**Transposição da Teoria da
Resposta ao Item: uma
abordagem pedagógica**

Eder Alencar Silva

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
MESTRE EM CIÊNCIAS

Programa: Mestrado Profissional em
Ensino de Matemática
Orientadora: Professora Dr^a. Lisbeth Kaiserlian Cordani

São Paulo, julho de 2017

Transposição da Teoria da Resposta ao Item: uma abordagem pedagógica

Esta versão da dissertação contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 23/06/2017. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Professora Dr^a. Lisbeth Kaiserlian Cordani (orientadora) - IME-USP
- Professora Dr^a. Mariana Cúri - ICMC-USP
- Professora Dr^a. Viviana Giampaoli - IME-USP

Agradecimentos

Primeiramente a Deus, por me conceder a vida e me permitir realizar este trabalho.

Às minhas mães – avó e tias – que me criaram e, mesmo com todas as dificuldades, conseguiram me dar condições para estudar.

Aos meus pais que, apesar de não terem me criado, contribuíram de alguma forma para o meu crescimento como ser humano.

À minha amada esposa Tatiana – amiga e companheira – que recém-casada teve toda paciência e compreensão da minha importância pessoal sobre a realização do mestrado.

À professora Dra. Lisbeth Kaiserlian Cordani – orientadora – pelo empenho, dedicação, paciência e aprendizado.

Aos professores do IME-USP, tanto da graduação quanto do mestrado, pelas contribuições dadas ao meu aprendizado durante os anos em que passei neste Instituto.

À Instituição em que trabalho atualmente, por acreditar em mim e possibilitar a conclusão do Programa de Mestrado Profissional em Ensino de Matemática do IME-USP.

Resumo

SILVA, E.A. **Transposição da Teoria da Resposta ao Item: uma abordagem pedagógica.** 2017. 179 f. Dissertação (Mestrado) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

Este trabalho tem por objetivo apresentar a Teoria da Resposta ao Item (TRI), por meio de uma abordagem pedagógica, aos professores da educação básica, que mencionaram esta necessidade por meio de pesquisa realizada pelo autor. Levar parte do conhecimento teórico que embasa esta teoria ao conhecimento do docente, principalmente a construção da curva de probabilidade de acerto do item, favorecerá a compreensão, a análise e o monitoramento do processo avaliativo educacional. Este material apresenta as principais definições e conceitos da avaliação externa em larga escala, além de fornecer insumos para a compreensão das suposições realizadas para aplicação da metodologia. Neste sentido, o texto foi estruturado de forma a apresentar didaticamente as etapas do processo de implementação de uma avaliação, desde a construção do item até a apuração e divulgação dos resultados. Todo enfoque será dado à construção do modelo da TRI com um parâmetro (dificuldade do item), também conhecido como modelo de Rasch, o que simplifica e facilita a compreensão da metodologia. O modelo utilizado nas avaliações externas em larga escala (modelo com três parâmetros) será introduzido a partir das considerações realizadas na abordagem que explicita o pensamento da construção do modelo de um parâmetro. Acredita-se que esta compreensão possa colaborar com o professor na exploração das habilidades/competências dos alunos durante os anos escolares.

Palavras-chave: Teoria da Resposta ao Item, TRI, Modelo de RASCH, Modelo logístico de dois parâmetros, Modelo logístico de três parâmetros, Avaliação Externa em Larga Escala.

Abstract

SILVA, E.A. **Transposition of Item Response Theory: a pedagogical approach**. 2017. 179 f. Dissertação (Mestrado) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

This study aims to present the Item Response Theory (IRT), through a pedagogical approach, to teachers of basic education, which mentioned this necessity through research conducted by the author. To take part of the theoretical knowledge that supports this theory to the teacher's knowledge, especially the construction of probability curve of item correct response, it will favor for understanding, analysis and monitoring the evaluation educational process. This material presents the main definitions and concepts of the external evaluation in large scale, besides providing inputs for understanding the assumptions made to apply the methodology. In this sense, the text was structured in order to present the implementation process stages of a large scale assessment, from the item construction to the results calculation and dissemination. The focus will be given to the IRT model construction of one-parameter (difficulty of the item), also known as Rasch model, since it simplifies and facilitates the understanding of methodology. The model used in the external assessment on a large scale (three-parameter model) will be introduced from the considerations made in the approach that explicit the thought of one-parameter model construction. It is believed that understanding can collaborate with teacher in exploration of the students' skills/competences during the school year.

Key words: Item Response Theory, IRT, RASCH Model, Two-Parameter Logistic Model, Three-Parameter Logistic Model, Large-scale External Assessment.

Sumário

Lista de Abreviaturas	vi
Lista de Figuras	viii
Lista de Tabelas	xii
Lista de Quadros	xiii
Introdução	14
1. Introdução à Avaliação Externa em Larga Escala	22
2. Processo de Implementação de uma Avaliação Externa em Larga Escala	26
2.1. População a Ser Avaliada.....	27
2.2. Matriz de Referência	30
2.2.1. Matriz de Referência do SARESP	31
2.3. Elaboração de Itens	35
2.4. Pré-Testagem ou Calibração do Item	40
2.5. Análise Estatística e Pedagógica dos Itens	41
2.6. Montagem dos Cadernos de Provas	42
2.7. Aplicação das Provas e Processamento da Base de Dados	47
3. Breve Introdução à Teoria da Resposta ao Item (TRI)	49
4. Abordagem Pedagógica sobre a Construção dos Modelos da TRI.....	53
4.1. Construção do Gráfico a Partir de Dados Empíricos	53
4.2. Curva Característica do Item (CCI)	76
4.3. Ajuste da Curva Logística	83
4.3.1. Modelo Logístico de 1 Parâmetro (Modelo de Rasch).....	87
4.3.2. Modelo Logístico de 2 Parâmetros	95
4.3.3. Modelo Logístico de 3 Parâmetros	105

4.4. Situações de Estimação dos Parâmetros dos Itens e das Habilidades	118
5. Cálculo da Habilidade de Um Respondente	120
6. Construção da Escala de Proficiência e Divulgação dos Resultados	132
6.1. Construção da Escala de Proficiência	132
6.2. Divulgação dos Resultados	143
Considerações Finais	145
Apêndices	148
A. Pesquisa Sobre a TRI Com Professores da Educação Básica	148
B. Exemplo Para Compreensão da Metodologia BIB	159
Anexo A.....	165
Questionário da Pesquisa Apresentada no Apêndice A	165
Referências Bibliográficas	170

Lista de Abreviaturas

ANA – Avaliação Nacional da Alfabetização

ANEB – Avaliação Nacional da Educação Básica

ANRESC – Avaliação Nacional do Rendimento Escolar

BIB – Blocos Incompletos Balanceados

CAEd – Centro de Políticas Públicas e Avaliação da Educação

CCI – Curva Característica do Item

CENP – Coordenadoria de Estudos e Normas Pedagógicas

CONBRATRI – Congresso Brasileiro de Teoria da Resposta ao ITEM

DE – Diretoria de Educação

EF – Ensino Fundamental

EJA – Educação de Jovens e Adultos

EM – Ensino Médio

ENEM – Exame Nacional do Ensino Médio

FDE – Fundação para o Desenvolvimento da Educação

IBGE – Instituto Brasileiro de Geografia e Estatística

IES – Instituição de Ensino Superior

IME – Instituto de Matemática e Estatística

INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

PIRLS – Progress in International Reading Literacy Study

PISA – Programme for International Student Assessment

PROUNI – Programa Universidade para Todos

SAEB – Sistema Nacional de Avaliação da Educação Básica

SARESP – Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo

SEE-SP – Secretaria da Educação do Estado de São Paulo

SIMAVE – Sistema Mineiro de Avaliação da Educação Pública

SME – Secretaria Municipal de Educação

TCT – Teoria Clássica dos Testes

TRI – Teoria da Resposta ao Item

UFJF – Universidade Federal de Juiz de Fora

USP – Universidade de São Paulo

WMLN – World's Most Literate Nations

Lista de Figuras

Figura 1: Evolução das proficiências médias em matemática do Brasil na Prova Brasil (1995 a 2015).....	17
Figura 2: Médias das proficiências da rede estadual do Estado de São Paulo em relação ao nível adequado por ano/série avaliado – matemática - SARESP... 18	
Figura 2.1: Processo de implementação da avaliação externa em larga escala.....	27
Figura 2.2: Relações entre habilidades, competências e conteúdos.....	31
Figura 2.3: Matriz de referência do SARESP da disciplina de matemática e suas tecnologias - 7º ano do ensino fundamental.....	34
Figura 2.4: Processo de elaboração do item.....	36
Figura 2.5: Curvas da probabilidade de acerto dos itens A, B e C, segundo as proficiências dos alunos do 7º ano do ensino fundamental em matemática. ...	42
Figura 2.6: Posicionamento dos itens A, B e C, em um banco de itens, de acordo com a respectiva dificuldade.	43
Figura 2.7: Agrupamento dos 104 itens em 13 blocos.	45
Figura 4.1: Distribuição do desempenho dos alunos em termos percentuais..	57
Figura 4.2: Distribuição das chances (<i>odds</i>) (acertos/erros) dos alunos.	61
Figura 4.3: Distribuição do logaritmo natural das chances (<i>odds</i>) (acertos/erros) dos alunos.....	64
Figura 4.4: Comparação entre as distribuições do percentual de acertos e o logaritmo natural das chances <i>odds</i> - (acertos/erros) dos alunos.	65
Figura 4.5: Apresentação das distribuições do percentual de acertos, das chances (<i>odds</i>) (erros/acertos) e do logaritmo natural das chances (<i>odds</i>) dos itens da avaliação.....	67

