



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Matemática

Uma proposta de exame de proficiência em Cálculo
Diferencial e Integral

por

Fabio Henrique Gomes

Brasília, 2016

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

GG633p Gomes, Fabio Henrique
Uma proposta de exame de proficiência em Cálculo
Diferencial e Integral / Fabio Henrique Gomes;
orientador Mauro Luiz Rabelo. -- Brasília, 2016.
65 p.

Dissertação (Mestrado - Mestrado Profissional em
Matemática) -- Universidade de Brasília, 2016.

1. Avaliação. 2. Cálculo Diferencial e Integral.
3. Proficiência. 4. Aprendizagem em cálculo. I.
Rabelo, Mauro Luiz, orient. II. Título.

Fabio Henrique Gomes

**Uma proposta de exame de proficiência em Cálculo
Diferencial e Integral**

Dissertação apresentada ao Departamento de Matemática da Universidade de Brasília, como parte dos requisitos para a obtenção do grau de

Mestre

Orientador: Prof. Dr. Mauro Luiz Rabelo

Brasília
2016

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Matemática

Uma proposta de exame de proficiência em Cálculo Diferencial e Integral.
por

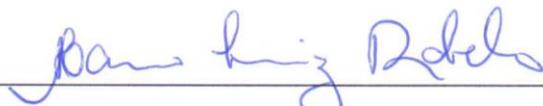
Fábio Henrique Gomes *

Dissertação apresentada ao Departamento de Matemática da Universidade de Brasília, como parte dos requisitos do "Programa" de Mestrado Profissional em Matemática em Rede Nacional – PROFMAT, para obtenção do grau de:

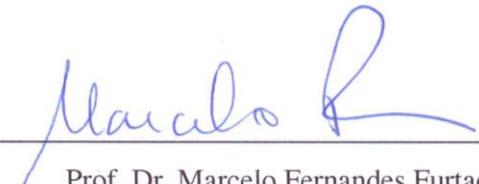
MESTRE

Brasília, 24 de junho de 2016.

Comissão Examinadora:



Prof. Dr Mauro Luiz Rabelo – MAT/UnB (Orientador)



Prof. Dr. Marcelo Fernandes Furtado – MAT/UnB (Membro)



Prof. Dr. Celius Antonio – MAT/UnB (Membro)

* O autor foi bolsista CAPES durante a elaboração desta dissertação

“Uma das grandes características, uma das grandes virtudes do professor é ele se expor perante seus alunos” Ubiratan D’Ambrosio

Dedicatória

A todos os que me educam no mundo de hoje e de ontem. Pois no contato com o mundo que me formo e fortaleço, assim a tais pessoas devo a vontade em continuar apreender e ressignificar o mundo. A essas que estiveram por vezes mais próxima como meus pais e minha irmã e a proximidade de tantas outras que às vezes nem imaginam o quanto trazem o fortalecimento para continuar caminhando. Vale ressaltar que a distância material, não nos deixa longe e significa uma proximidade através de boas energias.

Agradecimentos

À compreensão de todos que me aturaram no período de realização desse projeto, como os próprios professores, funcionários e alunos da jornada dupla de estudo e trabalho, assim como a Secretaria de Educação do Distrito Federal pela que concedeu dispensa parcial de alguns horários das coordenações.

À colaboração ao meu Orientador Prof. Dr. Mauro Luiz Rabelo, que me entusiasmou a realizar esse projeto, assim como a modificação que já se apresenta em minha práxis, assim como Prof. Dr. Rui Seimetz que me acolheu desde meu ingresso a cidade de Brasília orientando e acolhendo junto ao Departamento de Matemática. Ao MAT que abriu a oportunidade de participar das reuniões e em especial aos Professores Dr. Celius Antônio Magalhães e Dr. Marcelo Fernandes Furtado. Ao Rogério Moraes de Carvalho, pelas valiosas contribuições acerca de Testes Adaptativos.

Não esquecendo dos companheiros de turma do PROFMAT que sempre atenciosos e unidos deram uma cara colaborativa ao aprendizado no curso, em especial ao Josué Sergio de Souza que colaborou com a construção conjunta de uns dos capítulos, que foi parte importante do aprendizado científico, ao Thiago Yamashita Paiva e André Marcelino Marques que mais próximos deram apoio de encorajamento para chegar na finalização do curso e Augusto Hung pela sempre disponibilidade em ajudar na parte de formatação do trabalho.

A minha esposa Katia Motta Galvão Gomes, pela transcrição da entrevista pela paciência nas explicações sobre o projeto e todo o suporte do dia a dia. E aos amigos Carlos Augusto Casarin, Wilson Santos Marrola e Mariana Dutra Fogaça pelos bons conselhos na caminhada.

Em especial o programa do PROFMAT, que da união entre a CAPES e SBM puderam contribuir com a formação acadêmica, dado a natureza da democratização do acesso por diversos fatores que impulsiona a democratização da ciência.

Resumo

A dificuldade no aprendizado na disciplina de Cálculo Diferencial e Integral (CDI) e o elevado índice de reprovações e desistências de estudantes de diversos cursos de nível superior têm despertado a atenção nas últimas décadas, levando os educadores a repensarem os modelos de ensino, propondo novas práticas pedagógicas que façam uso, inclusive, das tecnologias hoje disponíveis. Nessa perspectiva, o Departamento de Matemática (MAT) da Universidade de Brasília implantou, nos últimos anos, algumas inovações no que diz respeito ao processo de avaliação dos estudantes de CDI e ao modelo pedagógico de aulas, com a intenção de melhor compreender e sanar as dificuldades dos estudantes. As inovações baseiam-se na concepção de avaliação de competências, com suporte da psicometria e da estatística. Neste trabalho, apresentamos, inicialmente, um breve histórico do trabalho recentemente desenvolvido pelo MAT com a implantação do cálculo magistral e a criação de testes de múltipla escolha como parte do processo de avaliação da aprendizagem em CDI, recorrendo-se à construção de uma matriz de referência para avaliação do desenvolvimento de habilidades e competências, bem como à Teoria da Resposta ao Item (TRI) e à Teoria Clássica dos Testes (TCT) para calibração dos parâmetros dos itens e obtenção das proficiências dos estudantes. O trabalho realizado pelos docentes do MAT constitui também uma etapa inicial da construção de um banco de itens para avaliação em CDI e criação de um futuro exame de proficiência em Cálculo. Fazemos aqui uma análise dos dois primeiros testes aplicados no primeiro semestre de 2015 e de seus resultados, assim como a preparação de uma escala de proficiência para servir de *feedback* aos participantes. Concluímos com a apresentação de um esquema simplificado de uma proposta de Avaliação Computacional Adaptada (CAT) para CDI.

Palavras-chave: Avaliação, Cálculo Diferencial e Integral, Proficiência, Aprendizagem em Cálculo.

Abstract

Educators has been increase the attention give to the learning difficulty in Differential and Integral Calculus (DIC) and the high disapproval rate and dropout of students from various higher education courses. They having been rethink their teaching models. As result, it is been proposed new pedagogical practices that make use of the technologies available today. In this perspective, the Department of Mathematics (MAT) at the University of Brasilia has implemented innovations in the evaluation process of students and the teaching model lessons both in order to better understand the student?s difficulties. This innovations was made based on the concept of evaluation skills, using psychometrics and statistics. In this dissertation we first did a brief history of the work developed by MAT with the implementation of masterful calculus class and the creation of multiple-choice tests as part of the evaluation process of learning in DIC, based on the construction of a reference matrix to assess the development of skills and competences, and based also on the item response theory (IRT) and the classical theory of tests (CTT) for calibration of the parameters of the items and obtain the proficiencies of the students. The work done by MAT are both an initial step to build a bank of items to evaluation in DIC and to create a proficiency exam in Calculus. We analyze the first two tests that were applied in the first semester of 2015. The tests results as well as the preparation of a proficiency scale can be used as a feedback to the participants. In conclusion we present a simple diagram of a proposed evaluation Computational Adapted Testing (CAT) to DIC.

Keywords: Evaluation, Differential and Integral Calculus, Proficiency, Learning in Calculus.

Sumário

Introdução	1
1 Breve comentário sobre avaliação	6
1.1 Avaliação	7
1.2 Matrizes de referência	14
2 Psicometria: Conceitos Básicos	17
2.1 Breve Histórico	17
2.2 Teoria Clássica dos Testes (TCT) e Teoria de Resposta ao Item (TRI) .	18
3 Histórico do projeto da aula magistral e do banco de itens para exame de proficiência em Cálculo	24
3.1 A implementação	27
3.2 Construindo os itens	31
3.3 O primeiro teste	36
3.4 As primeiras avaliações dos projetos	40
4 Análise dos Itens aplicados no primeiro semestre de 2015	45
5 Proposta de criação do CAT para implantação do exame de proficiência em Cálculo	55
6 Considerações finais	61

Lista de Figuras

4.1	Item 6 - P2 - 1/2015.	48
4.2	Análise Gráfica do Item 6 - P2 1/2015	49
4.3	Curva Característica do Item 6 - P2 - 1/2015	50
4.4	Item 16 - P2 - 1/2015.	50
4.5	Análise Gráfica do Item 16 - P2 1/2015	52
4.6	Curva Característica do Item 16 - P2 - 1/2015	52
4.7	Item 26 - P2 - 1/2015.	53
4.8	Análise Gráfica do Item 26 - P2 1/2015	53
4.9	Curva Característica do Item 26 - P2 - 1/2015	54
5.1	Curva Característica do Item 01 - P2 - 1/2015	56

Lista de Tabelas

1.1	Relações entre as habilidades (H_i) e as características do perfil (P_i) . . .	16
3.1	Distribuição dos itens segundo o parâmetro de dificuldade b	37
3.2	Parâmetros dos itens da primeira avaliação segundo a TRI e a TCT . . .	38
3.3	Percentuais de aprovação nas turmas dos cursos diurnos: semestres 1 e 2/2014 e 2/2015	39
4.1	Parâmetros da segunda avaliação do primeiro semestre de 2015	46
4.2	Distribuição dos itens segundo o parâmetro de dificuldade b	47
4.3	Faixa de distribuição para o Parâmetro de discriminação	47
4.4	Distribuição dos itens segundo o parâmetro de dificuldade b	47

Lista de Abreviaturas e Siglas

CDI	Cálculo Diferencial e Integral
MAT	Departamento de Matemática
TRI	Teoria da Reposta ao Item
TCT	Teoria Cássica dos Testes
CAT	Teste Adaptativo Computadorizado
IES	Instituições de Superior
UnB	Universidade de Brasília
USP	Universidade de São Paulo
UFF	Universidade Federal Fluminense
LDB	Lei de diretrizes e bases
SAEB	Sistema de Avaliação da Educação Básica
DCN	Diretrizes Curriculares Nacionais
CNE/CES	Conselho Nacional de Educação/Câmara de Educação Superior
ENEM	Exame Nacional do Ensino Médio
CESPE	Centro de Seleção e de Promoção de Eventos
REUNI	Reestruturação e Expansão das Universidades Federais

Introdução

A Educação Superior no Brasil vem passando por profundas transformações nas últimas décadas, decorrentes do aumento aproximado de 138% no número de alunos em Instituições de Ensino Superior (IES), ocorrido de 1996 a 2005, dos quais aproximadamente 64% nas IES Públicas, da criação de novas IES públicas e inserção de estudantes pertencentes a grupos sociais historicamente excluídos do ensino superior (CUNHA, 2007). Isso vem trazendo um novo desafio para as Universidades, conforme aponta Rabelo, Magalhães e Furtado (2015):

Naturalmente, esse quadro vem trazendo para o cenário acadêmico um novo perfil de aluno, levando as instituições de ensino superior (IES) a repensar estratégias de acolhimento e avaliação dos estudantes (RABELO; MAGALHÃES; FURTADO, 2015).

O estudante, após vencido o obstáculo de acesso à Universidade, nessa transição do Ensino Médio para o Superior, depara-se com a necessidade de adaptação a um meio acadêmico muito diferente do que encontrou em sua juventude, que exige dele mais autonomia, desenvolvimento do pensamento crítico, da capacidade de trabalho em equipes e redes, entre outras. Um dos desafios é propiciar a inclusão desejada, garantir um processo de crescimento com qualidade, equidade social e a permanência bem-sucedida e com dignidade (ALMEIDA; MARINHO-ARAÚJO; AMARAL; DIAS, 2012). Nesse processo existe o desafio de identificar e indicar caminhos para que os estudantes possam vencer suas fraquezas individuais para dar prosseguimento nos estudos e concluir com êxito a tão esperada formação profissional. Assim, as instituições precisam desenvolver estratégias de acolhimento de seus discentes para promover a adaptação necessária, em especial durante o primeiro ano do curso de graduação.

Na Universidade de Brasília (UnB) durante o primeiro semestre de 2014, no Departamento de Matemática (MAT), alguns professores, com o intuito de compreender melhor o aprendizado em matemática dos estudantes do primeiro ano de graduação, começaram a debater uma proposta de criação de um exame de proficiência em Cálculo

Diferencial e Integral (CDI). Para tal, criaram uma matriz de referência tridimensional, correlacionando as características do perfil profissional de formação desejado com as habilidades imprescindíveis para o alcance do perfil e os conhecimentos de matemática necessários para seu desenvolvimento (RABELO; MAGALHÃES; FURTADO, 2015).

A disciplina de Cálculo Diferencial e Integral (CDI) é geralmente ministrada no primeiro ano dos cursos de formação superior de Matemática, Física, Química, Engenharias, Administração, Economia, entre outros, ocorrendo variações de acordo com os projetos pedagógicos elaboradas pelas Instituições de Ensino Superior (IES) que ofertam esses cursos.

O tema da avaliação em CDI é de relevante interesse na educação superior devido às elevadas taxas de reprovação e de repetência, que chegam a levar a atrasos na conclusão da graduação, gerando desistência ou desligamento de um número perceptivo de estudantes, atrapalhando planos de formatura e profissionalização, e frustrando sonhos de obtenção de um diploma (RABELO; MAGALHÃES; FURTADO, 2015).

Estudos apontam que existem obstáculos à aprendizagem em Cálculo, decorrente principalmente da forma com que o estudante relaciona-se com o aprendizado em matemática antes de chegar à universidade. O que se observa com frequência é a ênfase na memorização de regras e métodos de resolução de problemas, com foco na repetição de modelos, contrapondo-se ao elevado nível de raciocínio exigido para o aprendizado significativo em CDI, que contempla um grande número de ferramentas e estratégias matemáticas, apropriação da lógica e maior abstração (MELLO; MELLO, 2007).

Um dos problemas que surgem na transição entre o ensino médio e a universidade é a grande quantidade de estudantes que escolhem a maneira de se promover nas avaliações escolares simplesmente pela memorização de procedimentos rotineiros de resolver determinados tipos de problemas de matemática. O uso dessa estratégia para aprender Cálculo é uma tarefa difícil, já que os procedimentos e problemas de Cálculo são numerosos e complexos. A aprovação em Cálculo por alunos que utilizaram essa lógica, apesar de mostrar que possuem grande capacidade, não é o suficiente para o sucesso no prosseguimento do curso que tem essa disciplina como obrigatória.

Os alunos estarão mais bem preparados quando demonstrarem a compreensão dos princípios e conceitos, aprenderem a aplicá-los em diferenciados contextos, desenvolvendo os procedimentos da matemática de forma coesa e coerente..

Nos cursos de Cálculo, exige-se a fluência na linguagem matemática específica dessa área. Espera-se que os estudantes trabalhem com exponenciais, logaritmos, linguagem de função, relações trigonométricas, teorema binomial, limites, derivadas, máximos e mínimos, integrais, entre outros. Deseja-se que eles transitem com naturalidade entre conteúdos relacionados, de forma quase espontânea, como no caso de triângulos para círculos trigonométricos e suas funções, ajam de forma instin-

tiva para serem mais produtivos em determinados contextos, como o reconhecimento das derivadas das funções elementares, da forma binomial e da série geométrica, por exemplo.

A aprovação em Cálculo não requer o nível de proficiência relacionado à fluência, naturalidade, espontaneidade, ações e reconhecimento com os conteúdos, apenas afirma ser prova de que se pode resolver os problemas padrões de Cálculo. E isso não é prova de que um aluno adquiriu a preparação necessária para o sucesso futuro. (BRESSOUD; MESA; RASMUSSEN, 2015).

Devido à carência de pré-requisitos para a disciplina de Cálculo, já que, em geral, é ministrada no primeiro semestre de diversos cursos, e o ingresso de estudantes com dificuldades de aprendizado em matemática é bem comum, o professor se vê na situação de ter de evitar exemplos que envolvam conteúdos considerados mais complexos. O professor, muitas vezes, deixa de usar relações trigonométricas mais complexas quando investiga se o aluno entendeu o processo de derivada, substituindo, por exemplo, a função $f(x) = \text{sen}(\text{arctg}(x))$, pela função $g(x) = \text{sen}(2x)$. Segundo expõe Rezende (2003), as duas questões teriam o mesmo grau de dificuldade, já que tratam da habilidade de resolver regra da cadeia, e avaliam o mesmo assunto, isto é, a derivada da composição de duas funções reais, mas a falta de compreensão pelos estudantes do significado da função $\text{arctg}(x)$, conteúdo que deveria ter sido aprendido no ensino médio, faz com que o professor acabe por evitar o seu uso. Rezende ainda comenta que ambas as questões estão avaliando apenas a parte técnica do Cálculo de derivadas, sem aprofundar na essência do conceito.

Isso relaciona-se com o que diz Penin (2001), quando comenta sobre a intencionalidade de um ensino voltado ao conteúdo enciclopédico, puramente informativo, fortemente vinculado como meio para o acesso aos vestibulares de grande concorrência, ao invés de se dar valor ao conhecimento e as competências mais gerais, que servirão tanto para inserção no mercado de trabalho quanto a continuidade dos estudos.

Um breve olhar sobre essa disciplina nos mostra dificuldades que os alunos apresentam no seu aprendizado, que, em um número significativo de casos, culmina com a não aprovação. Não é de hoje que se tem discutido no meio acadêmico a dificuldade no aprendizado de Cálculo. Ilustramos com alguns dados da disciplina de Cálculo Diferencial e Integral (CDI) da Escola Politécnica da Universidade de São Paulo (USP). No período de 1990 a 1995, os índices de reprovação situaram-se na faixa de 20% a 75%, enquanto no Instituto de Matemática e Estatística da USP, essa taxa ficou próxima a 45% (Barufi, 1999). Na Universidade Federal Fluminense (UFF), os índices são mais catastróficos e a reprovação fica na faixa de 45% a 95%, no período de 1996 a 2000 (Rezende, 2003). Esses relatos evidenciam um problema sério associado às disciplinas de Cálculo.

O elevado número de artigos que envolvem o tema, as conversas informais nos corredores dos Departamentos de Matemática e as reformas estruturais, apresenta que os professores não estão alheios a essas informações (NASSER, 2007).

Algumas reformas remontam à década de 80, quando se alterou a carga horária de cálculo de 4 créditos para 6 créditos, ou seja, de 60 horas para 90 horas semanais. Outros exemplos se referem à criação de disciplinas como matemática básica, conhecidas também por pré-cálculo ou cálculo 0, com o intuito de auxiliar e dar "embasamento" para os alunos poderem cursar o Cálculo com mais facilidade. Em alguns casos, como o de Arquitetura de algumas IES, os alunos tinham tanta dificuldade de aprovação na disciplina, que a decisão foi a mais radical possível: a sua exclusão da grade curricular. Isso ocorreu na Universidade de Brasília (UnB) e na Universidade Federal Fluminense (UFF) (REZENDE, 2003), por exemplo. Na USP, a partir de 1993, as disciplinas de CDI passaram a ser anuais, diferenciando-se das demais disciplinas (BARUFI, 1999). Isso mostra o desconforto que tais disciplinas causam para alguns cursos.

Tais observações não ocorrem apenas no nível nacional. Nos anos 1980, o famoso matemático Peter Lax, com um polêmico documento criticou os cursos de Cálculo da época e, então, viu-se deflagrar um movimento que ficou conhecido como "Calculus Reform" (REZENDE, 2003). Nos Estados Unidos da América, em 2010, o Cálculo continuava sendo um filtro para alunos de cursos superiores de ciências e engenharias. Pouco se sabia como era ensinado o cálculo nos programas mais eficazes e como promoviam os alunos em lugar de inibir a continuação de estudos. Com a expansão de vagas e o desigual aumento logístico para o ensino, os departamentos têm demorado para acolher as demandas da mudança demográfica, outros dados também merecem atenção em relação ao fracasso para minorias sub-representadas, estudantes economicamente desfavorecidos, mulheres e estudantes universitários que os pais não possuem formação acadêmica (BRESSOUD; MESA; RASMUSSEN, 2015).

No presente trabalho, analisaremos as mudanças estruturais que o Departamento de Matemática (MAT) da Universidade de Brasília (UnB) vem fazendo para melhorar o aproveitamento dos alunos na disciplina de CDI, fazendo um relato da trajetória do MAT nos últimos três anos. Abordaremos a criação da aula magistral e unificação da avaliação, a criação de um banco de itens para avaliação, embasado em uma matriz de referência formulada pelos docentes, o pré-teste e a construção da escala de proficiência para os alunos que cursam Cálculo 1. O trabalho inclui também uma entrevista semiestruturada com professores do departamento e análise de materiais utilizados.

