaNota/>> acesso em: 12/06/2015
- [11] Fonte: Portal INEP. Disponível em: <<http://portal.inep.gov.br/web/enem/sobre-o-enem>> acesso em: 12/06/2015
- [12] Fonte: Portal INEP. Disponível em: <> acesso em: 12/06/2015
- [13] Fonte: Portal INEP. Disponível em: <> acesso em: 12/06/2015
- [14] Fonte: Portal INEP. Disponível em: <> acesso em: 12/06/2015
- [B1] Bloom, Benjamin S. *Taxonomia de Objetivos Educacionais (1956). Avaliação da Aprendizagem escolar: estudos e proposições*, Boston, MA. Copyright (c) 1984 por Pearson Education.
- [B1] Bloom, Benjamin S. *Todas as nossas crianças Aprendizagem*, New York: McGraw-Hill.
- [B1] Bloom, Benjamin S. *Desenvolvimento de Talentos nos jovens*, New York: Ballantine Books.
- [15] Fonte: <http://www-history.mcs.st-and.ac.uk/Biographies/Bonferroni.html>. Disponível em: <> acesso em: 21/08/2015

[16] Fonte:https://en.wikipedia.org/wiki/Carlo_Emilio_Bonferroni.Disponível em: <> acesso em: 21/08/2015

[17] Fonte:<https://www.dicionarioinformal.com.br/avaliação/>.Disponível em: <> acesso em 17/07/2015

MEIER, Marcos; GARCIA, Sandra *Mediação da aprendizagem: contribuições de Fuerstein e de Vygotsky*, , Curitiba: Edição do Autor,2011.