Figura 4.6: Distribuição do logaritmo natural das chances (<i>odds</i>) dos itens e dos alunos.....	69
Figura 4.7: Percentual de acertos na avaliação devido à habilidade do aluno.	74
Figura 4.8: Curva ajustada ao percentual de acertos na avaliação devido a habilidade do aluno.	75
Figura 4.9: Previsão de acertos na avaliação com base na curva ajustada. ...	76
Figura 4.10: Distribuição dos alunos segundo sua habilidade estimada.	78
Figura 4.11: Percentual de acertos no item “k” por habilidade do aluno.....	78
Figura 4.12: Curva característica ajustada ao item “k”.	81
Figura 4.13: Comparativo entre as curvas características ajustadas para cada item.	83
Figura 4.14: Curva da distribuição normal acumulada - ogiva normal.....	86
Figura 4.15: Curva logística.....	87
Figura 4.16: Curvas características do modelo de Rasch.	90
Figura 4.17: Curvas da probabilidade de acerto do item, à medida que se eleva a dificuldade em 0,2 logito.....	91
Figura 4.18: Curvas da probabilidade de acerto do item, à medida que se diminui a dificuldade em -0,2 logito.....	91
Figura 4.19: Curva característica do item “k” ajustada às estimativas da probabilidade de acerto.....	93
Figura 4.20: Representação da dificuldade do item “k” por meio gráfico.....	94
Figura 4.21: Gráfico característico do item, discriminação ótima.	96
Figura 4.22: Gráfico característico do item, discriminação boa.	97
Figura 4.23: Gráfico característico do item, ausência de discriminação.....	99

Figura 4.24: Curvas características de itens do modelo de dois parâmetros para itens com a mesma dificuldade.	99
Figura 4.25: Curvas características de itens do modelo de dois parâmetros para itens com dificuldades distintas.	100
Figura 4.26: Curvas ajustadas às estimativas da probabilidade de acerto do item “k” pelos modelos de Rash e de 2 parâmetros.	105
Figura 4.27: Curvas características de itens do modelo de três parâmetros.	107
Figura 4.28: Curva ajustada às estimativas da probabilidade de acerto do item “k” para o modelo de 3 parâmetros.	116
Figura 4.29: Curvas ajustadas às estimativas da probabilidade de acerto do item “k” para os modelos de Rasch, de 2 e de 3 parâmetros.	117
Figura 5.1: Curvas características dos quatro itens considerados na composição da avaliação.	121
Figura 5.2: Curvas das probabilidades de acerto e de erro para o item “k”... ..	122
Figura 5.3: Curvas das probabilidades de erro para cada um dos quatro itens considerados na composição da avaliação.	122
Figura 5.4: Curva da probabilidade conjunta do caso 1.	124
Figura 5.5: Comparativo entre as curvas da probabilidade conjunta dos casos 1 e 2.	125
Figura 5.6: Curva característica dos itens que compõem a avaliação de matemática.	128
Figura 5.7: Curva da probabilidade de ocorrência do padrão de resposta do respondente em relação à habilidade.	129
Figura 6.1: Representação da transformação linear entre escalas (0,1) e SAEB para o 9º ano do ensino fundamental.	134
Figura 6.2: Exemplo de itens âncoras.	137

Figura 6.3: Esquema de ligação entre itens de anos avaliados.....	143
Figura A.1: Distribuição dos professores por sexo.	149
Figura A.2: Distribuição dos professores por sexo e formação.	149
Figura A.3: Formação em relação a pós-graduação.	149
Figura A.4: Modalidade da educação básica em que ministra aula na rede pesquisada.	150
Figura A.5: Anos completos na profissão de professor.	150
Figura A.6: Redes de ensino em que ministra aula, além da rede pesquisada.	150
Figura A.7: Importância dos resultados da avaliação externa na sala de aula.	151
Figura A.8: Conhecimento sobre a metodologia TRI.....	152
Figura A.9: Opinião sobre a melhor forma de mensurar o desempenho do aluno em avaliações do sistema de ensino.....	153
Figura A.10: Conhecimento sobre a construção do gráfico da probabilidade de acerto do item.....	154
Figura A.11: Interesse em saber sobre a construção do gráfico da probabilidade de acerto do item.....	156
Figura A.12: Contribuição da TRI para a prática pedagógica.....	157
Figura B.1: Banco de itens que contemplam todas as habilidades das matrizes de referência de matemática da educação básica.	159
Figura B.2: Seleção de itens que irão compor a prova de matemática do 7º ano do ensino fundamental.	160
Figura B.3: Agrupamento dos itens para composição das provas da avaliação.	163
Figura B.4: Composição dos cadernos de provas da avaliação.....	163

Lista de Tabelas

Tabela 4.1: Tabela de acertos e erros por alunos e itens.	55
Tabela 4.2: Tabela ordenada pelo total de acertos dos alunos e nos itens.	56
Tabela 4.3: Tabela ordenada com os percentuais de acertos dos alunos e nos itens.....	57
Tabela 4.4: Tabela ordenada com as chances (<i>odds</i>) dos alunos.....	61
Tabela 4.5: Tabela ordenada com o cálculo do logaritmo natural aplicado às chances (<i>odds</i>) dos alunos.....	63
Tabela 4.6: Tabela ordenada com o cálculo dos logaritmos naturais das chances (<i>odds</i>) dos alunos e itens.	66
Tabela 4.7: Tabela da habilidade ajustada do aluno (diferença entre a habilidade do aluno e a dificuldade do item).....	72
Tabela 4.8: Percentual de acertos por habilidade do aluno.	73
Tabela 4.9: Percentual de acertos no item “k” por intervalo de habilidades.....	80
Tabela 4.10: Estimativa da probabilidade de acerto do item “k” pelo modelo de Rasch.	92
Tabela 4.11: Estimativa da probabilidade de acerto do item “k” pelo modelo de 2 parâmetros.	103
Tabela 4.12: Parâmetros dos itens apresentados na Figura 4.27.....	107
Tabela 4.13: Estimativa da probabilidade de acerto do item “k” pelo modelo de 3 parâmetros.	114

Lista de Quadros

Quadro 2.1: Exemplo da metodologia BIB para o SARESP.....	46
Quadro 2.2: Agrupamento dos itens de acordo com percentual de acertos....	47
Quadro 2.3: Aplicação das provas segundo o ano avaliado e disciplinas.	48
Quadro 5.1: Casos possíveis de uma avaliação com 4 itens.	126
Quadro 5.2: Estimativa das habilidades dos respondentes da avaliação de matemática.....	130
Quadro 6.1: Proficiência dos respondentes da avaliação de matemática na escala SAEB.	135
Quadro 6.2: Descrição da escala de matemática para o ponto 400.	140
Quadro B.1: Quantidade de cadernos de provas, devido ao número de blocos a serem respondidos pelo aluno.	161

Introdução

As avaliações externas, no âmbito escolar, se justificam pela necessidade de monitorar o ensino, a partir da aprendizagem dos alunos, e fornecer subsídios com base em dados sobre os pontos a serem melhorados (ALAVARSE et al., 2013). Diversos instrumentos já foram criados para mensurar o desempenho, dentre os quais uma boa parte teve como base a Teoria Clássica dos Testes (TCT). Segundo Pasquali (2013):

“A TCT surgiu dentro da concepção monista materialista que imperava nas ciências em geral desde o empiricismo inglês do século XVII...”. (PASQUALI, 2013, p53)

Quando a avaliação em larga escala está baseada nos pressupostos dos postulados da TCT, o seu objetivo fundamental é o resultado geral obtido pelo aluno (GRÉGOIRE & LAVEAULT, 2002), utilizando-se do número total de acertos para medir o domínio em determinada área do conhecimento.

Segundo Sartes (2013), os fundamentos da TCT serviram como base para o desenvolvimento de inúmeros instrumentos de medidas e, principalmente, como referência para abordagens de medidas psicométricas. O uso da TCT em uma grande variedade de situações se deve à linearidade, estabelecida entre o resultado verdadeiro e o resultado observado, e aos seus fracos e poucos postulados (GRÉGOIRE & LAVEAULT, 2002).

Apesar da fácil aceitação e grande utilidade da TCT até os dias atuais, algumas limitações levaram à busca por outra metodologia. Dentre elas, baseando-se em Pasquali (2013), pode-se destacar:

- Os pressupostos da teoria estão baseados apenas no resultado geral (número de acertos) da avaliação;
- Os resultados observados na avaliação são dependentes da amostra de respondentes;
- As estimativas sobre o domínio do aluno em áreas do conhecimento dependem das questões selecionadas para compor a avaliação;
- Assume-se para todos os alunos um mesmo erro de medição;

- Avaliações mais longas são mais confiáveis do que as avaliações mais curtas;
- Os resultados de duas avaliações não podem ser comparáveis, a menos que sejam as mesmas ou sejam avaliações paralelas¹;
- A comparação entre alunos é realizada por meio da proporção de acertos.