Para tanto, buscou-se na literatura especializada artigos sobre a dificuldade em Cálculo para discutir tal procedimento, focando a análise em 4 trabalhos escolhidos por interesse: Bressoud, Mesa e Rasmussen (2015), Rezende (2003), Nasser (2007) e

Mello e Mello (2007). Apresentado por Bardin (1979), a análise de conteúdos como um conjunto de técnicas de análise de comunicações que visam obter indicadores por meio de procedimentos de descrição do conteúdo das mensagens, que permite a inferência de conhecimentos sobre essas mensagens.

A base de pesquisa para construir a avaliação de competência esteve focada nos estudos de Nagel e Richman (1976) e Perrenoud (1999). Para descrever as teorias psicométricas, foram utilizados, principalmente, os trabalhos Cohen, Swerdlik e Sturman (2014) e Rabelo (2013), enquanto, para a discussão acerca da matriz de referência como um processo metodológico para criação de um banco de itens, lançou-se mão de um recente trabalho de Marinho-Araújo e Rabelo (2015).

Com objetivo principal de apresentar uma proposta de criação de um exame de proficiência em Cálculo Diferencial e Integral, iniciou-se com a descrição do processo histórico das reformas pedagógicas ocorridas nos últimos três anos no Departamento de Matemática da UnB, utilizando-se de uma entrevista semiestruturada feita com dois atores principais desse processo, abordando os projetos existentes no MAT. As entrevistas foram analisadas segundo a proposta de análise de conteúdo proposta por Bardin (1979). Dando sequência ao trabalho, fez-se a análise quantitativa dos testes aplicados no primeiro semestre de 2015 no MAT e o que sua aplicação representou em termos de desempenho dos estudantes a partir da análise do erro. O trabalho culmina com alguns comentários acerca da construção de um modelo teórico de exame de proficiência em CDI e um relato da criação de um instrumento de feedback para os estudantes e docentes.

Este trabalho não abordará como o Cálculo é ou deveria ser ensinado. No entanto, é relevante não somente para os professores de matemática do ensino superior, mas também para os professores de ensino médio e gestores de políticas públicas, porque sinaliza para os problemas de preparação dos estudantes do ensino médio como suporte para que eles deem continuidade aos seus estudos. Não se pode esquecer que, segundo a LDB 9394/96 em seu artigo 35º, o ensino médio deve possibilitar ao estudante o prosseguimento nos estudos, como ser capaz de se adaptar com flexibilidade a novas condições de ocupação ou aperfeiçoamento posteriores, assim como o pensamento crítico e a compreensão dos fundamentos científico-tecnológicos dos processos produtivos, relacionando a teoria com a prática de cada disciplina. Naturalmente, chegamos à conclusão, conforme já nos apontam as avaliações de larga escala aplicadas no Brasil, como o Sistema de Avaliação da Educação Básica (SAEB) e a Prova Brasil, que a formação no nível médio em matemática deixa muito a desejar, deixando para as instituições de educação superior uma grande lacuna em termos de aprendizado em matemática a ser preenchida.

Capítulo 1

Breve comentário sobre avaliação

Ao tratar o tema da avaliação e os diversos atores envolvidos no processo, cabe ressaltar o sujeito que avaliará e sua configuração social e histórica, como relata D’Ambrósio (1997) em palestra na Universidade de Uberaba:

A educação mantém-se naquele estilo velho; com o paradigma estabelecido no tempo de Newton, segundo o qual para aprender é necessário ser ensinado e que ensino tem como consequência aprendizagem. Sabemos que a aprendizagem se dá de muitas formas e não como uma direta consequência do ensino, e que o ensino se faz de muitas formas além do tradicional “professor-aluno”.

Esse sujeito avaliador, com suas crenças e atitudes, muitas vezes usa o instrumento que tem em mãos como forma coercitiva frente aos avaliados, conscientemente ou não, do poder a ele atribuído. Para analisar esse processo de transição, onde se estabelecem novos conceitos e são mantidas algumas concepções, vamos referenciar no trabalho o modelo de avaliações do desenvolvimento de competências e outras ferramentas que possam ajudar a interpretar e mensurar, a exemplo da psicometria.

Como reflexão sobre a formação do professor de matemática investigada em sua tese de doutorado, Santos (2007) pontua:

A formação do professor de matemática é vista como resultado de um processo histórico-cultural que mantém ainda uma forte herança de elementos de uma sociedade colonial, corroborado pela não participação democrática do povo brasileiro em seu processo de constituição sócio-cultural numa sociedade capitalista e excludente. E o trabalho demonstra que os atuais processos de formação de professor de matemática ainda são fortemente sedimentados numa formação alienada aos ditames de uma sociedade de classes, que não permite ao futuro professor compreender e fazer uso da necessária

autonomia inerente à sua atuação, o que o faz atuar como um intelectual orgânico a serviço da consolidação da hegemonia da classe dominante. Nesse sentido, os constructos teóricos de P. Freire e de U. D'Ambrosio mostram-se como indicadores de encaminhamentos possíveis no processo de formação de um professor de matemática crítico/libertador e, por isso, consciente de sua tarefa como agente ativo na formação de um educando não especialista em matemática, mas inserido em sua realidade social como um sujeito transformador e em transformação, que encontra na matemática uma ferramenta para o processo dialético de sua própria construção (SANTOS, 2007, p. 8).

Dentro da linha de pesquisa sobre a formação de professores, verificou Santos (2007) que existe uma unanimidade de que a formação não está em conformidade com o que se espera socialmente do professor ou que há problemas na formação do professor. Também expõe que menos de 5% dos trabalhos sobre formação docente, produzidos na Faculdade de Educação da USP, no período de 1990 a 1998, relacionavam as questões políticas e profissionais da profissão docente.

Este trabalho não discutirá as questões profissionais e políticas da formação docente, porém cumpre lembrar o nosso eterno educador Paulo Freire (2014), quando diz que a não abordagem política tem sim um viés político e passa pela escolha política de defesa da manutenção da ordem vigente do mais forte.

1.1 Avaliação

Avaliar é um processo que envolve concepções, crenças, valores, princípios, teorias, conceitos, metas, desejos, trajetórias; quando tal processo tem como foco a educação torna-se potencialmente mais complexo, gerando intrincados desafios à sua verificação e registro (MARINHO-ARAÚJO e RABELO, 2015).

Nessa perspectiva é que aqui se compreende que a avaliação é um processo multifacetado, que exige muita clareza de sua intencionalidade por parte de quem avalia e de quem é avaliado. Quando se foca em um de seus instrumentos, a prova, deve-se considerar que um bom teste depende de sua solidez e de dois aspectos essenciais, entre outros, a fidedignidade e a validade. A fidedignidade envolve a consistência de mensuração, ou seja, em tese, um instrumento de mensuração perfeitamente fidedigno mede da mesma maneira com consistência e sua validade é se ele, de fato, mede o que se propõe a medir (COHEN; SWERDLIK; STURMAN, 2014, p. 123 e 124).

Entretanto, mesmo planejando e certificando as variáveis de estratégia, construção, instrumentos, conteúdos e finalidades que afetam a avaliação, nem sempre

correspondem aos critérios e fatores pretendidos, já que existe, uma rede de significados indiretos em relação ao avaliador como também ao avaliado, que podemos investigar. Vincula-se assim a avaliação na educação, com desdobramentos sociais, além do individual, ultrapassando a finalidade de classificação meritocrática. A avaliação possui, assim, características educativa, pedagógica, psicológica, com responsabilidades na formação humana e da cidadania, tanto do sujeito quanto do espaço social (MARINHO-ARAÚJO; RABELO, 2015).

A busca pela construção de competências necessárias ao perfil que se tem intenção em formar, deve-se relacionar com melhorias nas condições sociais, assim como nas alterações necessárias das propostas pedagógicas (MARINHO-ARAÚJO; RABELO, 2015).

Em Nagel e Richman (1976), vemos uma apresentação no que se refere à pedagogia do fracasso.

Honestamente adorariamos destruir este sistema de notas para fracasso — 1, 2, 3, 4, ... 10. É um pesadelo imaginar os exércitos de estudantes — do jardim da infância aos cursos superiores — que invadem os campos de batalha das salas de aula. Alguns vencerão o combate e receberão a medalha de um 10, mas, em compensação, muitos outros serão derrotados e fracassados. Assim como o Dr. Glasser (Schools Without Failure) e outros já salientaram, agora, ao reprovamos um aluno, não estamos apenas expulsando-o de nossa sala de aula, mas sim jogando-o para fora do mundo. São escassos os empregos para os despreparados, os sem talento e para os FRACASSADOS. Acredite ou não, estamos constantemente mandando nossos alunos de volta para a mesma batalha. Remendamos as pressas os nossos fracassos, suavizamos seus semblantes, murmuramos algumas palavras de conforto, todavia não fornecemos a estes soldados educacionais novos planos de guerra, novas estratégias, mais armamentos — e eles fracassam de novo, e de novo, e de novo (NAGEL e RICHMAN, 1976).

Para pactuar com esse pensamento nos vemos em uma opção de desenvolver um programa para competências dentro do atual sistema de notas, saindo do programa do fracasso.

A noção de competência é a capacidade de agir de maneira eficaz em uma determinada situação, com o uso de conhecimento, mas sem limitar-se a ele. Assim como o conhecimento é uma reprodução da realidade, que internalizamos e armazenamos ao sabor de nossa experiência, a competência ultrapassa a noção do conhecimento, dado o contexto que não basta o conhecimento para realizar uma atividade, precisa ter competência. Assim a competência utiliza, integra, ou mobiliza tais conhecimentos (PERRENOUD, 1999).

Perrenoud também complementa que o conhecimento e a inteligência não garantem uma resolução rápida a um processo inédito e complexo, pois seriam mais lentos e eficazes que uma pessoa que tem competência para resolver esse tipo de situação. A competência está relacionada além da inteligência operária em esquemas de analogias, processos intuitivos, entre outros fatores que aceleram a mobilização dos conhecimentos e favorecem a procura e criação de estratégias apropriadas (PERRENOUD, 1999).

Marinho-Araújo e Rabelo (2015) argumentam que foram inseridas no conceito de competência, na última década, as dimensões afetivas e sociais, relacionada ao desenvolvimento cognitivo e ao domínio de conhecimento científico e tecnológico. Acrescentam também que a definição de competência não é única, clara e consensual, a exemplo do que ocorre em vários documentos oficiais, nos quais há sobreposição dos termos competências/habilidades, usado de forma combinada, gerando distorções na identificação de cada item.

A competência se constitui na possibilidade de ampliação, de integração e de complementação de diversos recursos, entendidos tanto como capacidades cognitivas, afetos, desejos, que são subjetivos, e aos saberes, conceitos, posturas, atitudes. Assim, vai muito além das habilidades e dos conhecimentos. E tais elementos compreendem-se como componentes de uma competência, mas não de forma exclusiva a ela, pois são mobilizados por outros recursos e em contextos diferentes (MARINHO-ARAÚJO; RABELO, 2015).

A construção de competência está junto com a formação de esquemas de mobilização dos conhecimentos, em tempo real, com discernimento ao serviço de uma ação eficaz, e sua eficácia associada a uma postura reflexiva (PERRENOUD, 1999). Não podemos fixar o desenvolvimento da competência ao treinamento de um conjunto de capacidades ligado à ação técnica profissional. É necessário compreendê-la em uma dimensão ampla, contemplando não só aspectos racionais, cognitivos ou mentais, mas também processos intersubjetivos, afetivos, socioculturais (MARINHO-ARAÚJO; RABELO, 2015).

Com base nessa conceituação, conhecimento e competência, ou habilidade e competência não se confundem. Quando o tempo é limitado podem causar disputa de privilégio na dedicação de cada construção, pois o treinamento, que gera os esquemas de mobilização, requer tempo para o sujeito viver experiências e analisá-las (PERRENOUD, 1999). Assim, em um tempo limitado não se faz possível cobrir uma quantidade numerosa de conteúdos, senão deixando de lado a construção de uma fração considerável da bagagem de competências. Devido ao nível de complexidade das ações humanas, que exigem o aprofundamento ou a organização de conhecimentos, um tempo considerável é necessário para o desenvolvimento de competências (MARINHO-ARAÚJO; RABELO, 2015).

Como citado que a proficiência necessária para o sucesso em Cálculo, perpassa a proficiência necessária para a aprovação, é utópico acreditar na conciliação de que o aprendizado sequencial de conhecimentos culmina de forma espontânea sua integração operacional total em uma competência. E não podemos também arquitetar uma construção dialética entre o conhecimento e a competência.

A conceituação de competência é muito ancorada no mundo do trabalho, na inovação pela busca da eficiência dentro das empresas, e essa pode ter uma ligação com o surgimento dessa discussão no âmbito escolar (PERRENOUD, 1999; CASTIONI, 2010). Nos anos de 1970, o conceito de competência relacionava-se com o saber fazer, vindo de práticas de trabalhos sem a articulação com o conhecimento teórico, devido principalmente pela divisão técnica e fragmentada do trabalho (MARINHO-ARAÚJO; RABELO, 2015). Segundo esses autores,

As demandas por melhoria da qualidade dos produtos e de flexibilização dos processos de produção exacerbaram a competição e acabaram por utilizar a educação como meio de incrementar ou adaptar aquisições individuais às exigências do mercado. O conceito de competência acompanhou essas demandas, sendo redefinido a partir de critérios operacionais, observáveis, aparentes, mensuráveis. Dessa forma é que, especialmente no contexto das organizações ou empresas, o termo competência foi sendo associado a uma variedade de atributos como capacidades, aptidões, qualificações que seriam adequados e esperados à execução de determinadas atividades profissionais. Treinar e monitorar comportamentos favoráveis à consecução de metas passou a representar o que se entendia por desenvolver competências (MARINHO-ARAÚJO; RABELO, 2015).

Cita Perrenoud (1999) que o sistema educacional tem sido construído a partir de cima, onde um grau mais alto de formação induz os horizontes da formação dos anos anteriores. Observando o surgimento do trabalho com competência, nos primeiros ciclos de ensino, nesse contexto vemos ao contrário poucas faculdades que tenham documentados os objetivos de uma formação universitária e menos ainda formulado em uma linguagem de competência.

Em uma breve busca de cursos que trabalham por competência, encontramos previsto nas diretrizes curriculares nacionais (DCN) de diversos cursos que o perfil profissional de formação deveria ser alcançado por meio do desenvolvimento de diversas competências/habilidades. Apenas a título de exemplo, cita-se o seguinte trecho das DCN para os curso de Matemática, de acordo com o Parecer CNE/CES 1.302/2001:

Os currículos dos cursos de Bacharelado/Licenciatura em Matemática devem ser elaborados de maneira a desenvolver as seguintes competências e habilidades.

- a) capacidade de expressar-se escrita e oralmente com clareza e precisão;
- b) capacidade de trabalhar em equipes multi-disciplinares;
- c) capacidade de compreender, criticar e utilizar novas ideias e tecnologias para a resolução de problemas;
- d) capacidade de aprendizagem continuada, sendo sua prática profissional também fonte de produção de conhecimento;
- e) habilidade de identificar, formular e resolver problemas na sua área de aplicação, utilizando rigor lógico-científico na análise da situação-problema;
- f) estabelecer relações entre a Matemática e outras áreas do conhecimento;
- g) conhecimento de questões contemporâneas;
- h) educação abrangente necessária ao entendimento do impacto das soluções encontradas num contexto global e social;
- i) participar de programas de formação continuada;
- j) realizar estudos de pós-graduação;
- k) trabalhar na interface da Matemática com outros campos de saber.

A diretriz acrescenta ainda no que se refere às competências e habilidades próprias do educador matemático, o licenciado em Matemática deverá ter as capacidades de:

- a) elaborar propostas de ensino-aprendizagem de Matemática para a educação básica;
- b) analisar, selecionar e produzir materiais didáticos;
- c) analisar criticamente propostas curriculares de Matemática para a educação básica;
- d) desenvolver estratégias de ensino que favoreçam a criatividade, a autonomia e a flexibilidade do pensamento matemático dos educandos, buscando trabalhar com mais ênfase nos conceitos do que nas técnicas, fórmulas e algoritmos;
- e) perceber a prática docente de Matemática como um processo dinâmico, carregado de incertezas e conflitos, um espaço de criação e reflexão, onde novos conhecimentos são gerados e modificados continuamente;
- f) contribuir para a realização de projetos coletivos dentro da escola básica.

Ao mesmo tempo que o exemplo mostra a sinalização de formação para o desenvolvimento de competência, ele também exemplifica o conflito conceitual presente em um documento oficial, chancelado pelo Ministério da Educação, pois claramente há confusão sobre o que seja competência com o que é habilidade e mesmo conhecimentos.

Perrenoud ainda constrói que é ingênuo querer catalogar a busca do mundo escolar pela competência, por uma simples dependência ideológica em relação à política

econômica, como mostra a influência que vem por vários setores para essa construção (PERRENOUD, 1999, p. 12 a 14).

Essas alterações pedagógicas, nas estruturas curriculares da educação básica e nas diretrizes curriculares da educação superior, mostram a necessidade de investigar os indicadores da trajetória de formação educativa, expressa em competências, habilidades, conhecimentos e saberes fundamentais.

Há que se desenvolver processos avaliativos que consigam evidenciar a forma pela qual ocorre a articulação teoria e prática, bem como indicadores de como as competências se constroem, vinculadas às subjetividades individuais e sociais próprias às relações e aos contextos vivenciados. (MARINHO-ARAÚJO; RABELO, 2015)

Alguns itens relativos à construção de competência, apresentados por Nagel e Richman (1976), serão tratados nos próximos parágrafos. Iniciamos pelas pesquisas que apontam que 95% dos nossos alunos podem aprender quase tudo, se lhes for proporcionado tempo suficiente. Assim, na prática pedagógica, carece de modificação a tradicional organização do horário a fim de permitir tempo suficiente para o aluno obter maestria nos objetivos, no sentido de possibilitar a cada estudante o tempo que ele necessita para aprender. Cabe ao educador ser obstinado e vigoroso, não deixando que o pensamento do fracasso passe pela sua cabeça ou a dos alunos.

Para a construção coletiva de um trabalho de competência, deve-se buscar o apoio de outros professores, verificar com os pares quais as piores coisas que poderiam ocorrer abandonando o sistema de notas para fracasso, com isso o grupo de professores poderá ver que elas não são tão terríveis assim. E, quem sabe, também optem por trabalhar por competência.

O professor deveria ser um conselheiro na sala de aula, com objetivos claros e comunicados abertamente aos alunos, pois, assim, proporcionará um ensino planejado especialmente para sanar suas deficiências e levá-los a atingir maestria no curso. Seus itens passarão a ser um instrumento imparcial, uma vez que visam apenas testar os alunos naqueles objetivos comunicados abertamente.

Nessa jornada de trabalhar por competência, terá os procedimentos planejados para avaliação e que refletem adequadamente os objetivos. Podendo assim trabalhar os objetivos visto de forma completa ou o que não foi atingido ainda, sendo incompleto, abandonando a nota ou o reprovado. Não havendo recuperação como castigo, mas uma nova chance de avaliação para atingir o objetivo, construir o que não foi feito e não refazer o que já foi feito erroneamente. Acompanhar a “prontidão” diagnosticada para uma específica unidade de ensino, a fim de quando tiver que passar notas numéricas, publicar uma lista de competências e relacioná-las às notas.

Sem medo do fracasso, o professor pode pedir mais tempo extra nos estudos aos seus alunos, para que atinjam os objetivos. Se eles acreditam que podem ter um sucesso com horas a mais de estudos, poderão surpreender. Com um ensino aperfeiçoado, o professor descobrirá que os alunos não precisarão de todo o tempo previsto, para o encontro com o sucesso.

Este platô de um ensino para competência, relatado nos últimos cinco parágrafos, é comumente chamado de Ensino Relativo ao Critério (Criterion Referenced Instruction) (NAGEL; RICHMAN, 1976, P 47 a 50).

Na avaliação referenciada ao critério, o desempenho do testando é verificado no sentido de que ele poderá ou não realizar, o que aprendeu ou não, o que satisfaz ou não satisfaz de critério para inserção em um grupo determinado. A abordagem referenciada ao critério tem ampla aceitação no campo de programas educativos assistidos por computadores, uma vez que nesses programas, antes de encaminhar o testando para uma nova etapa é analisado o domínio do material naquele segmento (COHEN; SWERDLIK; STURMAN, 2014, p. 139).

A evolução do sistema educacional rumo ao desenvolvimento de competências é uma hipótese digna de maior atenção. Talvez seja essa a única maneira de “dar um sentido à escola”, para salvar uma forma escolar que está esgotando-se sem que seja percebida, de imediato, alguma alternativa visível (PERRENOUD, 1999, p. 33).

Essa evolução é difícil, pois exige importantes transformações dos programas, das didáticas, da avaliação, do funcionamento das classes e dos estabelecimentos, do ofício do professor e do ofício do aluno. Essas transformações suscitam a resistência passiva ou ativa por parte dos interessados, de todos aqueles a quem a ordem gerencial, a continuidade das práticas ou a preservação das vantagens adquiridas importam muito mais do que a eficácia da formação (PERRENOUD, 1999, p. 34). Por outro modo, Nagel (1976) posiciona-se argumentando que se pode utilizar a aula do mesmo modo que já havia sendo feito, apenas alterando a avaliação.