Foi a partir da procura de teorias alternativas que pudessem suprir as limitações impostas pela TCT (PASQUALI, 2013) que surgiu a Teoria da Resposta ao Item (TRI). A TRI versa sobre o item respondido, avalia o desempenho geral no teste e descreve como o desempenho relaciona-se com as habilidades medidas pelos itens na avaliação (HAMBLETON & JONES, 1993).

Na TRI, o desempenho de um aluno em cada item de um teste pode ser predito por meio de um conjunto de habilidades. Sendo assim, a habilidade é a causa e o desempenho é a consequência dessa habilidade ou efeito. A equação do modelo matemático da TRI que descreve essa relação leva em conta que quanto maior a habilidade do aluno, maior será a probabilidade de ele acertar o item e o inverso também é verdadeiro.

Segundo Marques (2012), em toda a América Latina a TRI é pouco utilizada e foi somente a partir de 1990 que esta técnica passou a ser conhecida e aplicada no Brasil, sendo hoje utilizada, por exemplo, em avaliações da Prova Brasil, ENEM e do SARESP (Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo). É importante que se diga que a TRI não veio para substituir toda a TCT, mas assumiu um protagonismo importante nos aspectos de análise dos itens e confiabilidade da medida (PASQUALI, 2013). Já em países do hemisfério norte, Marques (2012) destaca o uso da TRI pelos EUA, Canadá e Europa por substituírem grande parte das suas avaliações da aprendizagem baseadas na TCT pela TRI. Inclusive atualmente as principais avaliações externas como

¹ Segundo Pasquali (2013), dois testes são considerados paralelos quando eles medem a mesma coisa, mas possuem questões diferentes.

PIRLS² e PISA³, que visam comparar o nível educacional entre nações, usam a TRI.

Quando o objetivo é avaliar a aprendizagem escolar em um dado momento, a estratégia mais utilizada é a mensuração do desempenho dos alunos a partir de avaliações externas. Neste contexto, é de conhecimento público o fraco desempenho dos alunos brasileiros nas avaliações externas internacionais e nacionais. Políticas públicas são realizadas no Brasil a fim de melhorar a educação no país, obtendo-se até alguns resultados positivos, mas pouca coisa se conseguiu em relação à qualidade de formação do aluno no ensino básico.

Um estudo realizado com 61 nações pela Universidade de Connecticut, Estados Unidos, construiu o *ranking* WMLN (World's Most Literate Nations)⁴, divulgado em março de 2016, mostrando que, dentre essas 61 nações, o Brasil é o primeiro país no ranking da categoria que trata dos “insumos de entrada” da educação, categoria esta que leva em conta, por exemplo, os indicadores de quantidade de anos da educação básica, proporção de pessoas em idade escolar na escola e percentual do PIB destinado a educação. Porém, na categoria que avalia o aprendizado dos alunos, por meio das avaliações internacionais PIRLS e PISA, o Brasil ocupa a 55ª posição.

Mesmo as avaliações nacionais, como por exemplo, a Prova Brasil⁵, não trazem resultados animadores, principalmente para a disciplina de matemática⁶. Pois, se observa que o distanciamento entre as médias das proficiências obtidas, em relação às proficiências esperadas, aumenta à medida que se eleva o ano/série

² O PIRLS (Progress in International Reading Literacy Study) avalia a alfabetização dos alunos mais jovens (final do ensino fundamental I para o final do ensino fundamental II).

³ O PISA (Programa de Avaliação Internacional de Estudantes) é um exame multidisciplinar que avalia os resultados educacionais dos alunos de 15 anos.

⁴ O WMLN classifica as nações de acordo com o nível de educação e recursos de apoio. As classificações baseiam-se em cinco categorias: bibliotecas, jornais, entradas e saídas educacionais e disponibilidade de computadores. O *ranking* geral e por categoria pode ser acessado no site: <http://www.ccsu.edu/wmln/>.

⁵ BRASÍLIA (2016) traz maiores detalhes sobre a Prova Brasil. O Portal do INEP contém mais informações: <http://provabrasil.inep.gov.br/web/guest/educacao-basica/saeb/resultados>.

⁶ Este mau desempenho não é recente. Em 2007, Sadovsky (2007) fez uma discussão sobre o ensino da matemática, propondo revê-lo, interrogá-lo e analisá-lo.

avaliado(a). Ou seja, ao invés dos alunos se aproximarem do desempenho esperado para seu ano/série, eles estão ficando, ao longo do tempo, mais distante do que se espera.

O histórico das proficiências médias da Prova Brasil na disciplina de matemática em 20 anos desde a primeira aplicação, exposto na Figura 1, mostra que apenas os anos iniciais do ensino fundamental estão próximos do nível esperado (proficiente). Já o ensino médio encontra-se numa situação crítica, pois há uma tendência decrescente nas proficiências médias.

Figura 1: Evolução das proficiências médias em matemática do Brasil na Prova Brasil (1995 a 2015).

Fonte: Diretoria de Avaliação da Educação Básica – DAEB/INEP⁷.

As avaliações estaduais sobre o ensino também não mostram bons resultados. Mesmo o Estado de São Paulo, responsável por 28,7%⁸ do PIB brasileiro, apresenta dificuldades para chegar aos níveis esperados. A Figura 2 apresenta as médias das proficiências dos alunos e as proficiências esperadas para os

⁷ Diretoria responsável pela divulgação dos resultados das avaliações públicas federais, disponíveis em: <http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/inep-apresenta-resultados-do-saeb-prova-brasil-2015/21206>. Acesso em: 11/03/2017.

⁸ Informação disponível na página do Estado de São Paulo: <<http://www.saopaulo.sp.gov.br/conhecasp>>. Acesso em: 11/03/2017.

níveis adequado e básico de cada ano avaliado pelo SARESP.

Figura 2: Médias das proficiências da rede estadual do Estado de São Paulo em relação ao nível adequado por ano/série avaliado – matemática - SARESP.

Fonte: Baseado em Sumário Executivo – SARESP 2014 (SÃO PAULO, 2015b).

A troca de metodologia nas avaliações externas alterou também o objetivo em si da avaliação. Se com a TCT as avaliações tinham como objetivo avaliar a aquisição de conteúdos específicos de determinadas áreas do conhecimento, os sistemas de avaliação que usam a TRI têm como objetivo avaliar o desenvolvimento de habilidades e competências, sendo estas relacionadas com a aplicação do conhecimento em determinado nível de aprendizagem. E isto impacta diretamente o ensino na sala de aula, que deve deixar de ser “conteudista” para desenvolver habilidades e competências por meio dos conteúdos disciplinares (PERRY, 2009).

Segundo Barretto (2001b, apud PERRY, 2009, p.17) os sistemas de avaliação utilizam metodologias sofisticadas, como a TRI, apoiando-se em referenciais distintos dos utilizados pelos professores, o que pode impactar na compreensão da metodologia e contribuir para resistência ao uso dos resultados. A partir do comentário acima e com o intuito de verificar a posição de professores sobre sua

compreensão de alguns aspectos técnicos da TRI, como por exemplo a *construção da curva de probabilidade de acerto de um item*, foi realizada uma pesquisa pelo autor deste trabalho, em uma rede de ensino privada de São Paulo, cujos resultados estão no Apêndice A e o questionário disponibilizado no Anexo A. Os resultados obtidos sugerem que a elaboração de um trabalho com vistas a uma “simplificação da teoria” pode representar um auxílio ao professor para entender melhor a construção da Teoria da Resposta ao Item. Esta “simplificação da TRI” se tornou, então, o objetivo principal deste trabalho.

Este melhor entendimento sobre a metodologia é importante, uma vez que as escolas, representadas por seus dirigentes e professores, são cobradas para atingirem metas, estipuladas por órgãos superiores, fundamentadas na TRI. Por este motivo foi desenvolvido este texto que visa fazer a *transposição da TRI*⁹, tornando mais clara esta metodologia, explorando principalmente a construção da curva de probabilidade de acerto do item sob uma ***abordagem pedagógica***¹⁰, para que o professor se familiarize com os resultados e possa aprimorar os planos de ação voltados para o desenvolvimento das habilidades e competências dos alunos. A literatura é vasta na descrição de modelos teóricos, mas o que se pretende aqui é desenvolver a construção do pensamento sobre a metodologia.

A fim de facilitar esta compreensão, o texto terá como eixo de apoio a avaliação do SARESP e a disciplina de matemática, podendo ser estendida para outras disciplinas, em qualquer avaliação externa em larga escala que se utiliza da TRI. Não se pretende discutir aqui se a TRI é a melhor metodologia avaliativa na área educacional e nem tão pouco fornecer ferramentas para usar a TRI em sala de aula (pois não é o caso). Parte-se do ponto que a TRI é a metodologia mais utilizada nas avaliações em larga escala e que, a partir de seus resultados,

⁹ Usou-se a expressão *transposição da TRI* com o sentido de levar a conhecimento a Teoria da Resposta ao Item, por meio de uma abordagem pedagógica.