A avaliação de competência vem ao encontro do processo de revisão da prática avaliativa, como também do processo de ensino-aprendizagem. É preciso investigar aspectos relativos aos procedimentos de ensino; à organização de tempos, espaços e métodos; à negociação e à partilha de responsabilidades e encaminhamentos ao aprimoramento do processo educativo. A avaliação liga-se tanto sobre a perspectiva individual, como na sua dimensão coletiva. A abordagem por competência permite visualizar melhor a construção do conhecimento até o passo de atingir ações do cotidiano. Oportuniza aprender a identificar, mobilizar, gerenciar e utilizar esses conjuntos de recursos, articulados a habilidades, saberes, conhecimentos e outras características pertinentes. Tal processo nem sempre é rápido e prático, tem um caráter não limitado, por isso enriquecido por circunstância que podem ser únicas em outros momentos rotineiras e habituais. Na avaliação de competência, deve-se observar o processo dinâmico do

aprendizado, reservando tempo da vivência nas experiências e de análise dessas, assim conscientizar-se e ponderar sobre a utilização de recursos e habilidades, e sobre as distintas formas de mobilizá-los e avaliá-los (MARINHO-ARAÚJO; RABELO, 2015).

Com o intuito de fundamentar procedimentos em avaliação de competências, indicadores, ações e estratégias, vamos necessitar da utilização de matrizes enquanto instrumentos metodológicos orientadores à identificação e construção de tais competências, conforme será feito na próxima seção.

1.2 Matrizes de referência

As matrizes de referência são instrumentos utilizados como suporte teórico-metodológico para avaliação de competência. Na literatura da área, encontramos diferentes modelos epistemológicos que fundamentaram as concepções de competência e deram suporte para a construção dessas matrizes. Essas matrizes foram elaboradas de acordo com o contexto, o ambiente e as crenças da sua época (MARINHO-ARAÚJO; RABELO, 2015).

Para a construção da matriz, temos que analisar o objetivo almejado de formação que queremos avaliar. Temos que balizar o perfil profissional a partir das literaturas ou de análises do processo histórico de formação, como a consulta de documentos, projetos e legislações. Entrevistas, observações de especialistas e outras metodologias podem também dar subsídio para a construção do perfil.

Com o perfil a ser atingido, pré-estabelecido, identificam-se os recursos a serem mobilizados para atingir as competências esperadas. Tais componentes fundamentam a matriz de referência tridimensional, que relaciona as linhas horizontais, que contemplam as características harmônicas ao perfil esperado, com as colunas verticais, onde são inseridos os diversos recursos a serem avaliados. A interligação dessas resume os objetivos da avaliação, permitindo a ligação dos perfis com os diversos recursos a serem mobilizados (MARINHO-ARAÚJO; RABELO, 2015).

Algumas matrizes, como a do Sistema de Avaliação da Educação Básica (SAEB), são compostas por um conjunto de descritores, que exibem os saberes a serem apreendidos no processo de aprendizagem. Esses descritores articulam os objetos de conhecimento com as operações mentais necessárias para a habilidade avaliada (RABELO, 2013).

Entre as etapas fundamentais do processo de elaboração de uma matriz de referência, inserem-se a validação teórica e a validação empírica. A primeira contempla a análise semântica, por profissionais da área, que revisam a clareza do propósito enunciado em cada elemento descrito na matriz, enquanto, na segunda, avalia-se a capacidade dos itens em evidenciar as relações estabelecidas na matriz de referência,

feita também pelos especialistas da área, ou fazendo-se uma pré-testagem, utilizando processos estatísticos e psicométricos para dar uma maior confiabilidade às análises (MARINHO-ARAÚJO; RABELO, 2015).

Encontramos a apresentação da matriz de referência criada pelos professores do MAT, para auxílio na construção do banco de itens calibrados de CDI, que foi elaborada no primeiro semestre de 2014. Para a construção optou-se pela matriz construtivista, por revelar a dimensão processual e contextual na investigação das competências, que está evidente nas características do perfil de formação desejado. A proposta é de uma matriz tridimensional, que articula as características do perfil de formação com as habilidades que contribuem para que essas características se manifestem e os objetos de conhecimento de CDI por meio dos quais as habilidades serão avaliadas (RABELO; MAGALHÃES; FURTADO; MARINHO-ARAÚJO, 2015). Nessa matriz, são elencadas as seguintes habilidades esperadas para o alcance do perfil profissional estabelecido:

- H1. Identificar linguagens e traduzir sua plurissignificação;
- H2. Interpretar diferentes representações de um mesmo conceito, transitando por representações simbólicas, gráficas e numéricas, entre outras;
- H3. Inter-relacionar objetos de conhecimento em diferentes áreas;
- H4. Organizar estratégias de ação e selecionar métodos; H5. Ler e interpretar dados e informações e expressar-se com clareza e precisão;
- H6. Aplicar métodos adequados para análise e resolução de problemas;
- H7. Formular e articular argumentos adequadamente;
- H8. Fazer inferências indutivas, dedutivas e analógicas;
- H9. Formular hipóteses e conjecturas, prever resultados e fazer generalizações;
- H10. Analisar criticamente a solução encontrada para uma situação-problema;
- H11. Confrontar possíveis soluções para uma situação-problema.

Utilizando-se como base as diretrizes curriculares nacionais para os cursos de graduação, os professores do MAT estruturaram, para o perfil de formação desejado, as seguintes características:

- P1. Concebe a matemática como um corpo de conhecimentos rigoroso, formal e dedutivo, produto da atividade humana, historicamente construído;
- P2. Domina os conhecimentos matemáticos e compreende o seu uso em diferentes contextos interdisciplinares;
- P3. Identifica, formula e soluciona problemas;
- P4. Valoriza a criatividade e a diversidade na elaboração de hipóteses, de proposições e na solução de problemas;

P5. Analisa criticamente a contribuição do conhecimento matemático para a sua formação.

Durante o processo de validação semântica e empírica, duas das habilidades inicialmente listadas foram fundidas, tornando-se as onze habilidades e as cinco características apresentadas acima, conforme elucidam os autores.

Os objetos de conhecimento contemplam os seguintes tópicos: limites, derivadas, máximos e mínimos, integrais e aplicações, de acordo com o explicitado a seguir.

Os objetos de conhecimento foram selecionados dos conteúdos curriculares das disciplinas em que serão aplicados os testes e explicitados em quais cruzamentos ou células da matriz eles se inserem. Desse modo, a matriz poderá ser utilizada em outras disciplinas, já que as dimensões referentes ao perfil e às habilidades são transversais à formação. (RABELO; MAGALHÃES; FURTADO; MARINHO-ARAUJO, 2015)

Relacionando as características do perfil com as habilidades, são estabelecidas as possíveis células de ligação, de acordo com a Tabela 1.1, marcadas por um “X”.

A síntese desse complexo cruzamento é que dará origem ao item da prova, que precisará dar conta da articulação proposta em cada célula da Matriz.

Tabela 1.1: Relações entre as habilidades (Hi) e as características do perfil (Pi)

Características do perfil	Habilidades										
	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11
P1	X	X	X				X				X
P2	X	X	X		X	X	X		X	X	X
P3	X	X	X	X	X	X	X	X	X	X	X
P4		X	X	X		X	X	X	X		X
P5			X			X	X	X	X		

Fonte: RABELO, MAGALHÃES, FURTADO & MARINHO-ARAUJO, 2015.

No capítulo 3 iremos abordar como foi a utilização das matrizes para formar o banco de itens, o processo de calibração dos itens, contando um pouco da história do desenvolvimento desse processo no MAT. Para dar suporte a tal desenvolvimento, precisamos trabalhar um pouco sobre a TCT e TRI, o que será feito no próximo capítulo.

Capítulo 2

Psicometria: Conceitos Básicos

Cápítulo desenvolvido por:

Fabio Henrique Gomes

Josué Sergio de Souza

Professores de Educação Básica da SEEDF e egressos do PROFMAT

2.1 Breve Histórico

Psicometria é uma área da Psicologia que consiste no conjunto de técnicas utilizadas para quantificar um conjunto de comportamentos que se deseja conhecer melhor. Para Pasquali(2008), a psicometria caracteriza-se por expressar numericamente um fenômeno psicológico. Para ele,

a psicometria fundamenta-se na teoria da medida em ciências para explicar o sentido que têm as respostas dadas pelo sujeito a uma série de tarefas e propor técnicas de medida dos processos mentais.

É um ramo especializado da Psicologia que se dedica ao estudo e elaboração de testes de avaliação psicológica e ao desenvolvimento e aplicação dos conhecimentos estatísticos e de outros processos matemáticos à psicologia. É uma área da psicologia com uma concepção estatística que explica os comportamentos e aptidões por meio de testes cuja mensuração é feita através das respostas que os indivíduos fornecem a uma série de tarefas, tipicamente chamada de itens. Para isso, há duas teorias que serão descritas a seguir: a Teoria Clássica dos Testes (TCT) e a Teoria de Resposta ao Item (TRI).

Pasquali (2008) apresenta a psicometria nas suas origens etimológica e histórica. Embasada no método quantitativo, a psicometria representa a teoria e a técnica de medida dos processos mentais, aplicada, em especial, na área de Psicologia e da Educação. Historicamente, a psicometria tem suas origens na psicofísica dos psicólogos alemães Ernst H. Weber e Gustav Fechner. Porém, quem é considerado criador da psicometria é o inglês Francis Galton, que contribuiu criando testes para medir

processos mentais. Leon Louis Thurstone criou a análise fatorial múltipla e possibilitou diferenciar a psicometria da psicofísica. Daí, a psicometria ficou definida como medida do comportamento do organismo por meio de processos mentais (Lei do julgamento comparativo).

Muñiz (2003), Pasquali (2008) e Andriola (2009) descrevem as origens da TCT e da TRI, dos quais segue breve resumo. A TCT tem a sua origem intimamente associada ao nome de Spearman, discípulo de Galton, e foi axiomatizada por Gulliksen, em 1950. A TRI, por sua vez, foi inicialmente elaborada por Lord e por Rasch e, finalmente, axiomatizada por Birnbaum e por Lord. F. M. Lord, em sua tese doutoral dirigida por Gulliksen, na qual apresentou uma sistematização dos principais conceitos e modelos de TRI, que começaram a surgir com o trabalho de Thurstone publicado em 1925. G. Rasch, em 1960, propôs o modelo logístico de um parâmetro, no qual a probabilidade de acerto a um item é influenciada pela sua dificuldade. A. Birnbaum propôs novos modelos logísticos, fundamentados na ogiva normal de F. M. Lord, que permitiu a geração de novos procedimentos visando a sua aplicação prática. Um desses modelos, denominado modelo logístico de três parâmetros, será apresentado na próxima seção.

2.2 Teoria Clássica dos Testes (TCT) e Teoria de Resposta ao Item (TRI)

Alguns parâmetros importantes utilizados como suporte nessas teorias são a discriminação de um item, que tem relação com a diferenciação do sujeito com um grau mais alto ou mais baixo da capacidade de responder ao item, e a dificuldade, que pontua o atributo de um item não ser facilmente realizado, resolvido ou compreendido.

Vejamos um suposto teste sobre a dificuldade física em um hipotético teste de três itens: item 1, a pessoa anda de um cômodo a outro da casa; Item 2, a pessoa necessita de ajuda para levantar, sentar e caminhar; item 3, a pessoa é capaz de correr 1,5 km todos os dias. Quem responde verdadeiro ao item 1 e falso ao item 2 tem mais mobilidade de que quem responde verdadeiramente o item 2 e falso ao item 1. Na TCT poderíamos avaliar cada item com a mesma pontuação, já na TRI, entretanto, os itens podem ser contabilizados com pesos diferentes. Assim o item 3 receberia maior pontuação que os demais, e o item 1 mais pontos que o item 2 quando o critério é o da mobilidade (COHEN; SWERDLIK; STURMAN, 2014, p. 168 e 169).

O modelo clássico da psicometria tradicional (Pasquali, 1999) está fundamentado na TCT. Esta tem como objetivo a determinação das propriedades ou dos parâmetros do teste como um todo. As aptidões são medidas, em geral, pela soma das pontuações referentes às respostas dadas a uma série de itens, expressa no chamado

escore total, revelando que o objetivo é explicar o resultado final total.

A TCT é clássica por ser uma teoria bem estabelecida distinta de outras mais atuais, especificamente da que se denomina Teoria de Resposta ao Item (TRI).

A TCT considera os testes como um conjunto de estímulos comportamentais (itens) cuja qualidade é definida em termos de um critério; este, por sua vez, é representado por comportamentos presentes ou futuros. A TCT está apoiada no seguinte paradigma: o escore empírico ou bruto do sujeito é constituído de dois componentes: o escore real ou verdadeiro do sujeito avaliado (V) e o erro de medida (E). Daí, surge o modelo fundamental da psicometria, elaborado por Spearman e detalhado por Gulliksen, o qual confirma a tese de que o escore bruto de um examinado é a soma do escore verdadeiro e do erro ($T=V+E$). Deste modelo, decorrem alguns postulados básicos:

- o escore esperado é o escore verdadeiro. Isto decorre do conceito de esperança matemática do escore empírico, ou seja, se o sujeito responde infinitas vezes ao mesmo teste, ele terá infinitos diferentes escores empíricos, e a média destes infinitos escores será o escore verdadeiro, porque ela eliminaria os erros;
- não há correlação entre o escore verdadeiro e o erro, pois a correlação entre o escore verdadeiro e o erro é zero; portanto, não há nenhuma razão para supor que escores verdadeiros maiores terão erros positivos e escores verdadeiros menores terão erros negativos;
- erros em testes paralelos não são correlacionados.

O modelo da TCT é embasado em dados empíricos coletados de um conjunto de itens agrupados inicialmente de maneira intuitiva. A definição dos itens, que comporão o teste, é feita por meio da validade aparente, ou seja, escolhem-se aqueles itens que parecem estar medindo a mesma coisa. Na TCT, os parâmetros do item e da habilidade são dependentes da amostra e do teste.

O parâmetro importante a ser analisado, utilizando-se na TCT, é a dificuldade dos itens que compõem um teste. Esta pode ser definida como a porcentagem de examinados que respondem corretamente aos itens. O cálculo da dificuldade de cada item é feito dividindo-se o número de pessoas que acertaram o item pelo número total de pessoas que responderam. Para fins de avaliação de larga escala, os testes devem ser compostos de itens que alcancem todo o *continuum* da escala, ou seja, devem ter uma amplitude que inclua itens fáceis, medianos e difíceis (VIANNA, 1989).

Outro parâmetro é a discriminação dos itens, que se refere ao poder que um item possui para distinguir sujeitos com magnitudes de traços diferentes, do qual o item constitui a representação comportamental (PASQUALI, 1999). Quanto mais próximas forem as magnitudes do traço que o item puder diferenciar, mais discriminativo ele

será. Estatisticamente, esse conceito, na TCT, representa a correlação dos escores dos sujeitos no item com seus escores no teste total.

Dos diversos índices propostos para medir o poder de discriminação de um item, a correlação bisserial é o mais comumente utilizado nas avaliações de larga escala no Brasil. A correlação bisserial é uma versão da correlação ponto-bisserial mais apropriada quando uma das variáveis é dicotômica ou binária (assume apenas dois valores) (RABELO, 2013). Esta é uma medida de associação entre o desempenho no item e o desempenho no teste. A correlação bisserial é menos influenciada pela dificuldade do item e tende a apresentar menos variação de uma situação de testagem para outra (WILSON; WOOD; GIBBONS, 1991). Sua fórmula é:

$$r_b = \frac{\bar{M}_p - M}{S} \frac{p}{h(p)}, \text{ em que}$$

\bar{M}_p = média no teste dos sujeitos que acertam o item (p)

M = média total do teste

S = desvio-padrão do teste

p = proporção de sujeitos que acertam o item

$h(p)$ = é a ordenada da curva normal no ponto de divisão dos segmentos que contém as proporções p dos casos.

A Análise Gráfica do Item (AGI) é o gráfico que descreve, dentro dos pressupostos da teoria clássica, o comportamento do indivíduo frente a cada item que ele responde. Esta análise pressupõe que um aluno que dá uma resposta certa em um item de múltipla escolha adquiriu mais habilidades que um aluno que dá resposta errada e também que um aluno com mais itens certos adquiriu mais habilidades que um aluno com menos itens certos. Para elaborar esse gráfico, organiza-se os estudantes em ordem crescente de escore bruto no teste. Depois, calcula-se, por faixa de notas, a proporção de marcação por opção de resposta em um item fixado do teste. Em um item bem elaborado, o percentual de escolha da opção correta, correspondente ao gabarito, vá aumentando à medida que o desempenho global no teste também aumenta e que, em relação às incorretas, ocorre exatamente o inverso (RABELO, 2013).

A TCT é o modelo mais usado e aceito na literatura psicométrica hoje por causa da sua fácil utilização e simplicidade quando comparado à TRI. Além disso, para a aplicação da TRI são necessários bem mais indivíduos do que para a TCT.

A Teoria de Resposta ao Item (TRI) é a abordagem que concentra-se na relação entre a resposta de um indivíduo a um item de um teste e a posição desse indivíduo, em termos probabilísticos, da aptidão que está sendo medida, ou seja, possui como foco o estudo individualizado dos itens de um teste. Na TRI são medidos os traços latentes, que são características intrínsecas dos indivíduos que não podem ser medidas diretamente, isto é, não podem ser observadas fisicamente (COHEN; SWERDLIK; STURMAN, 2014, p. 165 e 170).

A independência do instrumento de medida em relação ao objeto medido nos teste de inteligências é um dos motivos para o surgimento da TRI. Embora seja uma teoria que surgiu por volta da década de 50 do século passado, foi nos anos 80 que começou a ser difundida, com o avanço do desenvolvimento de *softwares* para uso prático dos algoritmos (RABELO, 2013, p. 126). Portanto, a TRI preocupa-se com o estudo das características métricas dos itens, utilizando, para tanto, uma escala microscópica. Já a TCT tem seu foco direcionado ao próprio instrumento de medida e emprega, para tal, uma escala macroscópica (ANDRIOLA, 2009). Enquanto a TCT tem interesse em produzir testes de qualidade, a TRI se interessa por produzir tarefas (itens) de qualidade.

Do ponto de vista do resultado, a TCT o expressa pelo score bruto ou padronizado, enquanto a TRI o faz como proficiência estimada pelo modelo estatístico. Em relação à base numérica da comparação, a TCT permite comparação de escores apenas entre indivíduos submetidos à mesma prova enquanto a TRI permite a comparação de indivíduos submetidos a provas diferentes, mas incluídos em uma escala comum. Segundo Rabelo (2013, p.126), a

A utilização da TRI para análise de teste de conhecimento veio para sanar algumas limitações da TCT, principalmente no que diz respeito a discriminação dos itens, fidedignidade dos teste e comparabilidade de desempenho de indivíduos que se submetem a testes diferentes

O pressupostos da TCT são prontamente satisfeitos e caracterizados como “fracos”, em contrapartida com os da TRI caracterizados como “fortes”, “difíceis”, “rigorosos” e “robustos”. (COHEN; SWERDLIK; STURMAN, 2014, p. 166)

Há grande variedade de modelos da TRI, cada um projetado para lidar com dados apresentando certos pressupostos e características distintas. (COHEN, SWERDLIK, STURMAN, 2014, p. 168) O modelo apresentado para este trabalho é o modelo logístico unidimensional de 3 parâmetros que permite que seja estimado o nível de aptidão, a partir da relação da probabilidade de acertar um item em função da sua habilidade (Θ), da dificuldade (b), da discriminação (a) e da probabilidade de acerto ao acaso (c). A probabilidade de um indivíduo j responder a um item i , será definida por:

$$P(X_{ij} = 1|\Theta_j) = \frac{(1-c_i)}{1+\exp(-Da_i(\Theta_j-b_i))}$$

Em que

- X_{ij} é a resposta do indivíduo j ao item i (igual a 1, se está correta, e igual a 0, se incorreta);

- $a_i > 0$ é o parâmetro de discriminação do item, que para TRI, é definida como o poder do item para diferenciar indivíduos com magnitudes próximas da habilidade que está sendo aferida, na Curva Característica do Item (CCI) relaciona-se com um valor proporcional à inclinação da curva no ponto de inflexão, e apresenta-se negativa quando viola a monotocidade do item;
- b_i é o parâmetro de dificuldade (ou posição) do item, medido na mesma escala de habilidade, define o mínimo de proficiência que o estudante precisa para ter uma alta chance de acertar a resposta. É o valor necessário para ter uma probabilidade de acerto igual a $\frac{(1+c)}{2}$, ou seja quando o acerto ao acaso é zero, a dificuldade é o valor de habilidade necessário para ter uma chance de 50% para acertar o item. Como a dificuldade está na escala de proficiência que tem variação de $-\infty$ a $+\infty$, na prática estarão situados entre -3 e 3, pois, é aí que se localizam cerca de 99,7% dos casos e os itens fora deste padrão são considerados que sugerem problemas de concepção, itens com valores mais altos são considerados itens mais difíceis;
- $0 < c_i < 1$ é o parâmetro de acerto ao acaso, que indica a chance de acerto de um indivíduo que não sabe responder ao item e acaba respondendo corretamente mas de maneira aleatória. Ele indicará se de algum modo a opção verdadeira se mostra atrativa ao indivíduo de baixo desempenho. Corresponde na CCI ao ponto em que a assíntota horizontal inferior da curva intercepta o eixo das probabilidades, refletindo as chances de um estudante de proficiência muito baixa selecionar a opção de resposta correta;
- Θ_j representa a habilidade ou traço latente do j-ésimo indivíduo;
- \exp representa a função exponencial;
- D é um fator da escala, que é igual a 1 na métrica logística e igual a 1,7 na métrica normal.

Fazendo-se um paralelo com a AGI, o gráfico da função de probabilidade expressa na equação anterior será denominado de Curva Característica do Item (CCI), que terá o formato de uma sigmoide, curva em formato de “S” que fornecerá diversas informações sobre a qualidade do item.

A TRI impõe alguns pressupostos na construção dos itens, como a unidimensionalidade e a independência local, ou seja, cada item deve trabalhar uma aptidão de forma dominante e não ser dependente da resposta de outro item (RABELO, 2013). Outro pressuposto é da monotocidade, ou seja, a probabilidade da resposta correta a um item deve aumentar à medida que aumenta o desempenho dos indivíduos no teste

como um todo. Quando a TRI se classifica como robusta tem correlação com a mínima violação desses três pressupostos (COHEN; SWERDLIK; STURMAN, 2014, p. 170).

As utilizações de itens âncoras, ou itens de ligação faz com que seja possível comparar os resultados de uma prova com outras, ou mesmo aplicar provas diferentes para grupos distintos. Para a comparação dos resultados de forma mais direta constrói-se uma escala de proficiência, a proficiência estimada pelo modelo estatístico, que a TRI torna possível através da atribuição de escores e da análise. Uma base numérica de comparação, colocando os indivíduos mesmo de provas diferentes sobre a mesma escala (RABELO, 2013, p. 126 à 129).

Para construção da escala de proficiência é necessário escolher uma origem, que é estabelecido no valor médio, e uma escala de medida o desvio-padrão, por exemplo, ambos obtidos das proficiências dos indivíduos que responderam ao teste. Mantem-se uma relação de ordem entre os valores da escala, quando se comparam duas escalas diferentes, sempre relacionados com o valor médio e o desvio-padrão.

Os parâmetros dos itens e das habilidades são obtidos por sofisticados cálculos matemáticos, que somente podem ser feitos com o uso de computador, tomando-se como base uma matriz de respostas que traz em suas colunas as respostas dos itens que compõem o teste e cada uma de suas linhas refere-se a um indivíduo que respondeu ao teste. Esses cálculos são feitos por *softwares* específicos que utilizam de técnicas estatísticas e matemáticas de estimação, entre os quais se destacam o BILOG e o BILOG-MG.

Para uma análise qualitativa da escala numérica, na TRI, faz-se necessário um processo denominado de interpretação pedagógica das classes de proficiência. O processo consiste na identificação de itens âncoras, representativos de cada nível da escala, a partir dos dados psicométricos e na interpretação pedagógica feita por especialistas após a aplicação do teste relativa ao real significado da avaliação proposta em cada item. Com esse processo, localizam-se assim os itens na escala de proficiência assim como se faz com os indivíduos. Esse é um dos grandes ganhos da TRI, que permite o estabelecimento de um *feedback* qualitativo para os estudantes, interpretando-se o significado da proficiência numérica em termos de aprendizado. (RABELO, 2013, p. 143-145).

Capítulo 3

Histórico do projeto da aula magistral e do banco de itens para exame de proficiência em Cálculo

Para obter informações sobre os projetos que têm relação com o curso de Cálculo do Departamento de Matemática da Universidade de Brasília (MAT) foram realizadas pesquisas em artigos publicados e uma entrevista semiestruturada, com dois dos professores do MAT responsáveis pela implantação dos projetos.

Os entrevistados assinaram um termo de consentimento livre e esclarecido e responderam a 35 perguntas. Para análise, foi realizada a transcrição na íntegra e, em seguida, a textualização, de forma a realizar um texto estruturado para apresentação em uma forma mais agradável à leitura, respeitando-se, no entanto, o ponto de vista do entrevistado. A textualização e a transcrição foram enviadas para os entrevistados para que analisassem e validassem o material quanto a possíveis distorções em relação à recolha oral.

De forma breve, Rabelo, Magalhães, Furtado e Marinho-Araujo (2016) relatam que alguns professores do MAT, para compreenderem melhor o aprendizado em matemática dos estudantes nas disciplinas do primeiro ano dos cursos de graduação, decidiram aprofundar estudos sobre as metodologias de elaboração de itens e propor um exame de proficiência em Cálculo, a partir da criação de uma matriz de referência embasada no desenvolvimento de competências e da construção de um banco de itens calibrados, utilizando, para isso, a Teoria de Resposta ao Item (TRI) e a Teoria Clássica dos Testes (TCT).

No entanto, antes de iniciarem esse trabalho, alguns professores do MAT decidiram reorganizar a dinâmica de funcionamento de algumas turmas de Cálculo 1 e Matemática 1, criando a Aula Magistral, ou o Cálculo Magistral – uma proposta que será explicada mais adiante. Um pouco dessa história será contada a partir do relato

da entrevista que será detalhado a seguir.

Durante a entrevista, algumas citações revelam como surgiram as motivações para a criação do projeto, incluindo a troca de experiências entre os professores do Departamento e suas linhas de pesquisa. Nota-se na fala do entrevistado A que a ideia para o projeto da proficiência surgiu por volta do ano de 2014:

O próprio Mauro, que sabia da TRI, comentou com a gente há uns dois anos atrás.

Com a unificação das provas de Cálculo, a proposta do exame de proficiência se fortalece e passa a incorporar mais professores. Paralelamente a esse projeto, foi ganhando forma a metodologia da aula magistral, como um dos projetos do MAT para as disciplinas Cálculo 1 e Matemática 1, conforme foi explicitado pelo entrevistado B:

(...) esse projeto [da proficiência] nasceu junto com outro que foi da unificação das provas antes de 2014-1. Cada prova pra sua turma, por exemplo, são 20 turmas, vamos pensar só no Cálculo 1(...) Na verdade, antes do magistral, eu pelo menos já tinha conversado com o Mauro Rabelo algumas coisas sobre isso, mas sim, antes era impossível né? Se você não tem uma prova unificada não tem como construir um banco de itens.

Essa fala mostra que, de fato, os dois projetos foram sendo desenvolvidos e um colaborando com o andamento do outro. Para se usar a TRI, é necessário que um quantitativo razoável de estudantes responda cada item e, assim, somente a aplicação de testes unificados possibilitaria a criação de um banco de itens calibrados pelos parâmetros da TRI. Isso é confirmado pelo entrevistado A, o qual, na realidade, aponta que o projeto do Cálculo Magistral surgiu antes do projeto da proficiência:

Aconteceu também a parte do item. O banco de itens tem que necessariamente vir depois da unificação, isso é um pré-requisito. A unificação veio com o magistral, foi um projeto só, mas necessariamente veio antes do banco de itens: o Cálculo Magistral.

Os entrevistados fizeram referência aos projetos do departamento, do Cálculo Magistral, da aprendizagem colaborativa, da unificação das provas e da criação do banco de itens para a criação de um exame de proficiência em Cálculo Diferencial e Integral. À frente veremos o relato confirmando que, no segundo semestre letivo de 2013, ocorreu a experiência das aulas com aprendizagem colaborativa. O entrevistado B explica quando e como iniciou o projeto da aula magistral, deixando claro que, no caso do Cálculo 1, o professor ministra uma aula de 2 horas na semana, para duas turmas juntas, totalizando 120 alunos, e as demais 4 horas de aula ocorrem em turmas de 60 alunos, no formato de exercícios, com acompanhamento do docente:

(...) no 2013/2 começou o magistral, então com prova unificada, aquela história toda. O primeiro teste de fato objetivo aconteceu na prova de cálculo 1 em 2014/2 (...). Aí, no 2015/1, já foram nas 3 provas que existiam na época, que não tinha teste e no 2015/2 foi o último, são 3 provas (...) porque o magistral funciona assim: o estudante tem 6 horas de aula por semana, nas primeiras horas, a primeira aula, a cada aula são 2h, você tem 2 turmas juntas no anfiteatro, então são em média 120 alunos e é uma aula por dia da semana sem muitos detalhes, e nas 2 aulas seguintes, ele se separa, hoje em turmas grandes, ainda de 60 alunos e vão trabalhar. São colocados em grupos resolvendo exercícios e aí o professor que está acompanhando a turma fica conversando e tirando dúvidas pontuais, explicando um ponto. A ideia é que o aluno possa resolver o exercício e ir aprendendo. Eventualmente o professor vê que todos estão com uma dúvida, com a mesma dúvida, ele pode ir ao quadro, falar alguma coisa, mas a ideia é que o aluno trabalhe durante 4 horas em sala e o professor em 2.

Em um apanhado histórico da implementação do processo, o entrevistado B pontua que:

ele iniciou também, mas meio como a proficiência com experiências pequenas aqui com poucas turmas, até que o coordenador viu que já dava pra (...) apresentar o projeto, oficialmente para o departamento, aí houve uma discussão lá no Colegiado; houve uma votação e a proposta ganhou. Foi consenso.

De forma temporal, temos na fala do entrevistado A:

A unificação e o magistral foram em conjunto, aliás para ser um magistral simplificado né? Essa é a ideia, ok? E depois, o banco de itens. (...) Abraçando o projeto da TRI.

Na visão dos dois entrevistados, fica claro que os projetos estão associados, conforme explicitado na seguinte fala do entrevistado A:

(...) estão associados. De fato para você ter um banco de itens que funcione, tem que ser unificado, não tem como ser separado e a unificação veio junto com o magistral, então eles estão mais ou menos juntos, apesar de serem distintos. O Departamento vê muito como uma coisa só.

Acerca da participação de professores do MAT na elaboração e na implantação dos projetos, de forma conjunta ou individual, o entrevistado B relata que,

inicialmente, foram 2 ou 3 docentes, que, por afinidade, resolveram abraçar o projeto da aula magistral juntamente com os dois entrevistados, o que incluía a unificação das provas. O projeto tem apoio de diversos professores do coletivo do MAT, inclusive do chefe do departamento e da comissão de graduação, e, hoje, envolve 10 professores, como cita o entrevistado B no trecho a seguir.

Foi alguns 2 ou 3, que por afinidade, a gente conversou e resolveram abraçar o projeto junto com a gente.(...) Hoje nós temos 10, no início éramos 8.(...) foi legal, ninguém sabia nada, só o Mauro, todos de Cálculo. (...) Acho que preciso ressaltar que as pessoas, mesmo que não participaram, mas já viram, gostaram; a gente tem apoio da chefia, sempre teve.

O entrevistado A salienta que

Na graduação, (...) a comissão de graduação também tem apoio integral. E o grupo de professores não mudou muito não, o grupo ficou bem unido em torno disso, eram os professores na época de Cálculo 1. A maioria ficou entusiasmada, tinham professores que nem estavam envolvidos com Cálculo 1 e que se interessaram pelo projeto, mas houve reações negativas, de pessoas: Então a reação foi mista, como esperado. Teve pessoas que acharam interessante, se envolveram no projeto e outras que ficaram de fora. Na elaboração de itens, os professores de Cálculo 1, em geral, aplicam os itens, mas nem todos elaboram.

3.1 A implementação

Pedido para contar sobre implementação junto a todos os professores, o entrevistado B esclarece sobre a estrutura do Departamento, que possui a figura do coordenador da disciplina, o qual desempenha a tarefa de estruturá-la.

Você tem a disciplina de Cálculo 1, que é coordenada. Tem o coordenador e ele diz mais ou menos o que é que vai funcionar, e eu acho que funcionou bem dentro do Cálculo 1, mas, de fato, a maior parte das pessoas envolvidas está no projeto. O coordenador sempre tem sido alguém do projeto, como o coordenador tem um pouco de autonomia, dentro do Cálculo 1 tem alguns que reclamam, e tal. Mas especialmente porque isso [questões objetivas] diminui o trabalho, por exemplo, o professor que não participa da elaboração, ele tem uma vantagem, porque tem parte da prova que ele não precisa corrigir mais, a nota já chega pronta pra ele. Então talvez ele não goste, mas talvez ele não reclame porque não está afetando muito a vida dele.

A barreira que pode surgir da inovação das práticas docentes, leva a ser vista de forma suavizada pela troca ocorrida no modelo e a prática que facilita em parte o processo de docência. Tentando entender motivos da barreira de inovação, vamos analisar a visão que os professores têm, como uma forma de temer o modelo de avaliação objetivo. O entrevistado A deixa claro que o fato de associar a TRI ao projeto foi fundamental para os professores aceitarem a inserção dessas questões objetivas na avaliação. Cabe lembrar também que o projeto como citado anteriormente não contempla apenas a avaliação em formato de questões objetivas, sendo que, até o momento, estas representam menos de 30% do peso das provas. Acompanhamos, assim, a seguinte fala do entrevistado A:

Agora tem uma critica grande, se a questão de item funciona, se não funciona, uma critica às provas objetivas, a introdução de itens objetivos nas provas. É um tabu isso, sempre foi. O departamento sempre olhou com suspeita, essa questão de itens objetivos. E se não fosse a TRI, que dá uma estrutura diferente, certo, provavelmente, a gente até hoje estaria sem usar esses itens. Não foi ninguém assim, vou detonar esse projeto, não teve isso, foi difícil, foi discutindo, tinha opiniões a favor, opiniões contra, mas foi, de uma maneira geral, foi bem aceito.

Relativamente à validação do projeto, como uma decisão coletiva, fica claro que ainda precisa ser avaliado e aprovado pelo Colegiado do Departamento. Sua aprovação está hoje no âmbito da coordenação da disciplina, conforme explica o entrevistado B.

Não passou ainda por uma aprovação administrativa do departamento todo, por enquanto é um projeto do Cálculo 1, que, em algum momento, vai ser levado ao departamento, aí talvez pode aparecer resistência de algumas pessoas. Essa coisa da resistência dos colegas eu não ligo muito não, porque eu acho saudável. Acho que é discutível, acho que até muda um pouco... (...) não, eu diria que o mesmo grau de mudança dentro da universidade. Tudo o que você vai mudar, tem o mesmo grau de dificuldade, que é grande, que é alto. Mas com paciência você muda.

A fala anterior revela também o otimismo do entrevistado B, que acredita em propostas de mudança, a despeito de dificuldades e resistências por parte de alguns colegas.

Na seguinte fala do entrevistado A, fica claro como funciona, na prática, o sistema de avaliação, no que se refere à pontuação e à aprovação do aluno. Explica o entrevistado A que

Aplicamos 3 testes e 3 provas (...) apesar de ser a teoria de resposta ao item, a maneira de dar a nota ao aluno ainda é uma maneira clássica, pois conta-se o número absoluto de itens que o aluno acertou. A gente não conseguiu ainda fazer aquilo que começamos naquela reunião, que ia colocar o aluno dentro da escala. Isso ainda é uma segunda parte do projeto, né? Então tem um detalhe ainda, nós estamos no meio do caminho, as provas já são estruturadas segundo a TRI, mas a avaliação ainda é uma avaliação clássica, nós não conseguimos ainda fazer uma avaliação. (...) a estrutura da prova é o seguinte, são 10 itens ou 10 questões objetivas, que cobrem o conteúdo todo com aquelas habilidades etc, e 2 questões abertas e que o aluno tem resolver de fato, ele tem que escrever, tem que mostrar a habilidade de se expressar corretamente. Então, hoje é bem dividido. A nota é assim, é 70% da avaliação, 20% dos testes e 10% de umas atividades em sala que eles fazem. A prova é o seguinte, a prova são 10 objetivas, valendo 40% da nota, e 2 são subjetivas, valendo 60%.

O entrevista B explica que, em relação às avaliações, o estudantes têm dificuldade de compreender, inicialmente, a natureza dos itens que são aplicados:

Eu sempre atuei na aula de teoria, então o problema da aula de teoria é o seguinte, você fica muito distante do estudante, por que são 120 alunos, então eu não sei mas é assim o que eu posso dizer com respeito aos alunos: eles sempre acham muito estranho o primeiro teste ou a primeira prova. O tipo de questão, por que eu acho que eles estão acostumados com questões (...) que cobram conteúdo também né, ainda que no ENEM essa coisa toda, já tenha (...) um contexto, mas não sei o que, mas eles estão acostumados a pegar e fazer uma conta lá, então eles estranham muito, no primeiro, mas eu acho que rapidamente eles se adaptam e o que eu tenho observado é que o rendimento deles nessas questões objetivas é bem melhor do que na subjetiva.

Uma fala interessante do entrevistado A revela uma das ideias centrais da aprendizagem colaborativa:

(...) principalmente exigir do aluno que ele próprio pratique, acho até que é uma dificuldade, porque os outros fazerem é muito fácil, quando você vai fazer é que de fato aparece os problemas etc, então o magistral tem essa intenção de fazer com que o aluno pratique, né e é só quando ele pratica, que vê as dúvidas que tem — essa é que é a ideia.

Isso relaciona-se diretamente com o que Torres, Alcântara e Irara (2004) afirmam com relação à aprendizagem colaborativa:

Muitas pessoas acreditam que o conhecimento é uma entidade que se transfere de uma cabeça para outra. A aprendizagem colaborativa, no entanto, parte da ideia de que o conhecimento é resultante de um consenso entre membros de uma comunidade de conhecimento, algo que as pessoas constroem conversando, trabalhando juntas direta ou indiretamente (i.e., resolução de problemas, projetos, estudos de caso etc.) e chegando a um acordo.

Para esses autores, a aprendizagem colaborativa é “uma estratégia de ensino que encoraja a participação do estudante no processo de ensino aprendizagem e que faz da aprendizagem um processo ativo e efetivo”. Isso reforça a percepção do entrevistado A a respeito da metodologia proposta no segundo momento do Cálculo Magistral, quando se deixa para os estudantes a possibilidade de realizarem, em grupo, uma lista de exercícios cuidadosamente selecionada. Devido à peculiaridade dos conteúdos em Matemática, para que a aprendizagem seja significativa, as atividades em grupo devem ser criadas levando-se em conta o caráter investigativo da atividade, ou seja, a apresentação de atividades que contrastem com simples resoluções mecânicas de exercícios. Isso relaciona-se diretamente com a proposta de avaliação de competências que é subjacente à matriz de referência elaborada para dar suporte à elaboração dos itens para o projeto da proficiência.

Assim pontua o entrevistado A sobre a finalidade do projeto:

A gente quer resolver o problema do ensino de Cálculo, e é um meio pra isso. O fim é o aprendizado do aluno ele [projeto] é meio pra se determinar, pra se atingir esse objetivo (...) a expectativa, é que o aluno seja muito melhor formado, tenha uma estrutura diferente, seja crítico, pensar com clareza, isso a gente está buscando né? A gente espera que sim, mas acho que ainda tem um longo caminho aí. A gente ainda vai ter que trabalhar bastante para chegar lá.

Compartilhando a ideia de desenvolvimento do pensamento crítico do aluno sinalizada pelo entrevistado A, vale a pena trazer a tona a seguinte função da matriz de referência, abordada no capítulo 1, conforme explica Rabelo (2016):

Nessa concepção, os itens constituem desafios a serem enfrentados pelos estudantes e que permitem a mobilização dos conhecimentos e das habilidades em busca da solução para a situação proposta. Essa perspectiva coaduna-se com a avaliação de competências e de perfil profissional proposta por

Marinho-Araújo e Rabelo (2015), (...). Extrapolando a visão meramente tecnicista e considerando o contexto multifacetado e orgânico das práticas avaliativas no âmbito educacional, a experiência mostra que a Matriz traz organicidade ao processo e faz com que os professores elaborem cada item — ou mesmo cada opção, no caso da múltipla escolha —, com intencionalidade, um dos pressupostos fundamentais da avaliação. Além disso, os testes ficam mais equilibrados em termos de abrangência e de nível de complexidade, já que devem contemplar o máximo possível de habilidades e características de perfil elencadas na Matriz. (RABELO; MAGALHÃES; FURTADO; MARINHO-ARAÚJO, 2016)

Quando questionado se haveria apoio para realização do projeto e se este participa de algum tipo de financiamento para o seu desenvolvimento, o entrevistado A esclarece:

Recurso financeiro a gente não tem, a participação no projeto é inteiramente voluntária, contamos com uma estrutura muito boa do CESPE, na hora de ler os cartões, porque imprimir e fazer a leitura dos cartões sem isso, não teríamos como. Outra coisa boa foi o software para fazer essa análise. Aqui não tinha, então o CESPE nos autorizou a usar o software o A, o B e o C, eles nos ajudaram bastante. Aliás vale a pena mencionar que o apoio do CESPE tem sido essencial pra gente continuar.(...). Por exemplo, nos perguntaram: vocês ganham para elaborar essas questões? Ganhamos um trabalho, um final de semana inteiro disponível pra isso.

Fica evidente o apoio dado pelo Centro de Seleção e de Promoção de Eventos (CESPE), órgão que integra a Fundação Universidade de Brasília, e sua participação no projeto com a impressão e leitura das folhas de respostas. Além disso, o entrevistado B cita algumas reuniões realizadas com a equipe de especialistas do Centro, estatísticos e psicometristas que dominam a TRI:

Um suporte de Know how também. Fizemos umas 2 ou 3 reuniões lá com o pessoal da estatística deles.

3.2 Construindo os itens

Para o desenvolvendo dos itens de múltipla escolha, fez-se necessário o suporte teórico encontrado em Rabelo (2013), do qual extraímos alguns excertos e descrevemos a seguir.

O item de múltipla escolha é o tipo de questão que exige do estudante selecionar uma única opção entre as várias apresentadas. O gabarito ou chave de resposta será a única alternativa impreterivelmente correta, enquanto as outras opções ou alternativas devem ser incorretas, porém plausíveis. (RABELO, 2013). Expõe Cohen, Swerdlik e Sturman (2014, p. 251) que podemos ter outro formato de itens, onde cada alternativa terá a característica de medir um traço do indivíduo, assim cada estudante escolhe a alternativa que melhor representa a resposta sobre a si mesmo. Esse estudo compreende apenas o tipo de item encontrado em Rabelo.