¹⁰ Esta expressão foi baseada em Hambleton et al. (1991) que apresenta procedimentos não usuais para estimar os parâmetros da TRI, mas que, quando utilizados para fins pedagógicos, facilitam a compreensão do tema.

decisões são tomadas em prol da melhoria da qualidade do ensino, como por exemplo, a elaboração de políticas educacionais.

Com a finalidade de concentrar em um único material os principais conceitos e denominações utilizados, este texto se desenvolverá por meio da apresentação das etapas do processo de implementação de uma avaliação externa em larga escala, desde a construção do item até a apuração e divulgação dos resultados, com enfoque para a construção do modelo da TRI de um parâmetro (dificuldade do item), também conhecido como modelo de Rasch. Os demais modelos serão apresentados a partir das considerações realizadas na abordagem que explicita o pensamento da construção do modelo de um parâmetro.

A fim de apresentar a TRI sob uma abordagem mais didática, a dissertação está estruturada em capítulos, da seguinte forma:

- **Capítulo 1 - Introdução à Avaliação Externa em Larga Escala** explora o conceito, objetivos e foco desse tipo de avaliação, além de trazer um breve histórico no Brasil;
- **Capítulo 2 - Processo de Implementação de uma Avaliação Externa em Larga Escala** apresenta as etapas desse processo, destacando-se algumas delas, tais como: a determinação da população avaliada, matriz de referência avaliativa, elaboração dos itens da prova, calibração dos itens, montagem dos cadernos de provas, aplicação e constituição da base de dados;
- **Capítulo 3 – Breve Introdução à Teoria da Resposta ao Item (TRI)** traz um breve histórico sobre a TRI, apresentando suas hipóteses básicas, a suposição do conhecimento cumulativo e definições acerca de habilidade e item;
- **Capítulo 4 - Abordagem Pedagógica sobre a Construção dos Modelos da TRI** apresenta os princípios fundamentais da TRI e explora, a partir uma sequência de passos, utilizando uma abordagem pedagógica, a construção do gráfico característico da curva da probabilidade de acerto do item. Além disso, também apresenta o ajuste da curva logística por meio dos três modelos mais utilizados: 1, 2 e 3 parâmetros;

- **Capítulo 5 - Cálculo da Habilidade de Um Respondente** mostra como se chega ao cálculo da habilidade do respondente, utilizando-se de um exemplo com poucos itens para facilitar a compreensão, e fazendo uma simples comparação dos resultados obtidos pela TRI com a TCT;
- **Capítulo 6 - Construção da Escala de Proficiência e Divulgação dos Resultados**, também por meio de uma abordagem pedagógica, mostra como é obtida a proficiência do aluno a partir da habilidade encontrada pela TRI, apresenta a construção da escala e sua interpretação pedagógica, e como é feito para garantir a comparabilidade dos resultados com avaliações anteriores e dentre os anos avaliados. Apresenta também como a entidade responsável pelo SARESP divulga os resultados da avaliação.

As **Considerações Finais** trazem um fechamento do trabalho e a possibilidade futura de transformar este trabalho em um “guia introdutório” para um primeiro contato com a metodologia.

Capítulo 1

Introdução à Avaliação Externa em Larga Escala

Conforme o Guia de Elaboração de Itens do CAEd (JUIZ DE FORA, 2008), nos dias de hoje existem dois importantes processos complementares de avaliação:

- Avaliação Interna: voltada para o desenvolvimento dos processos de ensino e aprendizagem, e é realizada pelo professor. Esta avaliação pode ser do tipo diagnóstica ou formativa em que o professor, baseado nos trabalhos realizados em sala de aula, procura identificar o que os alunos aprenderam;
- Avaliação Externa: avalia o desempenho de um conjunto de estudantes, muitas vezes agrupados por escola ou por sistemas. É dada essa denominação porque esta avaliação é planejada, elaborada, corrigida e tem seus resultados apurados fora da escola. Diferentemente da avaliação interna, a avaliação externa geralmente possui como foco o sistema de ensino e não a aprendizagem individual do aluno, pois procura verificar as habilidades e competências dos alunos pré-determinadas para o ano/série avaliado.

Desta forma, uma avaliação externa em larga escala, pode ser compreendida como uma avaliação do ensino aplicada em grandes quantidades de estudantes, realizada por entes externos à escola. Ainda segundo o Guia do CAEd (JUIZ DE FORA, 2008), esse tipo de avaliação apresenta três objetivos básicos:

- Definição de subsídios para a formulação de políticas educacionais;
- Acompanhamento ao longo do tempo da qualidade da educação;
- Produção de informações capazes de desenvolver relações significativas entre as unidades escolares e órgãos centrais ou distritais de secretarias, bem como iniciativas dentro das escolas.

De acordo com Pimenta (2012), as avaliações externas começaram a ser desenvolvidas no Brasil na década de 1980, culminando na implantação de um sistema de avaliação da educação básica brasileira em 1991, conhecido como SAEB (Sistema Nacional de Avaliação da Educação Básica), propondo-se a fornecer dados sobre a qualidade dos sistemas educacionais do Brasil como um todo, das regiões geográficas e das unidades federadas (estados e Distrito Federal). Dentre os seus objetivos, pode-se destacar a identificação de problemas e as diferenças regionais do ensino e a apresentação à sociedade dos resultados dos processos de ensino e aprendizagem. Porém, como há uma quantidade grande de alunos matriculados nas escolas brasileiras, o SAEB é aplicado em uma amostra deste universo.

Em 1995, o SAEB passou por uma reestruturação metodológica¹¹, deixando de mensurar seus resultados pela TCT (Teoria Clássica dos Testes) e adotando a TRI (Teoria da Resposta ao Item), o que possibilitou a comparação dos resultados ao longo dos anos.

Visando complementar as avaliações realizadas em âmbito nacional, durante a década de 1990 alguns Estados brasileiros criaram os seus próprios sistemas avaliativos (PIMENTA, 2012). Foi neste contexto que o Estado de São Paulo em 1996 implementou o SARESP (Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo), sendo obrigatório à rede estadual e aberto para a rede municipal e escolas privadas, onde na Seção 2.1 serão dadas mais informações sobre esta avaliação.

Segundo o documento básico de referências avaliativas do SARESP (São Paulo, 2009), na educação, é fundamental definir uma matriz de referência em situações de aprendizagem e ensino, para que seja possível verificar o aprendizado do aluno em relação às expectativas de aprendizagem estabelecidas para o seu ano/série escolar, visto que os dados e informações gerados pela avaliação, tanto qualitativos quanto quantitativos, possibilitam a realização de uma análise sobre a qualidade do ensino e a tomada de decisão.

¹¹ Informação extraída no site do IBGE (<http://ces.ibge.gov.br/base-dados/metadados/inep/sistema-nacional-de-avaliacao-da-educacao-basica-saeb.html>) e do INEP (<http://portal.inep.gov.br/web/saeb/historico>). Ambos acessados em 12/06/2016.

Dependendo da estratificação dos resultados e da abrangência das conclusões obtidas na análise, o foco pode ser a turma, o ano escolar do ensino ou até mesmo a escola.

A análise dos resultados da avaliação externa em larga escala deve ser realizada no mínimo por dois agentes: a escola e o órgão responsável pelo sistema escolar avaliado. Este ato permite que a escola realize um diagnóstico do seu ensino e elabore um plano de ações educativas com foco em suas necessidades. Já ao órgão responsável, isto contribui para uma tomada de decisão mais assertiva e direcionada às metas desejadas para a melhoria da qualidade do sistema de ensino.

Segundo Klein e Fontanive (1995), o objetivo de uma avaliação externa em larga escala é informar o que um conjunto de alunos em diferentes séries sabem e são capazes de fazer, em um determinado momento, e acompanhar sua evolução ao longo dos anos. Os mesmos autores também colocam que se o objetivo fosse avaliar cada aluno, as avaliações externas em larga escala teriam de ter um grande número de itens, algo superior a 100, para cada disciplina avaliada, de modo a cobrir todos os tópicos do programa escolar. Mas esta situação é praticamente inviável, não só devido ao tempo para o aluno responder, mas também à operacionalização e ao custo.

Para que o aluno responda um número factível de itens, adequado ao tempo de prova e sem prejudicar os objetivos da avaliação externa em larga escala, aplica-se uma metodologia que agrupa em blocos os itens que cobrem o programa escolar. Estes blocos são combinados em diferentes arranjos e são estes arranjos que compõem os cadernos de provas. Apesar de um aluno responder uma parte dos itens, cada item é respondido por boa parte dos alunos e, dessa forma, a metodologia embasa a validade dos resultados da avaliação do sistema educacional em relação ao programa escolar. Esta metodologia associada à montagem do caderno de provas é chamada de Blocos Incompletos Balanceados, também conhecida como BIB. Informações adicionais, como referências e um exemplo sobre esta metodologia, poderão ser encontradas na Seção 2.6 e no Apêndice B.