Para teste padronizado, utilizando itens de múltipla escolha, é aconselhável uma produção do dobro de itens, a depender do número de itens do teste e que os itens cubram a matriz de referência ou conteúdo da avaliação. Isso dará uma ideia do teste, para, então, selecionar os itens que realmente formarão o teste. Depois desse processo é preciso verificação se os itens selecionados cobrem o domínio do teste, caso necessário é preciso refazer algum item para cobrir o domínio. (COHEN; SWERDLIK; STURMAN, 2014, p. 251)

No processo de produção de itens de múltipla escolha, costuma-se encontrar três tipos básicos: convencional, que aborda a resolução do problema de maneira rotineira, repetindo o conhecimento aprendido ou o reconhecimento de informações; textualizado, que se caracteriza por um equívoco do objetivo de criação da questão contextualizada, formados geralmente por textos que são desnecessários para resolução do problema e até criação de situações inusitadas, ficcionais, para tentar expressar uma realidade e não atingindo este objetivo; contextualizado, que possibilita a investigação de uma situação-problema, assim avaliando a competência e habilidade que o item descreve.

O item contextualizado é um motivador ao estudante para resolver a atividade proposta, enquanto itens abstratos ou técnico-científicos costumam demandar raciocínios elaborados e apresentar níveis de julgamento mais complexos. Cabe ressaltar que avaliações nesse sentido não devem induzir a abandonar treinamento de itens ou o ensino da técnica e dos procedimentos, porém ultrapassar esse aspecto para alcançar fatores externos, que possam se ligar a forma social, cultural e histórica.

Os avaliadores devem sempre estar atentos para não inculcar no item teor de preconceitos ou discriminação de gênero, origem, religião, cultura e outros. De acordo com o contexto do item, pode haver favorecimento de estudantes de determinadas culturas e mesmo gênero, por questões ainda colocadas pela sociedade como interesse do sexo masculino em detrimento a interesse do sexo feminino, favorecendo o acerto daquele grupo.

A estrutura de um item de múltipla escolha contempla três partes: texto base, enunciado e opções. Essas partes devem ser interligadas com coerência e coesão, de forma concisa, constituindo-se uma unidade de proposição, explicitando uma única

situação-problema. Os itens devem ser claros, objetivos, precisos e impessoais. Não devem trazer informações desnecessárias, devendo esclarecer conceitos e termos técnicos que não sejam objetivos de avaliação.

Em termos formais, o comando de um item poderá ser uma afirmação incompleta a ser completada por uma das alternativas. Desse modo, o estudante deve completar a frase do enunciado antes de analisar todas as opções. O comando também poderá ser uma pergunta direta, onde cada resposta, exibida em cada alternativa, é uma possível solução, mas apenas uma só é correta.

No processo de elaboração, não é indicado escolher um item pronto e tentar condicioná-lo a uma competência, mas, sim, escolher o que se pretende avaliar, elaborar o texto base, buscando referências com possíveis adaptações, sem mudar a ideia do texto e afirmando na referência se o texto é adaptado. Após a elaboração do texto base, cria-se o comando e somente depois são construídas as opções. Como orientação básica encontrada em Rabelo (2013), ressalta-se que o texto-base não deve ultrapassar 10 linhas e opções de respostas com mais de 2 linhas tornam o item complexo e extenso, uma vez que cada item é elaborado para ser resolvido em média em 3 ou 4 minutos, por um estudante de proficiência média.

Quanto à plausibilidade dos distratores, é esperado que cada opção seja uma resposta possível dentro do contexto do item para um tipo de estudante que não desenvolveu a competência avaliada ou não sabe resolver a situação proposta. Assim, uma opção plausível é aquela que apresenta um raciocínio possível de um respondente de baixo desempenho, que não domina aquilo que está sendo avaliado. É preciso tomar cuidado para não transformar um distrator em uma pegadinha, o “pega” é a opção que induz alunos de bom desempenho, que desenvolveram a competência avaliada, a responder o item por aquela alternativa que não é a resposta correta. Quando uma opção não é plausível, pouco ou nenhum aluno acaba por escolher este, podendo assim não ter necessidade de aparecer como opção no item.

Itens com enunciados negativos devem ser evitados, para não trazer confusões de escolha. Além de escolher uma opção incorreta, não mostra, necessariamente, que o estudante saiba o que é correto. Deve-se evitar o uso de termos ambíguos, confusos ou vagos. Palavras rebuscadas, informações irrelevantes ou obscuras dificultam a análise da real proficiência do indivíduo.

Os itens que objetivam avaliar o desenvolvimento de competências devem ser inéditos e não privilegiar determinados alunos que possam ter utilizado de fontes diferentes de estudos, mas principalmente que a avaliação da competência do estudante está no enfrentamento de situações inusitadas, ou seja, inéditas.

O item deve ser discriminativo, isto é, os estudantes devem acertar os itens de acordo com o nível de suas proficiências. Assim, o estudante de bom desempenho acaba sendo melhor, em cada item, que o estudante de mau desempenho. Os itens

extremamente fáceis ou extremamente difíceis pouco discriminam. Cada item deve ser independente dos demais, não sendo necessário acertar outro para realizar a tarefa deste.

As alternativas do item devem ser construídas de forma paralela. O paralelismo sintático significa, por exemplo, começar todos por verbos, ou por substantivos ou artigo. Com regra geral, não se deve deixar uma das alternativas com um chamamento de atenção pelo simples fato da sua estrutura ser diferente das demais. Nas opções numéricas, sugere-se apresentá-las em ordem crescente ou decrescente, sem discrepâncias que podem levar a conclusões sobre a resposta, mesmo antes de resolver o item.

A posição de ocorrência do gabarito deve variar no teste, para se ter um balanceamento da quantidade de vezes de cada opção correta.

Alguns termos de frequência, cuja interpretação pode variar entre os indivíduos, devem ser evitados, como: muito, pouco, bastante, às vezes. Recomenda-se também que sejam evitados, como elementos centralizadores do erro ou do acerto, os termos: somente, apenas, nunca, exclusivamente, unicamente, sempre, totalmente, todo, jamais, raramente, exclusivamente, pode, algum, nada, nenhum. Nesse caso, um estudante, mesmo sem o domínio do que está sendo avaliado, poderá descartar uma opção de resposta apenas por conter um desses termos.

Os itens devem ser avaliados e se não seguirem os critérios da denominada engenharia de construção de itens, deverão ser reformulados ou mesmo eliminados da avaliação. (RABELO, 2013). Entre as dificuldades encontradas nesse processo, menciona-se, de forma curiosa, o entrevistado A:

Quando levamos pra ele [Mauro] fazer uma revisão, ele falou que estava tudo errado. Levamos uma cacetada (SIC). Levou as questões, tinha cada questão que para resolver precisava de 30 minutos. Era pra desenvolver o domínio de habilidades e não apenas o conteúdo. Este foi o choque número 1. Porque a gente estava acostumado a fazer o contrário, que era só conteúdo e foi o que a gente mais apanhou na elaboração, mas com a orientação do Mauro, acabou sendo bem conduzido o processo, nas poucas interações que nós fizemos, a gente levou uma primeira sugestão muito diferente, foi revisto e, assim, em 3 ou 4 reuniões a gente conseguiu afinar um pouquinho e entendeu o processo. Até hoje a gente tem dúvida, qual é a habilidade que está sendo aplicada, qual não é, se classifica, se não, mas já está bem mais claro esse processo; isso o Mauro ajudou muito em relação a essa questão. Foi essencial que ele tivesse dado essa orientação pra gente.

Com o suporte da Matriz de Referência, o procedimento para a elaboração dos itens de avaliação tem a busca do perfil geral a ser alcançado e suas características como objetivo central. Assim, parte-se da escolha da habilidade a ser avaliada e do

conteúdo que dará suporte à formulação do item, que, de modo geral, consiste da resolução de uma situação problema. Nessa perspectiva, as dificuldades de desenvolvimento dos itens e as metodologias utilizadas mostram a alteração necessária ao modelo usual com que os professores do MAT estavam acostumados. A dificuldade para a elaboração ainda é grande, conforme explicita claramente o entrevistado A.

Acho que até hoje a gente não aprendeu ainda, assim com todas as letras, como construir esses itens. Ainda tem, por exemplo, na nossa reunião tinha lá a habilidade colocada e a gente leu o item e poxa, mas esse item não tem nada a ver com essa habilidade, até aquele momento a gente estava em dúvida de como elaborar essas coisas. Eu acho que vai ser difícil sempre, a gente não vai aprender a fazer esses itens, vai melhorar, mas vai acabar sendo uma coisa que a gente vai continuar aprendendo pra sempre. Acho que essa que é a ideia.

O entrevistado B pontua a dificuldade na elaboração do banco de itens e sua validação, de acordo com o trecho transcrito a seguir.

Eu acrescento uma coisa mais, mais específica do projeto, a logística do processo de elaboração, assim é penosa demais porque você não tem uma ferramenta para desenvolver as coisas, eu sei que se você tivesse um sistema que gerenciasse os itens seria mais fácil, hoje isso tudo é feito na mão, tabela, essa questão de recurso.

Assim concorda e completa o entrevistado A:

Precisa por exemplo ter um programa que gerencie a elaboração de itens, porque é muita gente elaborando, são muitos itens, tem que ser revisto, essa serve, essa não serve, isso tudo dependeria de um gerenciamento menos artesanal. E a gente não tem recurso pra isso, então acaba sendo feito manualmente mesmo.

Perguntado se as questões abertas seguiam o modelo tradicional, ou se houve alterações, esclarece o entrevistado A.

Estão focadas em conteúdo [questões abertas], mas também em habilidades, o aluno tem que expressar, tem que escrever, ele tem que dizer o que ele está fazendo, tem um pouco aí sim (...) o percentual de aprovação ficou muito próximo dos percentuais anteriores. Aliás, esse é um ponto interessante, apesar do critério ser inteiramente diferente etc, provavelmente os alunos aprovados neste critério sejam bem diferentes dos outros, o percentual em

si não mudou muito, o percentual de aprovação na metodologia antiga e o percentual de aprovação na metodologia moderna ficaram muito próximos e isso é um ponto que a gente tem que esclarecer.

Ainda sobre a experiência de construir os itens e as consequências de se seguir uma metodologia bem fundamentada, contribuindo para que a avaliação seja mais justa, o entrevistado A cita.

Eu estava muito orgulhoso de fazer questão e levei uma aula, a gente aprendeu quase que do zero, foi bom então a teoria que explica o que, como se diz, trabalha teoria com elaboração de itens objetivos, foi uma novidade porque a gente sempre imaginava que os itens objetivos, você punha um macaquinho lá pra responder os itens e se você treinar direito ele vai responder... e não é assim de jeito nenhum. Se você elaborar de maneira correta os itens, é possível fazer uma avaliação bem justa, profunda da sua situação, com itens objetivos. E principalmente essa ferramenta estatística não é o item ou a questão é você ver como isso funciona no grupo de alunos. Essa foi outra novidade também bem interessante: você colocar num contexto estatístico e tirar informações estatísticas do item, não informações pontuais. Foi muito bom, foi uma experiência bem interessante.

Percebe-se como é visto de forma positiva o suporte teórico e a metodologia de elaboração e análise dos itens. O entrevistado B discorre sobre o argumento científico que dá sustentação ao projeto e facilita a adesão e construção de opiniões favoráveis.

Esse projeto é diferente, você tem uma teoria por trás, não é só achismo, porque eu acho que em discussões coletivas quando você fica só falando o que você acha. (...)tem um monte de pesquisa falando, olha, eu falo especificamente de você poder ou não avaliar por questões objetivas. Na literatura, você vê que é possível. Dá trabalho? Dá! A discussão fica mais interessante.

3.3 O primeiro teste

Uma das recomendações dos teóricos de avaliação educacional, de acordo com Pasquali (1999), é da distribuição de níveis de dificuldade de itens no teste dentro de uma curva normal. Nessa ótica, os itens classificados como muito fácil deverão representar 10% do total; fácies, 20%; medianos, 40%; difíceis, 20%; e muito fácil, 10%.

A Tabela 3.1, construída e avaliada por Rabelo, Magalhães, Furtado e Marinho-Araújo (2016), mostra que a distribuição encontrada na primeira avaliação de Cálculo I aplicada no primeiro semestre letivo de 2015, não difere muito dessa recomendação, indicando que, pelo menos no que se refere à dificuldade, o conjunto de itens mostrou-se adequado.

Tabela 3.1: Distribuição dos itens segundo o parâmetro de dificuldade b

Classificação	Valore de b	Itens na faixa	%
Muito fáceis	até -1,28	2, 11, 19, 20	15,4
Fáceis	de -1,27 a -0,52	3, 4, 8, 12, 14, 22	23,1
Medianos	de -0,51 a 0,51	7, 9, 13, 16, 18, 26	23,1
Difíceis	de 0,52 a 1,27	1, 5, 6, 10, 15, 17, 21, 23, 25	34,6
Muito difíceis	1,28 ou mais	24	3,8

Como reflexo da observação e estudos realizados com base na TRI e as conclusões sobre a avaliação, segue o comentário do entrevistado A sobre o processo da aplicação da primeira avaliação centrada no desenvolvimento de habilidades e competências:

A primeira prova inclusive a gente estava morrendo de medo, fizemos uma prova considerada fácil pelos alunos, bem tranquila assim, meio... estava fácil demais, né? Aliás, o Mauro achou aquela prova a melhor de todas, (estava mais equilibrada, mais próxima ao que os alunos podiam responder de maneira razoável).

No primeiro semestre de 2015, o projeto foi desenvolvido por um grupo de 10 professores e envolveu 1173 estudantes matriculados em 24 turmas de Cálculo 1, contemplando estudantes dos diferentes cursos presenciais de graduação da instituição, que incluem: engenharias, matemática, física, química, biologia, geociências, geofísica, administração, economia, ciências contábeis, estatística e computação.

Utilizando os parâmetros oriundos da TCT e da TRI, foram analisados os resultados de desempenho, apresentando os parâmetros de discriminação (a), dificuldade (b) e de acerto ao acaso (c), fornecidos pela TRI, e o coeficiente bisserial (Bis) e a dificuldade (D) dados pela teoria clássica.

Para cada item, foram também geradas a Curva Característica do Item (CCI) e a Análise Gráfica do Item (AGI). Conforme comentam Rabelo, Magalhães, Furtado e Marinho-Araújo (2016), esses gráficos e os parâmetros gerados pela TRI e TCT ajudam a investigar se os itens cumpriram suas finalidades. Os autores afirmam que:

Via de regra, elabora-se um item pressupondo-se que ele vá cumprir determinada finalidade, mas, somente após análise dos parâmetros, descobrem-se desvios nas alternativas, problemas nos enunciados, falta de capacidade de discriminação, dificuldade superior à esperada, entre outros.

Com essa metodologia, são fortalecidos os processos de ação-reflexão-ação no espaço educativo, possibilitando ao professor a transformação não apenas de seu fazer docente mas também de si, em um exercício permanente para a melhoria da qualidade das atividades desenvolvidas durante a ação pedagógica.

A tabela 3.2 ilustra alguns dados encontrados na primeira avaliação, aplicada em abril de 2015 para 1173 estudantes de Cálculo I.

Tabela 3.2: Parâmetros dos itens da primeira avaliação segundo a TRI e a TCT

Questão	TRI			TCT	
	a	b	c	Bis	D
01	1.475	0.562	0.245	0.694	0.497
02	0.719	-1.816	0.212	0.585	0.886
03	0.714	-0.897	0.189	0.616	0.760
04	0.904	-0.897	0.182	0.691	0.776
05	1.697	1.018	0.123	0.740	0.291
06	0.629	0.925	0.197	0.571	0.450
07	1.148	-0.183	0.164	0.765	0.634
08	0.869	-0.757	0.194	0.663	0.758
09	0.955	0.170	0.240	0.667	0.593
10	2.076	0.970	0.109	0.760	0.280
11	0.901	-1.313	0.209	0.717	0.851
12	0.819	-1.191	0.225	0.670	0.829
13	0.865	-0.375	0.176	0.720	0.669
14	0.611	-1.254	0.224	0.518	0.806
15	1.629	0.671	0.143	0.750	0.413
16	0.516	0.428	0.241	0.526	0.564
17	1.619	0.885	0.081	0.787	0.287
18	0.869	-0.419	0.193	0.700	0.686
19	1.150	-1.439	0.207	0.728	0.883
20	0.692	-1.994	0.207	0.615	0.894
21	0.884	0.549	0.224	0.647	0.494
22	0.525	-0.686	0.223	0.540	0.711
23	1.235	0.914	0.128	0.696	0.325
24	0.827	1.869	0.182	0.493	0.275
25	0.764	0.621	0.174	0.666	0.455
26	1.140	0.422	0.209	0.674	0.497
Média	1.009	-0.128	0.1885	0.664	0.599

Os estudos apresentados em Rabelo, Magalhães, Furtado e Marinho-Araújo (2016) indicam que esse conjunto de itens teve um bom comportamento do ponto de vista psicométrico, com parâmetros de discriminação e dificuldade médias respectivamente iguais a 1,009 e -0,128, segundo a TRI, e 0,664 e 0,599, segundo a TCT.

Como o parâmetro de discriminação da TRI, apresenta-se abaixo de 0,65 em apenas 4 itens, esses itens devem receber atenção no sentido de verificar sua adequabilidade. Segundo o parâmetro de dificuldade, apenas um item estava na escala de muito difícil ($b \geq 1,28$).

Dispostos na tabela 3.3, publicada por Rabelo, Magalhães, Furtado e Marinho-Araújo (2016), estão os índices de aprovação das turmas dos cursos diurnos em 2014 e no primeiro de 2015. Nela, AP: aprovados, RP: reprovados e TR: trancamentos feitos pelos estudantes.

Tabela 3.3: Percentuais de aprovação nas turmas dos cursos diurnos: semestres 1 e 2/2014 e 2/2015

Turma	Semestre											
	2015/1				2014/2				2014/1			
	AP	RP	TR	%AP	AP	RP	TR	%AP	AP	RP	TR	%AP
A	51	9	2	82,26	49	7	3	83,05	49	9	0	84,48
B	18	40	1	30,51	23	36	0	38,98	31	36	1	45,59
C	40	14	7	65,57	22	30	8	36,67	36	25	4	55,38
D	32	26	4	51,61	33	28	3	51,56	21	44	4	30,43
E	25	13	1	64,10	18	13	0	58,06	18	28	3	36,73
F	32	13	1	69,57	39	12	0	76,47	44	16	0	73,33
H	31	28	1	51,67	14	27	2	32,56	43	17	1	70,49
J	41	16	4	67,21	40	11	0	78,43	38	20	1	64,41
M	38	23	0	62,30	21	21	0	50,00	38	26	2	57,58
N	26	19	0	57,78	11	22	5	28,95	36	16	2	66,67
O	23	29	3	41,82	32	31	0	50,79	18	26	15	30,51
Y	38	15	3	67,86	30	34	1	46,15	28	30	7	43,08
Z	35	15	2	67,31	29	23	9	47,54	40	25	0	61,54
TOTAL	430	260	29	59,81	361	295	31	52,55	440	318	40	55,14

Quando questionados sobre uma possível avaliação do projeto, os professores argumentam que ainda não existem conclusões, com metodologias científicas, mas por observação concluem que houve uma homogeneidade no número de aprovações, abordada pelo entrevistado B.

Assim, esse semestre uma turma, é separada por curso, então por exemplo, biologia, tinha 1 semestre que a biologia tinha, sei lá, 30% de aprovação no outro tinha 60%. O que aconteceu agora com a unificação, é que isso está equilibrado, então a gente já sabe mais ou menos qual vai ser a aprovação

de cada turma, não tem variações, mas quando você pega a média, como tem muitas turmas, ficou igual.

Já os autores citados anteriormente, fazem uma análise bem interessante e favorável ao projeto.

Verifica-se uma melhora em termos de desempenho médio dos estudantes do ano de 2014 para 2015. Naturalmente, ainda é cedo para se fazer inferências acerca desse resultado, mas isso já representa um alento, principalmente se for observado que o quantitativo de estudantes ingressantes que cursaram integralmente o ensino médio em escolas públicas aumentou em cerca de 12,5% de 2014 para 2015, a partir da denominada Lei das Cotas Sociais. Esses estudantes ingressam com escores inferiores nos exames de acesso e, em tese, teriam mais dificuldades de acompanhar as disciplinas do primeiro ano dos cursos (RABELO; MAGALHÃES; FURTADO; MARINHO-ARAÚJO, 2016).

3.4 As primeiras avaliações dos projetos

Durante a entrevista, foi interrogado como estavam sendo avaliados os projetos, como os alunos se comportavam e também como estava o contexto do departamento.

Eu acho bom pelo seguinte, porque eu tenho um desânimo muito grande de aula teórica, eu acho que você fica ali falando, falando e no fundo ninguém está te ouvindo, então quanto mais você colocar o aluno pra trabalhar é melhor. Talvez existam outras maneiras de se fazer, que acho que não precisava ser magistral, você podia sei lá dar 1h de aula e depois colocar o cara pra fazer exercícios, que é uma maneira também, qualquer coisa que você fizer pra fazer com que o aluno seja ativo, que entenda que ele faz parte de um processo muito mais do que o professor. É bom, então, eu avalio positivamente, sou um dos entusiastas do negócio. Agora assim, eu não tenho nenhum elemento, por exemplo, se falar pega a aprovação, não sei o que, eu não tenho nenhum elemento científico, mas nós já vamos para o ano 3, de fato assim, ele vigorou por 2 anos e voltou pra discussão e quando voltou a decisão, foi que ficasse outros 3, exatamente por conta disso. Então, depois daqui 2 anos e meio, acho, quando completar 5 anos, vai haver uma nova avaliação, mas, por exemplo, o que eu posso dizer é que a avaliação que foi feita depois de 2 anos foi uma avaliação passional assim, quem não gostava, é não teve nenhum

O entrevistado B expõe sua visão sobre o Cálculo Magistral, e, segundo ele, o projeto estaria ajudando para aumentar o interesse dos alunos. No entanto, relata a dificuldade de avaliar e comparar os projetos.

Os alunos gostam, eu acho, na média, os que eu converso gostam mais do que o sistema antigo. E, principalmente no começo, quando se implantou, eu acompanhava mais, eu tinha minha turma de teoria, eu ia à aula de exercícios pra ficar perguntando, já que tinha aluno repetente. Não sei se é uma comparação válida, o cara fez dos dois jeitos, eu acho que eles gostam mais dessa maneira. Eles reclamam um pouco, falam que é pouca teoria, muitos acham que deveria seguir aí duas horas de exercício, porque acham que a aula de teoria é corrida. Não sei o que, mas, assim, eu acho que em geral eles gostam. Eles não se opõem pelo menos.

Desse modo, foi exposta a afinidade do projeto Magistral com os alunos, melhorando a aprendizagem em Cálculo. Será relacionado à importância do estudo em matemática em um laboratório social, onde a troca de explicações é uma boa maneira de autoavaliação, o aprendizado coletivo impulsiona o julgamento pessoal de seu aprendizado, assim como também leva a avaliação pelos pares, itens que enriquecem tanto na diversificação dos modelos de avaliação, como na qualidade de avaliação, como citado na sequência pelo entrevistado A, que no caso da matemática está relacionado com a construção lógica do argumento matemático.

Ainda salienta que a avaliação significativa do desenvolvimento dos projetos de Cálculo, será de grande valia o exame de proficiência, que tornará possível a comparação entre alunos de anos diferentes a construção de suas proficiências, isso com o suporte da escala de proficiência que está sendo construída pelos professores do MAT. Assim, prossegue o entrevistado A:

Os dois momentos são essenciais, eu acho que é preciso ter momento individual, em que o aluno faz uma reflexão, tenta fazer, mas é essencial o outro momento, coletivo. Aí tem uma brincadeira que eu sempre falo com os meus alunos, o laboratório de matemática é a conversa, em química você tem lá a substância, você experimenta, faz umas fumacinhas, mas com a matemática não tem jeito, a única maneira de você testar se entendeu ou não alguma coisa é conversando, é essencial esse momento coletivo. Se você conseguir explicar ao outro um exercício, é porque você entendeu o exercício, mas são mentes diferentes. Quer dizer, você tem que explicar ao outro uma coisa que você entendeu; isso é essencial. Isso é onde peca muito ao método tradicional, que é muito teórico e muito individual. O aluno vai pra casa e estuda e depois ele vai fazer a prova sem um contato com

os outros alunos. Então acho que os dois processos são importantes, um complementa bem o outro. Anteriormente, você focava muito no individual e agora você está dando um pouco mais de atenção ao coletivo. Você está num grupo discutindo, mostra suas razões, porque que você está certo. É assim, em matemática eu acho que é o ponto principal nessa história. Então a experiência coletiva é assim, eu acredito numa das coisas mais, que a gente conseguiu ganhar mais nesse processo do magistral, essa troca de experiências entre os alunos (...). Uma maneira rasa de avaliar [a melhora no desempenho] seria olhar para o índice de aprovação, por exemplo, é muito complicado porque a régua mudou. O metro que você mede as coisas, antes era um e agora o metro é outro, então a gente não tem uma escala linear de um metro pro outro, então é muito complicado de fazer essa avaliação. O que a gente pode entender, ou que a gente é só uma opinião mesmo. Olha a gente acha que está melhor, não tem ainda uma ferramenta para isso. Talvez, mais no futuro, em que a gente tenha uma régua comum, tenha uma unidade de medida, por exemplo, o desempenho da proficiência, você consiga dizer de um semestre para o outro, se as coisas melhoraram ou não, porque a régua será a mesma, então se você tiver a proficiência implantada durante anos e isso está bem estabelecido, você pode comparar alunos posteriores com alunos anteriores e aí dizer, se melhorou ou não etc. Mas isso depende de instrumentos muito sofisticados. Atualmente a gente não tem ferramenta pra dizer isso claramente.

O entrevistado B, no trecho transcrito a seguir, lembra bem que o processo de avaliação através da vivência e da crença dos docentes aponta para o sucesso do projeto. Ele ressalta que a frequência do modelo magistral é mais alta, os alunos se preocupam mais em assistir a aula teórica, já que é a única da semana. O professor entrevistado aproveita para comparar o modelo tradicional com o magistral, no que diz respeito à possibilidade que esse último oferece de acompanhamento do aluno pelo professor, enquanto naquele o estudante acaba estudando muito sozinho e, muitas vezes, nem sabe como fazê-lo.

Tem maneiras de fazer medições, aferições parciais. No diurno, os professores das aulas práticas têm me dito que os alunos estão interessados pra caramba. Assim, muitas vezes, o cara vai até o final da aula. Numa aula teórica, especialmente quando são 3, o método tradicional; das 11:30 e não tem ninguém mais, começa a debandada. E eu acho que eles se dedicam mais, mas também por conta do cara tem um cronograma do que fazer. Você meio que pega na mão do cara, acho que precisa fazer isso, acho que quando você deixa o aluno solto, ele acaba se perdendo, não sabe direito

o que tem que fazer, nem como estudar, pq o cara abre o livro em casa, cada seção do livro tem 140 exercícios. Mesmo na aula teórica, hoje com o magistral, a frequência é altíssima, ele praticamente não perde. Só tem 1, ele sabe que se ele perder aquela ali, se(...)

Sobre a presença dos alunos nas aulas e os motivos que levaram à introdução de tarefas obrigatórias nas aulas de exercícios, o entrevistado B ainda acrescenta que:

O seguinte, quando o magistral começou, a aula de exercícios não tinha nenhuma atividade que pontuava para o aluno, ele fazia a prova depois, então o que a gente começou a perceber é que eles começavam a faltar muito nas aulas, principalmente as turmas mais fracas, porque o cara ia mal na primeira prova, aí já se arrebatava na segunda, já meio que desistia e aí foi observando isso que a gente introduziu aquele negócio de uma atividade e pontuava. Então ele resolvia uma lista e entregava, depois que ficou assim, a assiduidade ficou muito alta, muito alta, pouquíssimos faltam, na aula teórica ninguém falta, ninguém falta na aula teórica.

Ao ser solicitado para comentar sobre as mudanças no Departamento de Matemática, ocorridas nos últimos anos, o entrevistado A relata um pouco o processo histórico e sua visão dos rumos do MAT, apontando as consequências do projeto REUNI, um programa de reestruturação das universidades federais, ocorrido há cerca de 10 anos e que trouxe para o MAT um quantitativo expressivo de novos docentes. Esses docentes aceitam bem iniciativas de mudanças em metodologias pedagógicas.

O departamento passou muito tempo sem contratar professores novos. Eu acho que, de 1985 até 2000, não teve praticamente contratação alguma, foi assim, no período Fernando Henrique, enfim, teve uma coisa muito... O departamento era muito engessado por causa disso, poucas iniciativas, poucas mudanças, Com o REUNI, houve uma mudança sim, na contratação de novos professores, o que mudou o perfil do departamento e mudou também o perfil dos alunos. Então foi uma mudança, uma mudança drástica. Bem marcada assim, hoje o departamento é muito mais aberto a essas iniciativas, ele está mais envolvido em modernizar o processo, está mais interessado em ter soluções inovadoras. Muito mais do que antes desse processo, isso é verdade.

Complementa o entrevistado B que, de acordo com sua experiência, o departamento de matemática da UnB é protagonista nesses projetos que envolvem o Cálculo, já que a maioria das IES não está desenvolvendo nada parecido. O entrevistado revela também sua percepção sobre o baixo desempenho dos estudantes ingressantes, um pouco conformado com o fato de isso ser um processo histórico.

Concordo e acho que principalmente por causa da oxigenação daquelas pessoas novas e assim, tem uma vantagem e uma desvantagem, mas em matemática o cara é novo é jovem você nunca está com um professor novo que não seja jovem, então eu acho, isso é fundamental, então eu não sei se isso é uma coisa importante que ajudou, mas eu não sei até que ponto é determinante. Eu estou falando isso porque viajo muito por conta de outras coisas então, sinceramente, eu vejo o nosso departamento aqui protagonista desse tipo de coisa de inovação (...) os outros não estão fazendo, não sei o porquê. Eu acho que não só aqui tem projetos de Cálculo, no Pará, eles têm um negócio grande lá, seguindo outras ideias, diferente do que a gente faz, mas, de uma maneira geral, eu acho que as pessoas estão paradas... [Quanto aos alunos] eu não vejo diferença. não, eu não consigo perceber uma diferença grande não, eu acho que eles continuam entrando sem saber nada, igual antigamente não sabiam nada.

Comentando sobre os novos processos de seleção que oportunizam a entrada na Universidade de uma parcela da sociedade que até então não tivera acesso, salienta o entrevistado B que esses estudantes pouco ou nada sabem, referindo-se aos pré-requisitos em termos de conhecimentos matemáticos de parte significativa dos alunos que chegam à universidade, conforme foi explorado na introdução deste trabalho. Já o entrevistado A possui uma visão ligeiramente diferente, mais relacionada com a atitude passiva dos estudantes de hoje, conforme pode ser inferido do comentário a seguir.

Eu noto uma diferença grande, por exemplo, nos meus alunos. Anteriormente vinham fazer revisão e brigavam comigo, eu tinha que ser duro. Os alunos [hoje] nem vem fazer revisão, são passivos, são alunos que aceitam mais, não é que aceitam com mais facilidade, é com menos iniciativa, eles aceitam o que você diz, você dá uma nota zero pra toda a turma, ninguém vai falar nada, isso é uma diferença brutal, uma coisa que eu senti assim muito. (...). Hoje os alunos são muito passivos, mas assim demais, qualquer coisa que você faça está feito, eles não questionam, não tem assim uma, uma ideia do papel de aluno que eles deviam ter.. Mudou bastante. Não sei, não sei por que motivo, mas isso eu sinto bastante. Da qualidade mesmo, da bagagem, mudou muito.

Essa visão dos professores entrevistados sobre os estudantes ingressantes revela a necessidade de implantação de políticas de acolhimento dos discentes, em especial no primeiro ano da graduação, de modo a ajuda-los nessa passagem da educação básica para a superior, momento em que eles têm necessidade de desenvolver a autonomia, para enfrentarem os desafios de uma formação para a profissionalização, desenvolvendo competências que são imprescindíveis para sua atuação futura.

Capítulo 4

Análise dos Itens aplicados no primeiro semestre de 2015

No primeiro semestre de 2015, foram realizadas três avaliações nas turmas de Cálculo I do MAT. A Tabela 4.1 a seguir apresenta os parâmetros da segunda avaliação, contemplando o índice de discriminação (a), a dificuldade (b) e o acerto ao acaso (c) e, segundo a TRI, e o coeficiente bisserial (Bis) e a dificuldade (D), de acordo com a TCT.

Comparando os dados médios obtidos para os parâmetros da primeira avaliação com os da segunda avaliação, podemos verificar que o índice de discriminação segundo a TRI baixou de 1,009 para 0,853, indicando que a primeira cumpriu melhor a função de discriminar no continuum das proficiências que a segunda. Já a dificuldade subiu de -0,128 para 0,391, revelando que a primeira foi, em termos médios, mais fácil que a segunda. Naturalmente, em termos de dificuldade, o dado mais importante será obtido na Tabela 4.3, pois se refere à distribuição dos índices de dificuldade ao longo dos cadernos de prova. O resultado do maior grau de dificuldade da segunda está diretamente relacionado ao fato da ausência de itens muito fáceis (Tabela 4.3).

Observando-se também os dados da TCT, verifica-se que a discriminação média caiu de 0,664 para 0,599 e o parâmetro de dificuldade média baixou de 0,599 para 0,533, sinalizando aumento de dificuldade média do teste. Cabe, então, lembrar que o parâmetro de dificuldade da TCT é dado pela porcentagem dos respondentes que acertaram o item em relação ao número absoluto de respondentes, ou seja, o teste teve um aumento de dificuldade. Já na TRI, a dificuldade refere-se à proficiência mínima necessária para que se tenha uma probabilidade mediana de acerto.

Para melhor compreender os resultados, vamos olhar os itens em relação ao teste, comparado com a recomendação de Pasquali (1999) para TRI, sobre a distribuição do nível de dificuldade no teste. O conjunto de itens da segunda avaliação de Cálculo I mostrou-se com o dobro de itens na classificação muito difíceis e nenhum item

Questão	TRI			TCT	
	a	b	c	Bis	D
01	1,274	0,176	0,154	0,767	0,526
02	0,829	-0,998	0,203	0,617	0,791
03	1,055	0,456	0,160	0,686	0,468
04	1,055	0,456	0,160	0,733	0,601
05	0,424	2,149	0,213	0,431	0,367
06	0,359	0,421	0,238	0,402	0,575
07	1,213	-0,361	0,188	0,742	0,679
08	0,415	1,962	0,220	0,402	0,393
09	1,066	1,473	0,121	0,589	0,240
10	1,015	-0,070	0,182	0,670	0,604
11	0,955	-1,120	0,214	0,668	0,832
12	0,584	-0,934	0,217	0,579	0,758
13	0,685	0,124	0,214	0,586	0,590
14	0,637	1,055	0,212	0,591	0,439
15	1,164	1,568	0,135	0,553	0,239
16	0,381	0,754	0,219	0,433	0,527
17	1,127	0,210	0,148	0,715	0,524
18	0,92	-1,134	0,193	0,688	0,826
19	0,839	-1,246	0,193	0,641	0,831
20	0,591	-1,107	0,213	0,514	0,774
21	1,636	1,355	0,093	0,715	0,201
22	0,771	0,552	0,236	0,585	0,520
23	0,906	0,686	0,173	0,671	0,436
24	0,69	0,394	0,220	0,581	0,542
25	0,835	0,703	0,228	0,585	0,480
26	0,762	2,642	0,111	0,442	0,157
Média	0,853	0,391	0,187	0,599	0,535

Tabela 4.1: Parâmetros da segunda avaliação do primeiro semestre de 2015

na classificação muito fácil em comparação com a distribuição dos itens da primeira avaliação. Para que o teste fosse melhor equilibrado, em uma futura aplicação, sugere-se trocar 3 itens com classificação muito difícil por 3 itens pré-testados que estejam na faixa de classificação muito fáceis. Isso poderá ajudar a melhorar a interpretação futura da escala de proficiência a ser construída para dar feedback aos estudantes de Cálculo I.

Conforme comenta Rabelo (2013), alguns autores consideram como discriminativos os itens que apresentam índices de discriminação superiores a 0,70. Assim, no teste em análise, 65,4% dos itens estão classificados como proporcionando uma boa diferenciação entre os estudantes de proficiências diferentes na resolução de cada item, ou seja, 7 dos 26 itens desse teste precisariam ser melhorados do ponto de vista da

Tabela 4.2: Distribuição dos itens segundo o parâmetro de dificuldade b

Classificação	Valore de b	Itens na faixa	%
Muito fácies	até -1,28		0
Fácies	de -1,27 a -0,52	2, 11, 12, 18, 19, 20	23,1
Medianos	de -0,51 a 0,51	1, 3, 4, 6, 7, 10, 13, 17, 24	34,6
Dífceis	de 0,52 a 1,27	14, 16, 22, 23, 25	19,2
Muito dífceis	1,28 ou mais	5, 8, 9, 15, 21, 26	23,1

discriminação. Fazendo-se uma classificação por faixa de valores, Rabelo (2013) sugere a seguinte distribuição, apresentada na tabela 4.4.

Tabela 4.3: Faixa de distribuição para o Parâmetro de discriminação

Valores	Discriminação
$a = 0$	Nenhuma
$0 < a \leq 0,35$	Muito baixa
$0,35 < a \leq 0,65$	Baixa
$0,65 < a \leq 1,35$	Moderada
$1,35 < a \leq 1,70$	Alta
$a > 1,70$	Muito Alta

Podemos, assim, verificar com o auxílio da tabela 4.4, que nenhum item esteve na faixa de discriminação muito baixa, 7 itens na faixa baixa discriminação e 18 na faixa moderada. Nas faixas alta e muito alta, encontramos apenas 1 item.

Tabela 4.4: Distribuição dos itens segundo o parâmetro de dificuldade b

Classificação	Valore de a	Itens na faixa	%
Muito baixa	a partir de 0 a 0,35		0
Baixa	a partir de 0,35 a 0,65	5, 6, 8, 12, 14, 16, 20	26,9
Moderada	a partir de 0,65 a 1,35	1, 2, 3, 4, 7, 9, 10, 11, 13, 15, 17, 18, 19, 22, 23, 24, 25, 26	69,2
Alta	a partir de 1,35 a 1,70		0
Muito alta	a partir de 1,70	21	3,8

Podemos também utilizar o suporte da TCT para classificar os itens de acordo com o coeficiente bisserial, que representa a discriminação na teoria clássica, e verificamos que não houve nenhum item abaixo de 0,30, que são considerados de baixa discriminação, segundo recomendação encontrada em Rabelo (2013).

Foram geradas, para cada item, a Curva Característica do Item (CCI) e a Análise Gráfica do Item (AGI), interpretações geométricas relacionados com a TRI

e a TCT, respectivamente. A CCI apresenta a probabilidade de acerto em função da proficiência do respondente no teste como um todo, enquanto AGI ilustra os percentuais de respostas em cada opção de resposta de item específico para cada faixa de escore bruto no teste.

Através da análise desses parâmetros, podemos verificar se os itens atingiram os objetivos propostos, ou seja, se cumpriram suas finalidades. Quando esses objetivos não são atingidos, deve-se verificar se houve problemas em sua elaboração, como distorções no enunciado, contexto desfavorável para os avaliados, desvios de construção das alternativas, dificuldade superior à esperada, baixa discriminação do item, entre outros. Assim, pode-se fazer a correção desses itens para um futuro teste de validação, ou mesmo recomendar o seu descarte.

Vamos então apresentar e discutir alguns itens, com um intuito de compreender melhor como os parâmetros nos orientam. Na figura 4.1, vamos apresentar o item 6 do 2º teste de 1/2015. A escolha deste item deve-se ao seu menor parâmetro de discriminação, em uma tentativa de compreender o porquê desse ocorrido. Na sequência, será comentado o item 16, que é a segunda menor discriminação, também inferior ao valor recomendado de 0,7. Apresentaremos também o item 26, que apresenta, segundo o modelo TCT, a discriminação com valor positivo em uma das opções incorretas, o que não é permitido, pois sinaliza ser atrativa para alunos de bom desempenho.

Exemplo 1:

T5 - TVI

H2 - Interpretar diferentes representações de um mesmo conceito, transitando por representações simbólicas, gráficas e numéricas, entre outras.

Nível de dificuldade: *Mediana*

Se o gráfico de $f : [1, 5] \rightarrow \mathbb{R}$ é como abaixo, então a quantidade de pontos de mínimo local de f é

- (a) 0.
- (b) 1.
- 2.
- (d) 3.
- (e) 4.

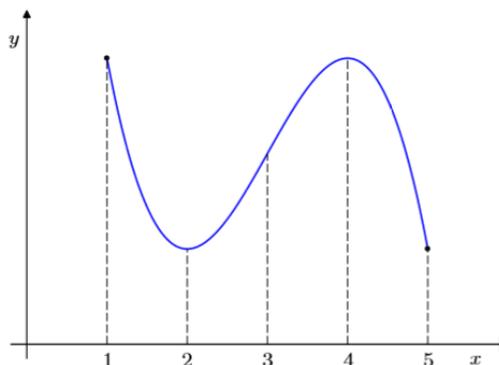


Figura 4.1: Item 6 - P2 - 1/2015.

Com base na TCT, apenas 6,6% dos respondentes escolheram as alternativas ‘a’, ‘c’ ou ‘d’, resultado da soma das proporções de respondentes destas, sugerindo que essas alternativas carecem de plausibilidade. Além disso, mais da metade dos respondentes acertaram o item, 57,5%, enquanto 35,8% optaram pela alternativa ‘b’. Esses dados estão apresentados na figura 4.2 a seguir.