Após aplicação das provas e apuração dos resultados, são elaborados documentos¹² para divulgação do desempenho das escolas. No SARESP, por exemplo, é divulgado um boletim por Escola, sumário executivo e relatórios pedagógicos das disciplinas avaliadas. Em geral, estes documentos trazem tradicionalmente dois indicadores sobre o desempenho dos alunos para efeitos de análises: o primeiro é a média e o segundo é o percentual de estudantes em cada nível definido na escala de pontuação. A média é uma maneira de sintetizar o resultado da turma, da escola, do Município, e etc. Há outras maneiras de sintetizar os resultados dos alunos, não exploradas nesses documentos de divulgação, como por exemplo, a mediana¹³. Já o percentual de estudantes por níveis da escala de pontuação fornece informações sobre a quantidade de alunos que, dentre o total avaliado, já consolidaram (ou não) as habilidades esperadas. Não será discutida neste texto a escolha dessas medidas e nem se são boas para comparação, pois não é o objetivo aqui.

¹² Estes documentos podem ser encontrados no site da Fundação para o Desenvolvimento da Educação (FDE) de São Paulo. Site: <http://saresp.fde.sp.gov.br/2015/#> (acessado em 10/07/2016). Para acessar anos anteriores, trocar 2015 pelo ano desejado.

¹³ Valor que ocupa a posição central dos dados ordenados (MAGALHÃES & LIMA, 2013).

Capítulo 2

Processo de Implementação de uma Avaliação Externa em Larga Escala

Na Figura 2.1, encontra-se o processo de implementação da avaliação externa em larga escala, baseado no esquema do Guia de Elaboração de Itens do CAEd (JUIZ DE FORA, 2008). Cada retângulo representa uma etapa desse processo, que nada mais é do que uma atividade a ser desenvolvida. Este esquema ajudará na compreensão das diferentes etapas, desde a concepção da avaliação até a elaboração do relatório sobre os resultados. Cada uma das atividades do processo de implementação disposta na Figura 2.1 será apresentada a seguir, com exceção das quatro últimas, que por estarem diretamente ligadas ao tema deste trabalho, serão tratadas nos capítulos à frente da seguinte forma:

- Obtenção dos resultados, utilizando a TRI e Análise dos resultados: Capítulo 5;
- Construção e interpretação da escala de proficiência e Elaboração dos relatórios gerais e pedagógicos: Capítulo 6.

Figura 2.1: Processo de implementação da avaliação externa em larga escala.

Fonte: Baseado no esquema do guia de elaboração de itens do CAEd (JUIZ DE FORA, 2008).

2.1. População a Ser Avaliada

É de extrema importância a definição do público alvo para que o mesmo seja levado em conta na elaboração dos itens da avaliação. Vale ressaltar que não está se falando em levar em conta o perfil do público respondente (socioeconômico, regional, ensino público ou privado, etc.), mas sim o público

esperado para cada ano avaliado da educação básica. Em seguida, é necessário decidir se a avaliação será censitária (abarcando toda a população alvo) ou por amostragem (seleção de parte da população). Caso seja por amostragem, existe a necessidade de elaborar um plano amostral para aplicação da avaliação. Tal plano requer estudo especialmente realizado com esse intuito, e não será abordado neste trabalho.

A título de exemplo, abaixo seguem algumas avaliações externas em larga escala com o respectivo público alvo definido bem como sua periodicidade:

- SARESP: o Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo possui como público alvo os alunos matriculados no 3º, 5º, 7º e 9º anos do ensino fundamental e no 3º ano do ensino médio do Estado, de todas as escolas das redes municipais e estadual, da rede de ensino do Centro Estadual de Educação Tecnológica “Paula Souza” - CEETEPS, da rede de ensino do Serviço Social da Indústria - SESI, e outras escolas estaduais não administradas pela Secretaria da Educação, bem como as escolas particulares que aderirem à avaliação. De acordo com a resolução SE-49, de 1º-9-2016, em 2016 o SARESP avaliou de forma **censitária o 3º, 5º e 9º anos do ensino fundamental e 3º do ensino médio, e por amostragem apenas o 7º ano do ensino fundamental**. As áreas curriculares avaliadas foram Língua Portuguesa e Matemática. Essa avaliação é **censitária** para todos os anos avaliados, com exceção do 7º ano do ensino fundamental cuja aplicação é por amostragem, e sua periodicidade é anual.
- SIMAVE: o Sistema Mineiro de Avaliação da Educação Pública possui como público alvo (desde 2015) os alunos matriculados no 3º, 5º, 7º e 9º anos do ensino fundamental e na 1ª e 3ª séries do ensino médio de todas as escolas das redes municipais e estadual do Estado de Minas Gerais. Esta avaliação é **censitária** e sua periodicidade é anual para o 3º ano do ensino fundamental; e bianual para os demais anos.
- ANEB: a Avaliação Nacional da Educação Básica possui como público alvo os alunos matriculados na 4ªsérie/5ºano e 8ªsérie/9ºano do Ensino Fundamental de todas as escolas (pública ou privada) do país. Esta

avaliação é **amostral** e sua periodicidade é bianual.

- ANA: a Avaliação Nacional da Alfabetização possui como público alvo os alunos matriculados na 2ªsérie/3ºano do Ensino Fundamental de todas as escolas das redes municipais, estaduais e federal. Esta avaliação é **amostral** e sua periodicidade é anual.
- ANRESC (Prova Brasil): a Avaliação Nacional do Rendimento Escolar possui como público alvo os alunos matriculados na 4ªsérie/5ºano e 8ªsérie/9ºano do Ensino Fundamental de todas as escolas das redes municipais, estaduais e federal. Esta avaliação é **censitária** e sua periodicidade é bianual.
- ENEM: o Exame Nacional do Ensino Médio possui como público alvo os alunos concluintes ou regularmente matriculados na 3ª série do Ensino Médio, em escolas públicas ou privadas, que vão se submeter a processo seletivo em Instituição de Ensino Superior (IES). Esta avaliação é **censitária**¹⁴ e sua periodicidade é anual.

É importante saber que existem avaliações externas voltadas para a avaliação do sistema de ensino (e não do aluno), como por exemplo, SARESP, SIMAVE, ANEB, ANA, e ANRESC, e outras avaliações em que o foco é realmente o aluno (com objetivo de classificar), como é o caso do ENEM.

Na pesquisa realizada com professores (Apêndice A), o que chamou a atenção na leitura dos comentários foi o SARESP ser mais citado pelos professores do ensino fundamental e o ENEM pelos professores do ensino médio. É claro que isto se deve ao vestibular, pois o ENEM é a principal via de acesso às instituições públicas de ensino superior e às vagas do PROUNI nas instituições privadas. Isto sugere que, para não haver confusão, é preciso deixar claro o foco da avaliação externa que se tratará ao falar da TRI, que no caso deste texto será o SARESP.

Em geral, as avaliações externas são utilizadas para aferir a qualidade do ensino, mediante avaliação periódica pré-determinada, em que os desempenhos dos

¹⁴ Apesar de ser censitária, a avaliação não é obrigatória e sua inscrição é por adesão.

alunos da educação básica em diferentes componentes curriculares são medidos e sua evolução pode ser acompanhada ao longo do tempo.

2.2. Matriz de Referência

Segundo o Guia do CAEd (JUIZ DE FORA, 2008), as provas das avaliações externas em larga escala na área educacional buscam aferir, isto é, mensurar de certa forma a proficiência dos estudantes nas determinadas áreas de conhecimento, em anos específicos do ensino fundamental e/ou médio. É por meio destes resultados que se verificará se o objetivo da avaliação externa em larga escala está sendo alcançado, em relação à qualidade do ensino, pois o desempenho na avaliação relaciona os conteúdos ensinados pelos professores e o nível de aprendizado dos alunos (domínio dos alunos sobre as competências e habilidades avaliadas).

Para que sejam elaborados os itens da avaliação, antes é necessário definir e relacionar as habilidades e competências para cada área de conhecimento a ser avaliada. São essas habilidades e competências que compõem a Matriz de Referência. Uma vez definida a matriz, ela norteará a elaboração de itens que irão compor o caderno de prova da avaliação externa em larga escala.

Geralmente, as Matrizes de Referência são resultados do estudo de Parâmetros Curriculares, Diretrizes Curriculares e da reflexão realizada por especialistas em educação sobre as habilidades consideradas essenciais em cada etapa do ensino avaliado. A matriz contempla implicitamente o conteúdo curricular e o nível de operação mental, necessários na verificação da habilidade avaliada (JUIZ DE FORA, 2008).

Como o SARESP é uma avaliação externa em larga escala, foco desse material, a partir deste ponto os exemplos didáticos terão como referência essa avaliação, em especial a disciplina de matemática.

2.2.1. Matriz de Referência do SARESP

Apesar da Matriz de Referência do SARESP ter sido elaborada por especialistas em avaliação, ela passou pelo olhar e recebeu críticas e sugestões dos autores da Proposta Curricular do Estado de São Paulo, de especialistas da Coordenadoria de Estudos e Normas Pedagógicas (CENP) da SEE/SP e por professores coordenadores das áreas envolvidas na avaliação. Ou seja, a Matriz de Referência passa por várias “mãos” até chegar à sua versão final (SÃO PAULO, 2009).