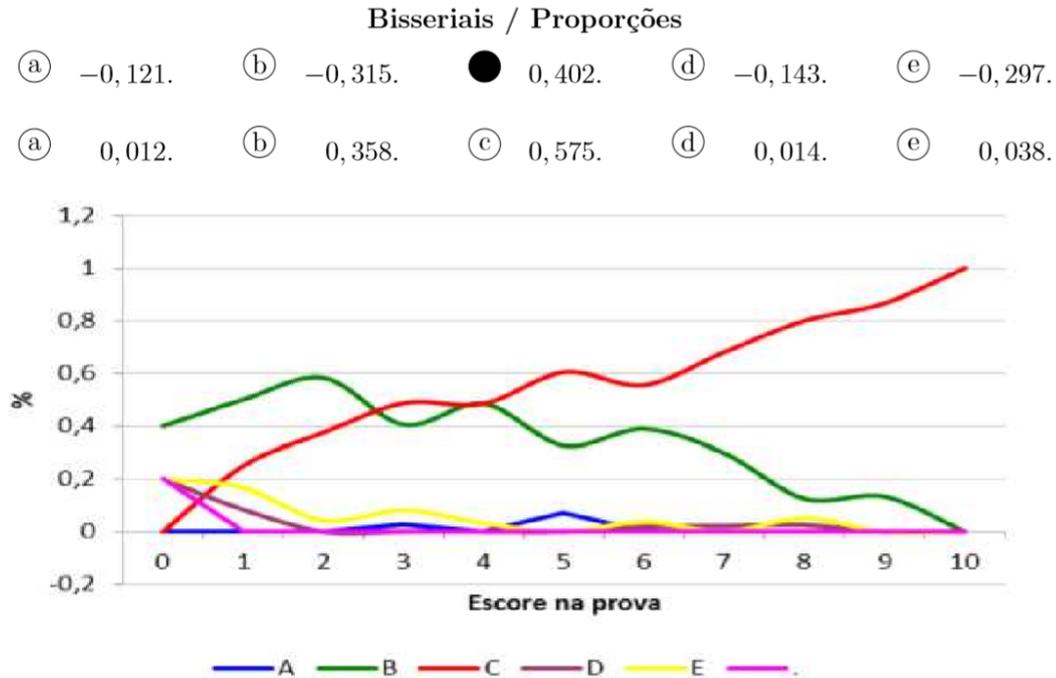


Figura 4.2: Análise Gráfica do Item 6 - P2 1/2015

Na AGI apresentada na figura 4.2, verificamos que mesmo alunos que tiveram um alto escore na avaliação continuaram a escolher a alternativa ‘b’, sinalizando para os docentes um feedback correspondente à análise do erro cometido, especificamente no que se refere ao conceito de mínimo local. O resultado do bisserial dessa alternativa é de -0,315, ou seja, ela não foi atrativa para os estudantes de maior desempenho. O bisserial de 0,402 correspondente à resposta correta sinaliza que a questão apresentou um razoável poder de discriminação, o que será contrastado com a discriminação dada pela TRI.

A escolha da alternativa B, de maneira equivocada, pode ter relação com a construção de que no extremo do intervalo fechado existe o ponto de mínimo local, ou, então, que o estudante não tenha, na matemática básica, apreendido o conceito de intervalo aberto e fechado.

Com base na CCI, figura 4.3, vemos que o acerto ao acaso está em 0,238, o que colabora com a hipótese de haver alternativas não plausíveis, pois o aluno terá mais de 20% de chance de acerto ao acaso em um item de 5 opções. No entanto, esse valor não chega a ser preocupante, sendo ainda considerado baixo e dentro dos padrões

aceitáveis.

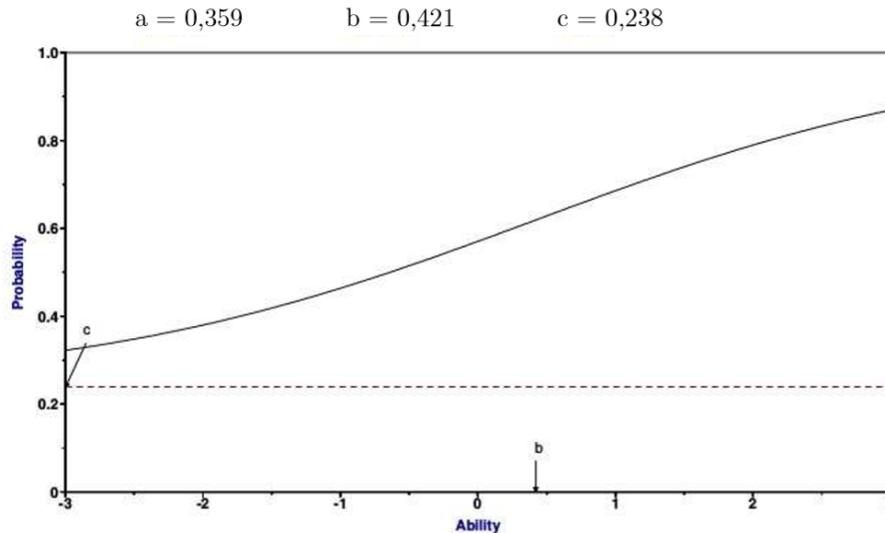


Figura 4.3: Curva Característica do Item 6 - P2 - 1/2015

Também podemos ver que o formato da curva sugere problemas na elaboração do item, pois a curva não se apresenta como uma sigmoide perfeita, aproximando-se mais de uma reta. Esse comportamento é característico de um índice baixo de discriminação, representado pelo valor de $a = 0,359$. Isso indica que o item não cumpriu com o propósito de avaliar, devido a indivíduos com proficiências muito diferentes terem probabilidades próximas de acertar o item. Quanto mais baixo for valor do parâmetro de discriminação, e quanto mais aproxima-se do zero, mais parecido com um segmento horizontal será a CCI (Rabelo 2013). Assim, a avaliação geral do item sugere que ele seja reelaborado.

Exemplo 2

Apresenta-se, na sequência, a figura 4.4 referente ao item 16 do 2º teste do 1/2015.

T1 - Limites ordinários

H4 - Inter-relacionar objetos de conhecimento em diferentes áreas.

Nível de dificuldade: *Difícil*

Suponha que a posição de um móvel, no instante $t > 0$, seja dada por $s(t) = f(t^3 - 3t)$. Se a função f tem derivada positiva, em quantos instantes o carro tem velocidade nula?

- (a) 0. ● 1. (c) 2. (d) 3. (e) 4.

Figura 4.4: Item 16 - P2 - 1/2015.

Segundo a AGI, figura 5, a alternativa ‘e’ quase não foi escolhida como resposta pelos estudantes, menos de 1%, podendo ser descartada e substituída por outra que tenha mais plausibilidade. Pelo gráfico da AGI, verifica-se que o comportamento das respostas sinaliza que, para os respondentes, a alternativa ‘e’ nem existia, ou seja, a questão tinha apenas 4 alternativas. As demais alternativas incorretas ‘a’, ‘c’ e ‘d’ tiveram escolhas respectivamente iguais a 10,8%, 24,5% e 10,8%, sugerindo que todas têm plausibilidade. A alternativa correta foi escolhida por 52,7% dos respondentes, sinalizando tratar-se de uma questão de média complexidade. Vemos, no gráfico da AGI, que os estudantes com escore no teste acima de 5, escolheram preferencialmente o gabarito ‘b’, porém uma proporção considerável de estudantes com escore maior ou igual a 5 também escolheram como resposta a alternativa ‘c’, não chegando a ser colocada como um atrativo para os alunos de bom desempenho porque seu coeficiente o bisserial da alternativa que foi -0,158.

Como observado na figura 4.5, estudantes com menor desempenho no teste escolheram com maior frequência o distrator ‘d’, seguido por ‘a’ e ‘c’. Poderíamos supor que o estudante que escolheu o distrator ‘d’ igualou a função $(x^3 - 3x)$ a zero, porém sem se dar conta que existe uma composição de função, ou mesmo, procedeu de forma equivocada, desconhecendo a regra da cadeia. A escolha do distrator ‘a’ pode ocorrer quando o estudante entende de maneira errônea que, pela hipótese de a derivada de f ser positiva, isso significa que não possui ponto crítico, não se dando conta da necessidade de utilizar a regra da cadeia. A escolha do distrator ‘c’ pode estar relacionada com o fato de o aluno compreender o problema mas não se atentar nas definições do intervalo da função. Essas são apenas algumas inferências possíveis acerca das escolhas. Se as questões fossem dissertativas, poderíamos compreender melhor as tentativas de resoluções que os estudantes fariam.

Podemos verificar pela CCI, figura 4.6, assim como ocorreu na CCI do exemplo 1, que a discriminação do item não foi satisfatória, pois a curva não se aproxima de uma sigmoide, curva em formato de “S”, mostrando a necessidade de revisão do item. Naturalmente, isso é comprovado pelo baixo valor de $a = 0,381$.

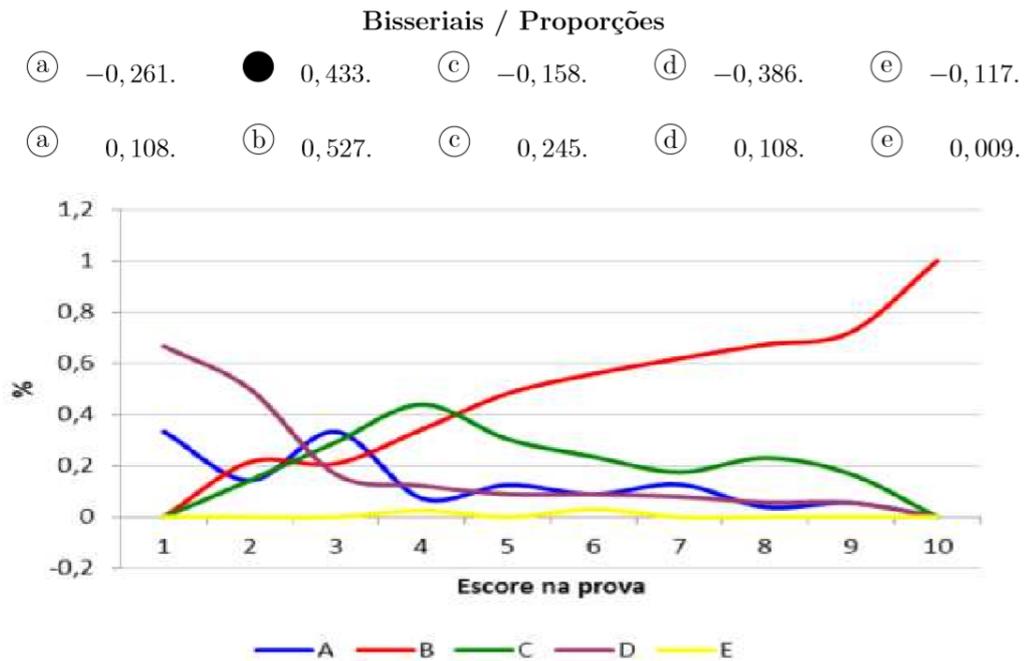


Figura 4.5: Análise Gráfica do Item 16 - P2 1/2015

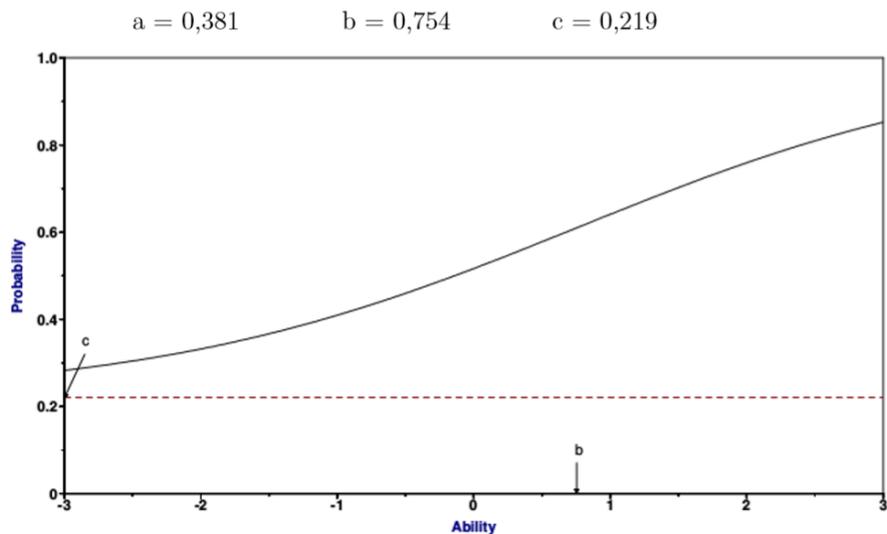


Figura 4.6: Curva Característica do Item 16 - P2 - 1/2015

Exemplo 3

Apresenta-se, na figura 4.7, o item 26 do 2º teste do 1/2015.

Segundo o gráfico da AGI, a alternativa 'd' apresentou preferência nas respostas dos estudantes que obtiveram escore total na prova de 2 até próximo a 9. O coeficiente bisserial positivo dessa alternativa aponta erro em sua construção como atrativa para estudantes de alta proficiência. Quando olhamos para as proporções de marcações, temos que 49,2% dos estudantes que responderam a questão escolheram

T6 - Derivadas: definição e regras básicas

H7 - Aplicar métodos adequados para análise e resolução de problemas.

Nível de dificuldade: *Difícil*

O número de pontos críticos da função $f(x) = 2 \sin x + \cos(2x)$, no intervalo $(-\pi, \pi)$, é igual a

- (a) 1. (b) 3. (c) 5. (d) 2. ● 4.

Figura 4.7: Item 26 - P2 - 1/2015.

essa alternativa como resposta e apenas 15,7% escolheram o gabarito, representado pela letra 'd'. Isso nos mostra que foi difícil, entretanto parte desse resultado tem o viés já citado do bisserial positivo que atraiu os alunos para a alternativa equivocada.

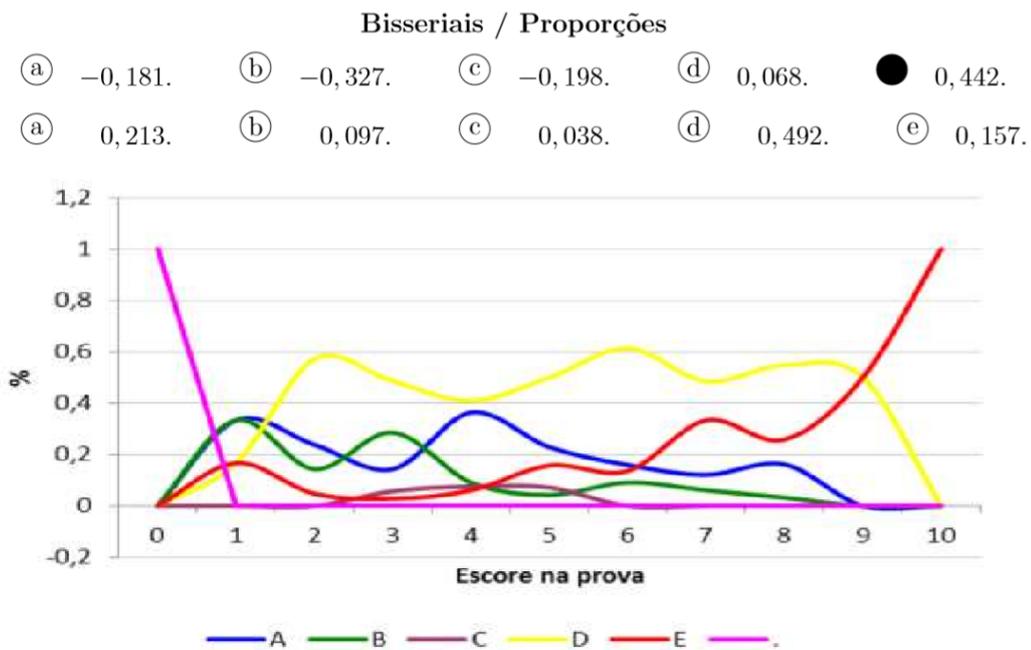


Figura 4.8: Análise Gráfica do Item 26 - P2 1/2015

Podemos verificar pela CCI, ilustrada na figura 4.9, que o item teve baixa chance de acerto aleatório, o que é bom. Considerado um item de moderada discriminação ($a = 0.762$), sua dificuldade foi muito elevada, já que $b = 2,642$, o que indica mais de dois desvios-padrão acima da média. Geralmente itens muito difíceis são de baixo valor de discriminação, como se vê no formato da curva, já que a sigmoide vai ocorrer bem mais á direita no eixo das proficiências, quase nem aparecendo no gráfico o ponto de inflexão da curva. Trata-se de um item com discriminação moderada, mas muito difícil. A curva rosa, que se identifica como '.', representa as questões deixadas

em branco ou anuladas.

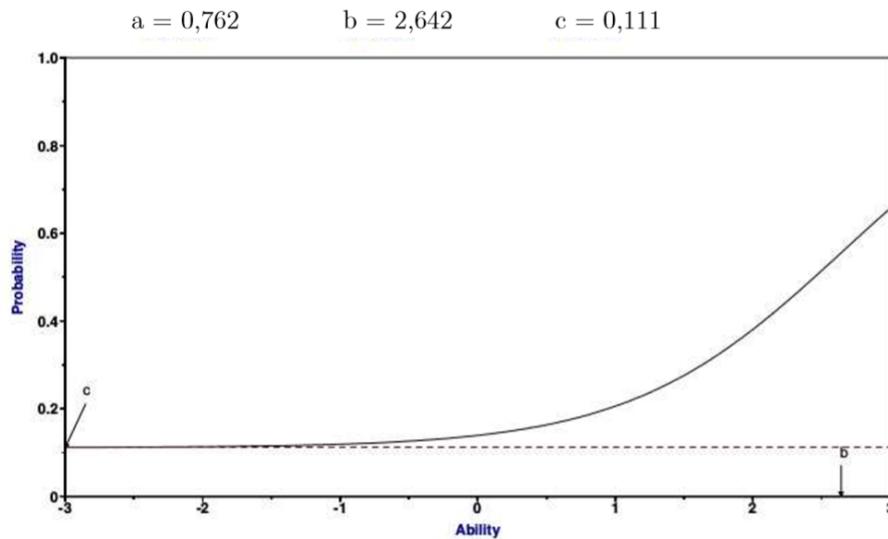


Figura 4.9: Curva Característica do Item 26 - P2 - 1/2015

Inicialmente, de acordo com as recomendações da engenharia de construção de itens descrita em Rabelo (2013), as respostas numéricas nas alternativas devem ser apresentadas em ordem crescente ou decrescente. Desse modo, sugere-se que as opções sejam reorganizadas, sem a necessidade de “embaralhar”, o que poderia induzir a algum tipo de resposta. Em relação à análise dos erros, se pensarmos no que levou os alunos a escolher a opção ‘d’, poderíamos fazer alguns inferências: o aluno pode ter tentando resolver o problema de forma gráfica e acabou tirando conclusões imprecisas, ou tentou resolver o problema mas confundiu-se quando foi usar a regra da cadeia, caindo em situações equivocadas, entre outras possibilidades.

Capítulo 5

Proposta de criação do CAT para implantação do exame de proficiência em Cálculo

Para a construção e interpretação de uma escala de proficiência dos estudantes, após as análises feitas com o uso da TRI, que gera um valor de proficiência para cada aluno e parâmetros para cada item, cada item passa a ser considerado âncora de determinado nível da escala, de acordo com certos padrões estabelecidos para isso. Após ser inserido na escala, procede-se à interpretação pedagógica de cada item, fazendo-se o que se denomina “interpretação da escala” de proficiência. Isso permitirá comparar alunos, submetidos a testes diferentes em tempos diferentes, inclusive comparar diferentes turmas, atribuindo a elas a localização de seus estudantes na escala, o que pode amparar futuras alterações em metodologias de aprendizagem.

Para ser considerado item âncora de determinado nível da escala, os indicadores associados ao item devem atender aos seguintes requisitos: apresentar 65% ou mais de respondentes que estão em torno daquela faixa da escala acertando o item, menos de 50% dos estudantes que estarão no nível anterior acertam o item e a diferença entre os percentuais dos que acertam seja maior que 30%. Alguns autores consideram muito rígidas essas exigências e utilizam uma proposta alternativa para inserir mais itens na escala: alterar o índice de menos de 50% dos estudantes que estarão no nível anterior acertam o item para menos de 65% desses e, ao invés da diferença dos percentuais dos que acertam, apenas verificar o ajuste da CCI pode ser classificado como bom. (RABELO, 2013). No início do ano de 2016, os professores do MAT fizeram algumas reuniões do tipo como oficinas pedagógicas com o objetivo de analisar o banco de itens já existente para trabalhar na construção de uma escala de proficiência para as avaliações de Cálculo. O trabalho consistiu em agrupar os itens em faixas da escala de meio desvio-padrão cada uma e realizar a interpretação pedagógica da escala, por meio

da interpretação dos itens. Essa construção se iniciou com a escolha dos itens já analisados e que estavam dentro dos parâmetros aceitáveis. Utilizando a curva (CCI) da TRI, separou-se os itens nas faixas da escala de proficiência. Por exemplo, na CCI do item 2, da segunda avaliação de 2015, que apresenta o parâmetro de dificuldade igual 0,176, conforme figura 10, determinou-se o nível mínimo de proficiência necessário para acerto de pelo menos 65% dos respondentes. Com esse processo descobre-se para que faixa da escala o item será considerado âncora.

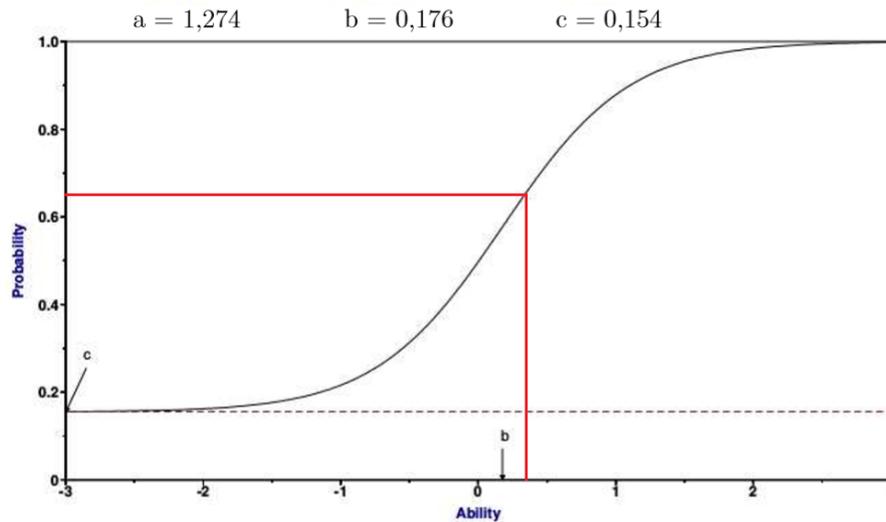


Figura 5.1: Curva Característica do Item 01 - P2 - 1/2015

Encontrando-se a proficiência do item, definiu-se, então, cada faixa em meio desvio-padrão e procedeu-se a análise pedagógica para verificar os conteúdos, habilidades e competências que cada item da faixa representava. Um limitante dessa análise, devido a sua construção ser realizada através de itens âncoras, é se existem itens já testados em número suficiente em todas as faixas dessa escala. Itens que estiverem com o nível de dificuldades muito próximo à fronteira deverão ser descartados para a construção da escala, pois esses não agregam informações úteis para definir em qual faixa o avaliando se apresenta. (OLIVEIRA; FRANCO; SOARES, 2007).

Após a análise do desempenho dos alunos e da releitura dos itens, os professores do MAT estão construindo a interpretação pedagógica das escalas de proficiência, o que gerará uma régua que funcionará como feedback para os alunos de Cálculo I, o que, para o futuro, poderá também dar suporte na decisão de aprovação do aluno pelo significado do que ele sabe, olhando para a interpretação da escala de proficiência e não, simplesmente, para um número arbitrariamente escolhido. Após uma análise do que seja o nível adequado ou suficiente para aprovação em Cálculo I, pensando-se, inclusive, na continuidade em disciplinas que exigem o Cálculo como um requisito, será estabelecida a proficiência para aprovação.

A metodologia permite que os estudantes que resolvem o teste sejam classificados em determinada faixa ou nível de proficiência, que estará vinculada a um conjunto de conceitos aprendidos e de habilidades desenvolvidas. Isso possibilitará a devolutiva de um feedback que representa uma boa probabilidade de um estudante que está em uma faixa específica da escala ter desenvolvido habilidades e dominar conteúdos expressos na faixa de proficiência em que se encontra e em todas as faixas abaixo dessa.

É interessante observar que a análise via TRI e TCT que trará a informação necessária para distribuição dos itens nas faixas ou níveis da escala, considerando seus níveis de dificuldade fácil, médio e difícil, de acordo com as respostas dos alunos aos itens, e não a priori como uma análise dos especialistas da área. O que fica a cargo dos especialistas é definir qual é o papel do item em relação às habilidades, competências e conteúdos avaliados.

A utilização das metodologias de análise de itens e o suporte computacional possibilitaram realizar essa escala. Dado os resultados que estão sendo estabelecidos com a construção do banco de item no MAT, podemos analisar uma nova proposta como um caminho a seguir nas avaliações de Cálculo, aproveitando a inclusão digital e assim os novos caminhos e possibilidades.

Pensando no suporte computacional para avaliação dos itens, podemos repensar a avaliação feita com lápis e papel e verificar a possibilidade de implantação do Computerized Adaptive Testing (CAT) ou teste adaptativo computadorizado (TAC), que consiste em um teste computadorizado e adaptado. O termo adaptado relaciona-se ao modo que o computador irá adaptar o teste à capacidade do testando (COHEN; SWERDLIK; STURMAN, 2014 p.15). Dessa forma, não necessariamente os estudantes realizarão o mesmo teste, pois a escolha dos itens estará disposta ao candidato que responde as questões, ou seja, conforme a proficiência do candidato estiver sendo testada, ele receberá itens que melhor ajudarão a definir sua proficiência, deixando assim de realizar itens que possuem um nível de dificuldade muito abaixo do que o candidato está habilitado a realizar, fazendo com que o testando não perca tempo e energia com questões desnecessárias.

Nas avaliações, um ponto importante que se discute é o tempo de prova, sendo que conforme o estudante demora muito com um teste, o teste pode sofrer um viés de cansaço do testando, assim deixando de avaliar a sua real proficiência.

Dessa forma, os itens são apresentados com base no desempenho nos itens anteriores, de acordo com a administração computacional do teste, podendo não ofertar um item de um nível superior até que o indivíduo não acerte dois ou mais itens dessa mesma proficiência na escala, ou mesmo propor que após um número de itens corretos respondidos na mesma escala dizer que o desempenho do aluno convergiu para essa faixa de proficiência. Isso se dá de acordo com o mecanismo escolhido para a máquina

de funcionamento do CAT. Para tanto é necessário um bom número de itens âncoras, ou seja, calibrados e relacionados a escala de proficiência.

Ao final do teste, não será dada a pontuação do estudante como um score, pois acertar um número de itens não o deixa em uma classificação idêntica a outro que acertou o mesmo número de itens, lembrando que o objetivo é definir qual a escala de proficiência do indivíduo, segundo os pressupostos da TRI.

O CAT tende a diminuir o efeito chão e o efeito teto. O efeito chão consiste em errar todos os itens e o efeito teto em acertar todos os itens. Isso porque, a partir do primeiro item, com erros sucessivos, o teste é adaptado para buscar itens de menor proficiência, para então poder catalogar alguma proficiência do estudante e isto tende ao limite em que o banco de itens é construído. Da mesma forma, o testando que acerta todos os itens de uma avaliação, não precisará passar necessariamente por itens de proficiência muito baixa, economizando itens e tempo e terá de forma objetiva voltado a itens de maior proficiência e também de maior dificuldade, assim discriminando melhor tanto entre os testandos de alta proficiência como os de baixa proficiência. (COHEN; SWERDLIK; STURMAN, 2014 p.15) Tal capacidade computacional é apresentada por esses autores com o nome de ramificação de itens.

Além dessas vantagens, o teste computacional é mais abrangente também na construção dos itens, ampliando as possibilidades de investigação de desenvolvimento de competências, pois poderá conter imagens com melhor definição, vídeos e construções inovadores através da realidade virtual, inclusive com a proximidade do que temos em jogos virtuais, onde o estímulo dos estudantes podem ser ampliados. Aponta Liden e Glas (2000) que um teste de perfil deverá seguir (1) Atender todas as restrições de teste; (2) conter todos os itens anteriormente administrados; (3) ter o máximo de informação na estimativa da atual capacidade; e que o item a ser aplicado ao testando, será aquele com a informação máxima no teste de perfil. As próximas ações estão atualizando a estimativa capacidade do examinando, retornando os itens não utilizados para a banco de itens, e repetindo o procedimento.

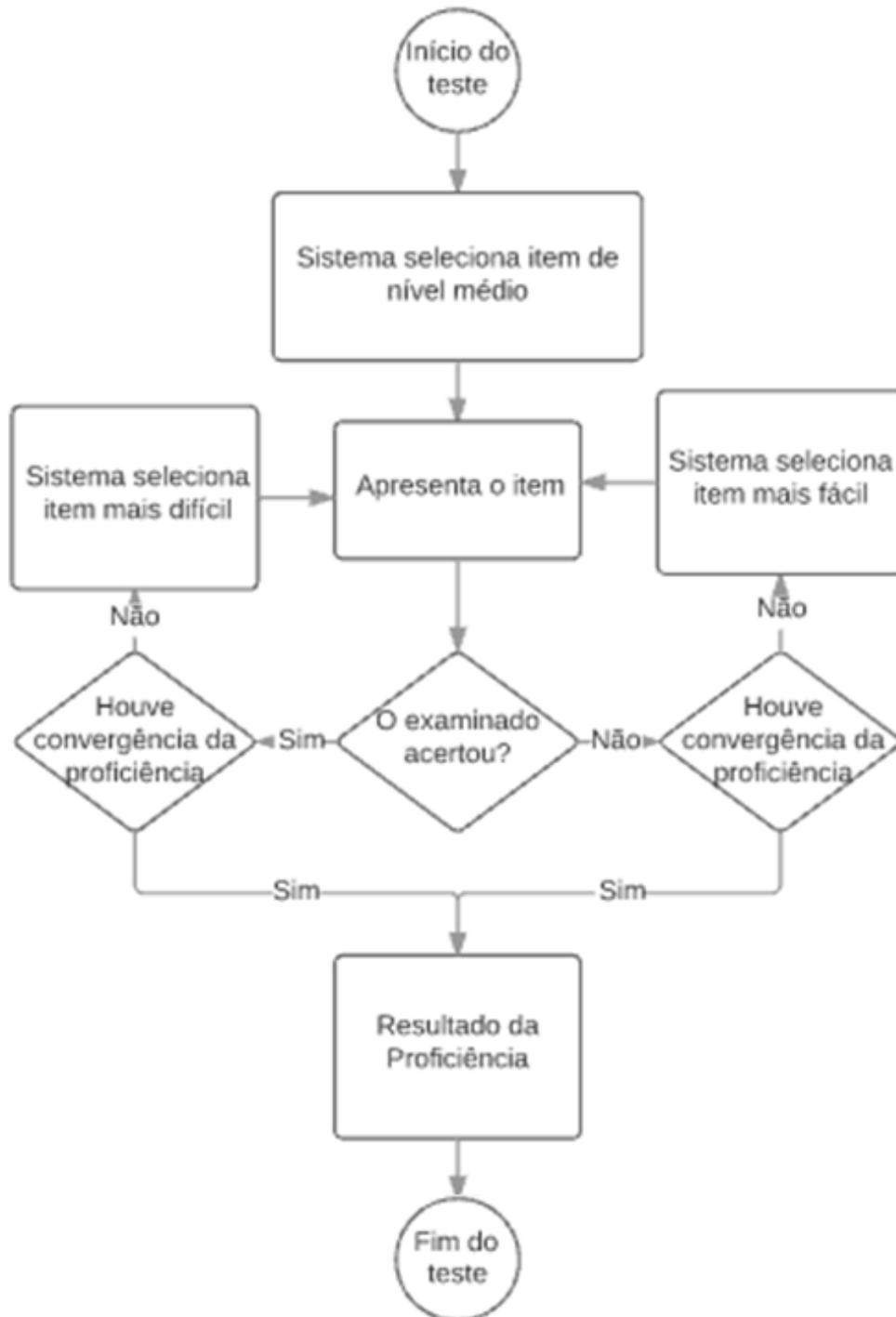
A prática de boa parte dos programas de CAT existentes consiste em adequar o comprimento do teste. Embora a regra de parada com base em determinado nível de precisão para o estimador de habilidade é forma desejável de um ponto de vista estatístico, parece impossível garantir recortes das mesmas especificações para todos os examinandos, de um teste com comprimento aleatório. (LINDEN; GLAS, 2000).

Dado que a proficiência verdadeira é desconhecida, tudo que se pode esperar é um mecanismo de seleção de itens com um resultado que deve se aproximar da proficiência real. Os especialistas dizem que CAT tem o recurso de produzir um teste viável convergindo para o valor ideal da proficiência dos examinandos.

A velocidade de convergência depende do tamanho e da natureza do conjunto de restrições. Para um teste de adaptação severamente restrito, a convergência

pode ser mais lenta do que para um teste a partir de apenas algumas restrições. Para obter mais informações, orienta-se a leitura do capítulo 2 de Linden e Glas (2000).

No que se segue, será apresentado um fluxograma que ilustra as linhas gerais de funcionamento do CAT. Obviamente não estão presentes os detalhes de funcionamento do algoritmo, mas o esquema traz uma proposta simplificada de construção de um CAT.



Assim, a proposta de uma próxima etapa a ser desenvolvida no MAT diz respeito à construção de um CAT para Cálculo I, o qual fará uso do banco de itens que está sendo construído e calibrado nos últimos três anos, que já constitui um bom ponto de partida para o teste computadorizado. Naturalmente, há um problema de logística a ser resolvido no que diz respeito à disponibilização de laboratórios de computação para aplicação dos testes, com garantia de estabilidade da rede de internet, como o suporte para número de alunos que estarão realizando o teste no mesmo momento, para que o teste funcione de modo online, e não off-line.

Capítulo 6

Considerações finais

Ao iniciar os estudos com o objetivo de desenvolver um modelo de exame de proficiência em Cálculo Diferencial e Integral, para o Departamento de Matemática da Universidade de Brasília, deparei com o primeiro e principal ponto que é condutor da pesquisa, a dificuldade dos alunos no aprendizado em matemática, que está fortemente relacionado com o aprendizado do Ensino Básico.

Dado a problemática do ensino em matemática, vemos que o suporte a metodologias mais precisas para avaliação pode ajudar a desenvolver um sistema para os professores avaliarem e compararem com mais clareza as metodologias utilizadas nas aulas. O trabalho aqui apresentado limitou-se à análise das respostas dos estudantes aos itens fechados, de múltipla escolha, elaborados a partir de um minucioso planejamento que teve como suporte uma Matriz de Referência criada especialmente para isso. Não chegamos a considerar as questões dissertativas, onde existe maior liberdade de elaboração das respostas pelos estudantes, até mesmo porque o objetivo era conhecer as potencialidades do banco de itens de múltipla escolha que está sendo construído pelo Departamento de Matemática da UnB.

Assim, realizou-se um levantamento sobre o ato de avaliar, considerando os atores principais - professor e alunos, com suas subjetividades. Para tanto, explorou-se a perspectiva de uma educação voltada ao desenvolvimento de competências, o processo denominado de engenharia de construção de itens e metodologias de análise de desempenho em testes embasadas na TCT e na TRI. Para acompanhar o processo pedagógico e conhecer um pouco da história das mudanças recentes implantadas no MAT, foi realizada uma entrevista semiestruturada com professores do departamento, o que levou à conclusão de que as práticas atualmente adotadas caminham em um sentido muito positivo, uma vez que o aluno tem maior interação com as aulas, em uma postura mais ativa, desde a implantação do sistema de aula magistral, composto de aulas magnas seguidas de duas aulas de atividades realizadas em grupo, sob a orientação de professores e monitores. Os resultados de desempenho de estudantes

a partir das mudanças feitas têm sugerido melhorias nos indicadores das taxas de sucesso/aprovação nas turmas de Cálculo, principalmente se considerado o cenário atual em que a Universidade tem recebido maior quantitativo de estudantes que ingressam por meio das políticas de inserção social, que inclui reserva de 50% de vagas em seus cursos de graduação para candidatos que estudaram integralmente o ensino médio em escolas públicas.

Ao final do trabalho, foi comentado acerca da possibilidade de implantação de um modelo de criação de feedback para os alunos, mas não foi exibido o modelo em si, que era um dos objetivos específicos, uma vez que o grupo que desenvolve a experiência no MAT ainda não chegou ao ponto de interpretar a escala de proficiência dos estudantes. Aproveitando a oportunidade da construção do banco de itens para Cálculo I, o trabalho encerrou-se com um comentário sobre a criação de um exame computacional adaptado, que, além de contribuir com a logística de avaliação para muitas turmas, também apresenta a vantagem de se ter uma prova personalizada para cada aluno, tentando tratar com equidade os participantes das avaliações.

Julgamos que o trabalho atingiu o objetivo de registrar esse esforço que o MAT tem feito para melhorar o aprendizado dos estudantes em Cálculo I e de ajudar os professores a elaborarem itens de avaliação que sejam mais calibrados do ponto de vista da dificuldade e da discriminação, aproximando-se mais a avaliação de seu propósito nobre de servir para regular a aprendizagem dos estudantes. A caminhada é longa, mas os resultados preliminares apontam que o MAT está indo na direção correta.

Referências Bibliográficas

ALMEIDA, L.; MARINHO-ARAÚJO, C. M.; AMARAL, A.; DIAS, D. Democratização do acesso e do sucesso no Ensino Superior: uma reflexão a partir das realidades de Portugal e do Brasil. *Avaliação*, 17 (3), 899-920, 2012.

ANDRIOLA, W. B. *Psicometria Moderna: características e tendências*. Est. Aval. Educ., São Paulo, v. 20, n. 43, maio/ago. 2009.

BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70 2006.

BARUFI, M. C. B. A. *Construção/negociação de significados no curso universitário inicial de Cálculo Diferencial e Integral*. Tese de Doutorado. São Paulo: FE-USP. 1999.

BRESSOUD, D. ; MESA, V. ; RASMUSSEN, C. *Insights and Recommendations from the Mathematical Association of America National Study of College Calculus*. 2015.

CASTIONI, R. *Educação no mundo do trabalho: qualificação e competência*. São Paulo: Francis, 2010.

COHEN, R.J. SWERDLIK, M.E. e STURMAN, E.D. *Testagem e Avaliação Psicológica: Introdução a testes e medidas*. 8º edição Ed. artmed 2014.

CUNHA, L. A. (2007). *A universidade crítica: o ensino superior na república populista*. 3 ed. São Paulo: Editora UNESP.

D'AMBRÓSIO, U. *A era da consciência: aula inaugural do primeiro curso de pós-graduação em ciências e valores humanos no Brasil*. São Paulo: Editora

Fundação Petrópolis, 1997.

FREIRE P. Pedagogia da autonomia: saberes necessários à prática educativa 48^o Ed. – Rio de Janeiro. Editora: Paz e Terra, 2014.

LINDEN W.J. ;LINDEN V.D.; Glas A. W.C. Computerized Adaptive testing, Theory and Practice Ed. Kluwer Academic Publishers, 2000.

MARINHO-ARAÚJO C.M.; RABELO M.L. Avaliação educacional: a abordagem por competências. Avaliação, Campinas; Sorocaba, SP, v. 20, n. 2, p. 443-466, jul. 2015.

MELLO, M.H.C.S; MELLO, J.C.C.B.S. Reflexões sobre o ensino de cálculo. COBENGE, XXXV Congresso Brasileiro de Educação em Engenharia, 2007. Curitiba. Paraná, 2007.

MUÑIZ, J. Teoria Clássica dos Testes. Madri: Ediciones Pirámide, SA, 2003.

NAGEL T.S. e RICHMAN P.T. Ensino para competência, Uma estratégia para eliminar o fracasso; trad. Cosete Ramos, M. A. – Editora Globo – Porto Alegre, 1976.

NASSER, L. Ajudando a superar obstáculos na aprendizagem de cálculo. . In: Encontro Nacional de Educação Matemática, 2007. Belo Horizonte. Anais. Belo Horizonte. 2007.

OLIVEIRA, L.K.M; FRANCO, C.; SOARES, T.M. Projeto GERES / 2005: Novos indicadores para a construção e interpretação da escala de proficiência. Revista Iberoamericana sobre Qualidade, Eficácia e Mudança em Educação 2007. Volume 5, numero 2 Disponível em: <http://www.rinace.net/arts/vol5num2e/art11_htm.htm>. Acesso em: 04 mai. 2016; 2007

PASQUALI, L. Psicometria: teoria dos testes na psicologia e na educação. 3. ed., Petrópolis-RJ: Vozes, 1999.

PASQUALI, L. Psicometria. REEUSP, p. 992-999, 2008.

PENIN, S.T. Texto: Didática e Cultura: O Ensino comprometido com o social e a contemporaneidade. In: Ensinar a ensinar. Castro, A. D. & Carvalho, A.M.P. São Paulo: Pioneira/Thomson Learning. Exposição Orale Discussão em Grupo, 2001.

PERRENOUD P. Construir as Competências desde a Escola; trad. Bruno Charles Magne. – Porto Alegre: Artes Médicas Sul, 1999.

PORTAL MEC, Parecer CNE/CES nº 1.302, de 6 de novembro de 2001 Diretrizes Curriculares Nacionais para os Cursos de Matemática, Bacharelado e Licenciatura. Disponível em: <<http://portal.mec.gov.br/cne/arquivos/pdf/CES13022.pdf>>. Acesso em: 26 mai. 2016.

RABELO M.L., MAGALHÃES C.A., FURTADO M. MARINHO-ARAÚJO C.M. A Matriz de Referência como Suporte para Avaliação de Competências em Cálculo Diferencial e Integral, 2016

RABELO, M. L.; MAGALHÃES, C. A.; FURTADO, M. Criação de bancos de itens calibrados como subsídio para avaliação de cálculo diferencial e integral, 2015

RABELO, M. L.. Avaliação Educacional: fundamentos, metodologia e aplicações no contexto brasileiro. Coleção ProfMat. Rio de Janeiro: SBM, 2013.

REZENDE, W. M.. O Ensino de Cálculo: Dificuldade de natureza Epistemológica. Tese de Doutorado. São Paulo: FE-USP, 2003.

RISTOFF, D. O novo perfil do campus brasileiro: uma análise do perfil socioeconômico do estudante de graduação. Avaliação, 19 (3), 723-747, 2014.

SANTOS, B. P. Paulo Freire e Ubiratan D'Ambrosio: Contribuições para a formação do professor de matemática no Brasil. Tese de Doutorado. São Paulo: FE-USP, 2007.

TORRES, Patrícia L.; ALCANTARA, Paulo R.; IRARA, Esrom A. F. Grupos de Consenso: Uma Proposta de Aprendizagem Colaborativa para o Processo de Ensino-Aprendizagem. Revista Diálogo Educacional, v. 4, n.13, p. 129-145, 2004.

VIANNA, H. M. Introdução à Avaliação Educacional. São Paulo: Ibrasa,

1989.

WILSON, D. T.; WOOD, R.; GIBBONS, R. Testfact: test scoring, item statistics, and item factor analysis. Chicago: Scientific Software, 1991.