A principal finalidade da matriz de referência é o seu poder de sinalização das estruturas básicas de conhecimentos a serem construídas em cada etapa da escolaridade básica (SÃO PAULO, 2009). Para uma compreensão das relações estabelecidas entre os três fundamentais aspectos da matriz (habilidades, competências e conteúdos), uma vez que estes aspectos estarão representados na Matriz de Competências do SARESP, na Figura 2.2 é apresentado um esquema triangular, retirado do documento Matrizes de Referência para Avaliação SARESP (SÃO PAULO, 2009), que expressa essas relações nos níveis de desempenho da escala de proficiência nas disciplinas avaliadas.

Figura 2.2: Relações entre habilidades, competências e conteúdos.

Fonte: Matrizes de referência para avaliação SARESP: Documento básico (SÃO PAULO, 2009).

A Figura 2.2 pode ser interpretada da seguinte maneira:

- Vértices: são os pontos valorizados na Matriz de Referência. Segue uma descrição de cada um desses pontos:
 - Habilidades: “saber-fazer”, relacionadas à prática do trabalho mental. São elas que descrevem as estruturas mais gerais da inteligência e que possibilitam saber do que o aluno precisa para responder o item na avaliação;
 - Conteúdos: são os conteúdos do currículo considerados pela matriz. Esses conteúdos estão necessariamente relacionados com as habilidades explicitadas pela matriz;
 - Competências (Cognitivas): compostas por habilidades, são um conjunto de operações mentais que o aluno realiza para estabelecer as relações com e/ou entre os objetos, situações, fenômenos, pessoas, etc.
- Eixo Habilidades-Conteúdos: verificação de conteúdos disciplinares mediada pelo uso de habilidades. A ligação entre estes aspectos é realizada por meio da Escala de Proficiência. É esta escala que sintetiza o domínio dos conteúdos e as habilidades alcançadas pelo estudante e que permitirá inferir sobre o nível de domínio das competências avaliadas.
- Eixo Conteúdos-Competências: a avaliação de competências reflete o compromisso assumido no currículo, pois a missão da escola estadual é caracterizada como:

“...um lugar e um tempo em que competências fundamentais ao conhecimento humano são aprendidas e valorizadas”. (SÃO PAULO, 2009, p.12)

À medida que os alunos dominam competências, isto garante sua aprendizagem efetiva. Este eixo é interligado pelos níveis de desempenho, pois é por meio deles que se verifica o que os professores estão ensinando e o quê os alunos estão aprendendo.

- Eixo Habilidades-Competências: neste eixo, os conteúdos disciplinares são o elo de ligação entre as habilidades e as competências. É por meio

destes conteúdos disciplinares que o aluno desenvolverá as habilidades e competências para resolver os itens das avaliações externas em larga escala.

É relevante destacar que a Matriz de Referência representa um recorte dos conteúdos curriculares, contemplando algumas competências e habilidades a eles associadas. Ela abrange de forma generalista os principais fundamentos das disciplinas avaliadas e as principais competências esperadas dos alunos, descrevendo de forma mais geral as estruturas da inteligência que, quando bem avaliadas, evidenciam o quadro real do efetivo desenvolvimento dos alunos (SÃO PAULO, 2009).

A Matriz de Referência do SARESP, organizada por ano da Educação Básica, foi construída e deve ser interpretada levando em consideração a hipótese de que o conhecimento é cumulativo (TODD & WOLPIN, 2003 e 2007). Esta hipótese é de suma importância na interpretação dos modelos da TRI, sendo apresentado da seguinte forma no documento básico do SARESP:

“... os alunos que estão situados em um determinado nível dominam não só as habilidades associadas a esse nível, mas também as proficiências descritas nos níveis anteriores – a lógica é a de que quanto mais o estudante caminha ao longo da escala mais habilidades ele terá desenvolvido. ” (SÃO PAULO, 2014b, p.5)

Com o objetivo de analisar e entender a estrutura de uma Matriz de Referência, na Figura 2.3 será apresentada a matriz da disciplina de matemática para o 7º Ano do Ensino Fundamental.

Figura 2.3: Matriz de referência do SARESP da disciplina de matemática e suas tecnologias - 7º ano do ensino fundamental.

COMPETÊNCIAS DO SUJEITO			
	GRUPO I Competências para observar	GRUPO II Competências para realizar	GRUPO III Competências para compreender
OBJETOS DO CONHECIMENTO (CONTEÚDOS) Tema 1 – Números, operações, funções, iniciação à Álgebra	H01 Reconhecer as principais características do sistema decimal: contagem, base, valor posicional.	H05 Fazer cálculos que envolvam adições e subtrações de frações.	H02 Estabelecer relações entre números naturais tais como “ser múltiplo de”, “ser divisor de” e reconhecer números primos e números compostos.
	H04 Representar medidas não inteiras utilizando frações.	H07 Fazer cálculos que envolvam adições e subtrações de números decimais.	H03 Resolver problemas que envolvam as quatro operações básicas entre números inteiros (adição, subtração, multiplicação e divisão).
	H06 Representar quantidades não inteiras utilizando notação decimal.	H09 Efetuar cálculos com potências.	H13 Aplicar uma ordem de operações ao resolver problemas (parênteses, multiplicação, divisão, adição e subtração).
	H08 Compreender a relação entre as representações fracionária e decimal de um número.	H10 Efetuar cálculos com multiplicação e divisão de números decimais.	H15 Expressar e resolver problemas por meio de equações.
Tema 2 – Espaço e forma		H11 Efetuar cálculos com adição, subtração, multiplicação e divisão com negativos.	
		H12 Ler e escrever expressões algébricas correspondentes a textos matemáticos escritos em linguagem corrente e, vice-versa.	
		H14 Resolver equações do 1º grau.	
	H16 Identificar formas planas e espaciais em situações do cotidiano e por meio de suas representações em desenhos e em malhas.	H17 Classificar formas planas e espaciais.	
Tema 3 – Grandezas e medidas / Proporcionalidade	H18 Identificar figuras espaciais a partir de suas planificações.	H19 Determinar área e perímetro de uma figura utilizando composição e decomposição de figuras.	
	H20 Identificar simetria axial e de rotação na leitura das representações dos objetos no dia a dia e das figuras geométricas.	H21 Identificar elementos e classificar poliedros.	
	H26 Identificar a soma das medidas dos ângulos de um triângulo (180°) e de um polígono de n lados (por decomposição em triângulos).	H22 Realizar medidas usando padrões e unidades não convencionais ou de outros sistemas de medida dados.	H27 Resolver problemas que envolvam medidas de ângulos de triângulos e de polígonos em geral.
		H23 Aplicar as principais características do sistema métrico decimal: unidades, transformações e medidas.	H29 Resolver situações-problema que envolvam grandezas direta ou inversamente proporcionais.
		H24 Reconhecer ângulos como mudança de direção ou giros, identificando ângulos retos e não retos.	H32 Usar desenhos de escalas para resolver problemas do cotidiano que incluam distância (como em leitura de mapas).
		H25 Efetuar cálculos que envolvam medidas de ângulos.	
		H28 Reconhecer situações que envolvam proporcionalidade.	
		H30 Reconhecer o conceito de razão em diversos contextos: proporcionalidade, escala, velocidade, porcentagem etc.	
		H31 Reconhecer pi como uma razão constante da geometria.	

COMPETÊNCIAS DO SUJEITO			
	GRUPO I	GRUPO II	GRUPO III
	Competências para observar	Competências para realizar	Competências para compreender
OBJETOS DO CONHECIMENTO (CONTEÚDOS) Tema 4 – Tratamento da informação / Probabilidade / Estatística	H36 Identificar o gráfico adequado para representar um conjunto de dados e informações. (gráficos elementares – barras, linhas, pontos).	H33 Resolver problemas que envolvam probabilidade de eventos simples.	
		H34 Identificar e interpretar informações transmitidas por meio de tabelas.	
		H35 Identificar e interpretar informações transmitidas por meio de gráficos.	
		H37 Utilizar diagramas de árvore para resolver problemas simples de contagem.	
		H38 Resolver problemas que envolvam a ideia do princípio multiplicativo de contagem.	

Fonte: Matrizes de referência para avaliação SARESP: documento básico (SÃO PAULO, 2009).

As habilidades na Matriz de Referência do SARESP estão organizadas por tema do conteúdo (objeto do conhecimento) e por grupo de competências (compostos por habilidades). O Leitor que desejar se aprofundar na compreensão dessa organização, poderá fazê-lo consultando o documento básico da Matriz de Referência do SARESP (SÃO PAULO, 2009).

2.3. Elaboração de Itens

A Matriz de Referência é o ponto inicial e fundamental na construção de itens (JUIZ DE FORA, 2008). É a matriz que orienta a elaboração de itens a fim de que estes tenham a qualidade esperada. Todos os itens elaborados passam por diversas etapas que envolvem vários profissionais da educação. Na Figura 2.4, é apresentado um fluxograma do processo de elaboração de itens, adaptado do esquema do Guia do CAEd (JUIZ DE FORA, 2008). Este fluxo pode ser aplicado à construção de item para qualquer disciplina.

Antes de elaborar um item deve-se compreender as habilidades e/ou competências esperadas, sendo essas associadas aos conteúdos curriculares, e o quê, exatamente, elas pretendem avaliar. Geralmente, os itens elaborados para uma avaliação em larga escala são de múltipla escolha, e possuem apenas uma alternativa correta.

Figura 2.4: Processo de elaboração do item.

Fonte: Adaptado do esquema do guia de elaboração de itens do CAEd (UFJF, 2008).

No processo apresentado pela Figura 2.4, existem duas etapas sob a responsabilidade da entidade organizadora que devem ser destacadas para melhor compreensão do processo:

1. “*Tratar os Itens do Pré-Teste Estatisticamente*”: nesta etapa acontece o que se chama de calibração dos itens, isto é, a estimação dos parâmetros do modelo frente às respostas dos alunos que compuseram a amostra, que será mais explorado na Seção 4.3. Aqui também se verifica a qualidade do item, ou seja, se ele é bom ou ruim para diferenciar os alunos, se houve item que todo mundo acertou/errou, analisando se o problema está na elaboração do item, na dificuldade/facilidade, se houve problema com a amostra de respondentes, se as alternativas (chamadas descritores) estão realmente bem elaboradas, etc.
2. “*Selecionar os Itens que Irão Compôr a Prova*”: geralmente, o pré-teste é realizado com um número de itens acima da quantidade necessária de uma prova, pois, se ele for realizado com a quantidade exata de itens que irão compor a prova corre-se o risco desta avaliação ser aplicada com itens a menos do que se esperava. Ou aumentar o custo para pré-testar mais itens para completar o que falta. Depois que os itens foram pré-testados, eles costumam ir para o que se chamam de banco de itens e é a partir deste banco que se faz a seleção para a prova da avaliação.

Para fins didáticos e a título de exemplo de aplicação do processo de elaboração de itens apresentado, foi adaptado um item do *SARESP 2013* para identificar os elementos importantes na sua elaboração:

- a) **Escolhe-se uma habilidade** para desenvolver o item, a partir da Matriz de Referência. Por exemplo: a partir da Matriz de Referência do SARESP, na disciplina de Matemática e suas Tecnologias para o 7º Ano do Ensino Fundamental, apresentada na Figura 2.3, foi selecionada a habilidade H08:

“Compreender a relação entre as representações fracionária e decimal de um número”

Esta habilidade refere-se ao Tema 1 (Números, Operações, Funções, Iniciação a Álgebra) e pertence ao Grupo I (Competências para Observar). Portanto, este item deve explorar de forma representativa a relação entre a fração e o número decimal.

- b) Com base na habilidade selecionada, deve-se **construir o enunciado**. Segundo o Guia (JUIZ DE FORA, 2008), o enunciado é um estímulo para que o aluno mobilize recursos cognitivos para buscar uma solução ao problema apresentado e responder ao que é solicitado pelo comando da resposta. Um exemplo de enunciado criado com base na H08 é:

“Existem várias maneiras de se representar três décimos. Em forma de fração fica $\frac{3}{10}$, em porcentagem fica 30%.”

- c) Em conjunto com a construção do item, deve-se **optar pelo auxílio ou não de um suporte**. Suporte é tudo aquilo considerado no item que tenha o objetivo de facilitar a contextualização (JUIZ DE FORA, 2008). Por exemplo: textos, tabelas, gráficos, calendários, placas de sinalização, rótulos de embalagens, bulas de remédio, mapas, maquetes, fotos, etc.

Este item não teve o auxílio de suporte.

- d) Em seguida, deve-se **elaborar o comando para resposta**. Isto é, dar o comando para que o item seja respondido. Nesse exemplo, o comando de resposta é complementar, solicitando ao aluno completar a frase com o formato decimal da fração:

“Na forma decimal fica: ”

- e) Por último, deve-se **construir as alternativas de respostas**, mas elas não devem ser construídas de qualquer forma. Na construção das alternativas, duas coisas devem ser levadas em consideração: o gabarito e os distratores. O Gabarito é a resposta correta do item, que deve ser única e retratar a capacidade do aluno em relação à habilidade cognitiva avaliada. Já os distratores, são as demais alternativas (incorretas). Eles devem produzir informações importantes, apontando possíveis formas de raciocínio dos alunos. Por exemplo: os distratores podem conter soluções

com erros que os estudantes costumam cometer e oferecerão informações sobre as dificuldades encontradas.

Distrator: (A) 0,1.

Distrator: (B) 0,003.

Distrator: (C) 3,0.

Gabarito: (D) 0,3.

Após a execução dos passos acima, o item completo é apresentado da seguinte forma:

Existem várias maneiras de se representar três décimos. Em forma de fração fica $\frac{3}{10}$, em porcentagem fica 30%. Na forma decimal fica:

(A) 0,1.

(B) 0,003.

(C) 3,0.

(D) 0,3.

Fonte: Baseado em item do SARESP 2013.

Apenas para exemplificação, a seguir é apresentado um item que faz o uso de suporte como auxílio. A habilidade avaliada no item é a H35 – Identificar e interpretar informações transmitidas por meio de gráfico. Refere-se ao Tema 4 (Tratamento da Informação / Probabilidade / Estatística) e pertence ao Grupo II (Competências para Realizar - esquemas procedimentais).

Uma empresa possui 50 funcionários, os quais se distribuem da seguinte forma com relação ao grau de escolaridade:

Observando o gráfico, é correto afirmar que o número de funcionários do ensino médio é:

Distrator: (A) a metade do ensino fundamental.

Distrator: (B) a metade do ensino superior.

Gabarito: (C) o dobro do ensino fundamental.

Distrator: (D) o dobro do ensino superior.

Fonte: SARESP 2013.

Caso o Leitor deseje conhecer uma elaboração de item de forma mais detalhada, o Guia de Elaboração de Itens do CAEd (JUIZ DE FORA, 2008) exibe um passo a passo dessa construção.

2.4. Pré-Testagem ou Calibração do Item

Após a elaboração, validação de especialistas da área da educação e revisão da norma culta da língua portuguesa, os itens são pré-testados, ou seja, respondidos por uma amostra de alunos que possuem características similares ao público alvo. Este processo também é conhecido como calibração do item, e, dependendo do modelo a ser ajustado, estima os parâmetros característicos

desse item. Estes parâmetros permitem estimar a probabilidade de acerto do respondente no item, assegurando que alunos com conhecimentos distintos tenham graus diferentes de acerto.

Os modelos disponíveis para a TRI, ou seja, as equações matemáticas que descrevem o desempenho dos alunos em termos de parâmetros, podem ter: um parâmetro (dificuldade do item), ou dois parâmetros (dificuldade do item e discriminação do item) ou ainda três parâmetros (dificuldade do item, discriminação do item e acerto ao acaso). Como já dito na Introdução, este trabalho vai concentrar sua análise inicial no modelo de um parâmetro (dificuldade do item) para desenvolver um passo a passo que explicita a filosofia de construção do modelo. As seções 4.3.2 e 4.3.3 apresentarão os demais modelos a partir das considerações realizadas na abordagem da construção do modelo de um parâmetro.

2.5. Análise Estatística e Pedagógica dos Itens

Uma vez que os itens foram respondidos, eles passam por um processo de análises, inclusive pedagógicas, e critérios estatísticos a fim de verificar se são itens consistentes no sentido de avaliar e mensurar o desempenho do aluno. Somente depois da aprovação desses itens, é que eles irão compor um “banco de itens” de onde eles poderão ser selecionados para compor a prova da avaliação em larga escala. Para saber mais detalhes sobre a construção e organização de um banco, ver Andriola (1998) que organizou um banco de itens destinados à avaliação do raciocínio verbal, utilizando a Teoria da Resposta ao Item (TRI).

2.6. Montagem dos Cadernos de Provas

Todos os itens pré-testados e aprovados passam a fazer parte do chamado *Banco de Itens*. É a partir desse banco que as provas da avaliação em larga escala serão compostas.

Antes de iniciar a montagem dos cadernos de provas, o Leitor será familiarizado sobre como os itens são posicionados, organizados, em um banco de dados. Na etapa “*Tratar os Itens do Pré-Teste Estatisticamente*” (Figura 2.4), após o pré-teste ter sido realizado em alunos com características similares ao público alvo, são obtidos alguns dados sobre cada item. Um destes dados é a dificuldade que será utilizada para ordenar os itens de forma crescente no banco. O Leitor não deve se preocupar, neste ponto, de como se obteve a curva e a dificuldade de cada item, pois, isto será tratado na Seção 4.3. Diante disso, considere os itens A, B e C, três itens de matemática do 7º ano do ensino fundamental, cujas curvas da probabilidade de acerto são exibidas por meio da Figura 2.5.

Figura 2.5: Curvas da probabilidade de acerto dos itens A, B e C, segundo as proficiências dos alunos do 7º ano do ensino fundamental em matemática.

A dificuldade dos itens A, B e C é, respectivamente, 232, 258 e 269. O posicionamento desses três itens em um banco é exemplificado pela Figura 2.6. Em resumo, dentre esses três itens o mais fácil é o A e o mais difícil é o C.

Figura 2.6: Posicionamento dos itens A, B e C, em um banco de itens, de acordo com a respectiva dificuldade.

Como existem diversas avaliações no país, a montagem dos cadernos de provas pode variar de acordo com o foco da avaliação. Por exemplo, o Exame Nacional do Ensino Médio (ENEM) tem como foco de avaliação o aluno, identificando em qual(is) área(s) do conhecimento ele precisa reforçar o seu grau de desenvolvimento. Em avaliações com o mesmo foco do ENEM, os itens são selecionados para compor uma única prova, pois os candidatos devem responder aos mesmos itens. Mas, se todos os candidatos respondem aos mesmos itens, por que os cadernos de provas do ENEM possuem cores diferentes? As cores utilizadas nos cadernos de provas do ENEM servem para diferenciá-los quanto à ordem em que os itens aparecem, reduzindo a influência da disposição dos itens no desempenho do candidato, e também por questões de segurança, a fim de evitar fraudes entre candidatos. Já o SARESP possui como foco a avaliação do sistema de ensino e, nesse sentido, não há necessidade de todos os alunos responderem aos mesmos itens. Até porque, a seleção de itens contempla todas as habilidades da Matriz de Referência da educação básica e isto exige um grande número de itens a ser respondido, o que inviabilizaria a avaliação, conforme já dito, devido ao tempo para responder, operacionalização e custo.

As avaliações externas em larga escala com o mesmo foco do SARESP (avaliar o sistema de ensino de tal forma que cada aluno responda a um pequeno número de questões) possuem a desvantagem de não poder medir o desempenho individual dos alunos, ou seja, avaliar o aprendizado de cada um, pois cada um responde a uma prova diferente e a uma baixa proporção de itens em relação ao total que representa todas as habilidades da Matriz de Referência.

Para garantir que, mesmo os alunos respondendo a uma parte da quantidade total de itens, seja possível tirar conclusões acerca do sistema de ensino, como por exemplo, da disciplina de matemática e de cada um de seus tópicos (Números Inteiros, Álgebra, Geometria, Tratamento da Informação, etc.), algumas avaliações, como o SARESP, utilizam a metodologia BIB¹⁵ – Blocos Incompletos Balanceados. Segundo Bekman (2001), esta metodologia aplicada na montagem dos cadernos de provas pode ser descrita como:

“distribuir um certo número **b** de **blocos** de itens em diversos **cadernos c** de forma que cada caderno não contenha a totalidade dos blocos. Estes cadernos **incompletos** são submetidos aos alunos que recebem, individualmente, um subconjunto do total de blocos de itens que está sendo aplicado na sala de aula, ou seja, cada aluno recebe apenas uma fração f_u do total de blocos submetidos à sala de aula”. (BEKMAN, 2001, pg 121)

A metodologia BIB impõe ordem, número de vezes em que cada bloco aparece, quantidade de questões que serão respondidas e determina o número de cadernos de provas que serão respondidos. O Apêndice B traz um exemplo, utilizando-se de um caso simples, para ajudar na compreensão da metodologia BIB. Além disso, Bekman (2001) apresenta uma série de exemplos de aplicação do BIB que o Leitor interessado poderá consultar.

No SARESP, de acordo com o relatório pedagógico de matemática (SÃO PAULO, 2014b), para esta disciplina há um total de 104 itens que contemplam

¹⁵ Para maiores informações sobre a metodologia BIB consultar Bekman (2001).

as 38 habilidades da Matriz de Referência em cada etapa de escolaridade. A Figura 2.7 apresenta estes 104 itens agrupados em 13 blocos com 8 itens cada.

Figura 2.7: Agrupamento dos 104 itens em 13 blocos.

Aplicando o BIB para 13 blocos, obtém-se que devem ser confeccionados 26 cadernos de provas, em que cada um contém 24 questões de múltipla escolha (3 blocos dos 13), exemplificados no Quadro 2.1. Com essa metodologia, a avaliação é composta por diferentes cadernos de provas (que possuem itens em comum entre si).

Quadro 2.1: Exemplo da metodologia BIB para o SARESP.

Nº	Caderno	Ordem em que os BLOCOS aparecerão nos cadernos		
		1º	2º	3º
1	A	IV	XIII	VI
2	B	I	IX	XI
3	C	VII	III	XIII
4	D	IV	XI	I
5	E	III	VI	XI
6	F	VI	VIII	IV
7	G	II	X	VI
8	H	X	IV	II
9	I	V	VII	X
10	J	XIII	IX	I
11	K	IX	III	XII
12	L	XI	VI	III
13	M	V	XII	III
14	N	XII	X	VII
15	O	III	I	IX
16	P	IX	XIII	V
17	Q	XII	VII	II
18	R	VIII	II	XII
19	S	XI	VIII	IV
20	T	I	V	VIII
21	U	VI	I	X
22	V	VIII	IV	XIII
23	X	X	XI	V
24	W	XIII	V	VIII
25	Y	VII	II	IX
26	Z	II	XII	VII

Fonte: Baseado em exemplo de Bekman (2001).

Os 104 itens possuem diferentes graus de dificuldade e são agrupados pelo percentual de acertos, geralmente obtidos na etapa de calibração, conforme Quadro 2.2.

Quadro 2.2: Agrupamento dos itens de acordo com percentual de acertos.

Grupo de Dificuldade do Item	Percentual de Acertos Para Classificação no Grupo
Muito Fácil	86 a 100%
Fácil	66 a 85%
Médio	36 a 65%
Difícil	16 a 35%
Muito Difícil	0 a 15%

Fonte: Relatório pedagógico de matemática do SARESP (SÃO PAULO, 2014b).

Segundo Rabelo (2011), para permitir a comparabilidade com resultados de edições passadas, as provas de um ano incluem itens selecionados de aplicações anteriores, os chamados itens de ligação. Para que seja possível comparar os desempenhos, Andrade & Valle (1998) citam que pelo menos 6 itens devem ser comuns entre as provas. No caso do SARESP, geralmente 13 dos 104 itens são de ligação, o que dá em média um item de ligação por bloco (SÃO PAULO, 2014b).

O processo de tornar comparáveis os desempenhos dos respondentes, seja com edições anteriores ou dentro do ano avaliado ou entre os anos avaliados, é chamado de equalização. Devido à importância deste assunto, Andrade et al. (2000) dedicam um capítulo do livro exclusivo a este tema e discutem sobre os possíveis casos práticos utilizados para fazer a equalização. É este processo que garantirá que os desempenhos dos respondentes de vários anos avaliados estarão em uma mesma métrica, ou seja, em uma mesma escala de medida.

2.7. Aplicação das Provas e Processamento da Base de Dados

Antes da aplicação das provas, os participantes (ou responsáveis) preenchem os formulários contextuais. Estes formulários visam descrever o perfil do aluno respondente, coletando informações, como por exemplo, o nível socioeconômico

e grau de instrução dos pais. No SARESP, geralmente as provas são aplicadas em dois dias consecutivos, conforme Quadro 2.3.

Quadro 2.3: Aplicação das provas segundo o ano avaliado e disciplinas.

Data	Prova	Anos/Série
1º	Língua Portuguesa (Linguagens)	3º e 5º anos EF
Dia	Matemática	7º e 9º anos EF e 3ª série EM
2º	Matemática	3º e 5º anos EF
Dia	Língua Portuguesa (Linguagens)	7º e 9º anos EF e 3ª série EM

Fonte: Manual do Apoio Regional e Agente Vunesp (SÃO PAULO, 2015a).

De acordo com o Manual do Apoio Regional e Agente Vunesp (SÃO PAULO, 2015a), participam do SARESP os alunos do 3º, 5º e 9º anos do Ensino Fundamental – EF e do 3º ano do Ensino Médio - EM, das escolas estaduais do Estado de São Paulo (todas), municipais e particulares (que aderiram) e uma amostra de 10% dos alunos das escolas estaduais que oferecem o 7º ano. Os componentes curriculares a serem avaliados são Língua Portuguesa (Linguagens) e Matemática.

Para evitar que um caderno de prova do SARESP seja respondido por proporções bem distintas do que outros, os cadernos de provas já vão identificados para o local de aplicação, constando os seguintes dados: Aluno, Turma, Turno, Escola, Município, DE/SME e Rede de Ensino. Seguindo os instrumentos de orientação (SÃO PAULO, 2015a), os agentes da entidade organizadora aplicam as provas da avaliação nos alunos participantes.

Terminada a aplicação das provas, a entidade organizadora recolhe os instrumentos de avaliação e faz a importação dos dados para o sistema, onde estes serão processados sob a metodologia TRI.

Capítulo 3

Breve Introdução à Teoria da Resposta ao Item (TRI)

A Teoria da Resposta ao Item (TRI) teve a sua formalização teórica iniciada por Frederic Lord entre os anos de 1950 e 1960, contando com uma importante contribuição de Georg Rasch mediante uma aplicação de um modelo dicotômico¹⁶ na área de avaliação educacional (PASQUALI & PRIMI, 2003).

Apesar do embasamento teórico da TRI ter se desenvolvido na década de 1950, de acordo com Pasquali (2013), foi somente a partir dos anos de 1980 que a TRI passou a ser predominantemente utilizada na Psicometria devido à complexidade de seus modelos matemáticos, cujas estimações dos parâmetros seriam inviáveis sem os programas de computador desenvolvidos nessa mesma época. No Brasil, a TRI passou a ser utilizada em 1995 no Sistema de Avaliação da Educação Básica (SAEB), que tem por objetivo avaliar a qualidade do sistema educacional do país (ANDRADE et al., 2010).

Quando se deseja verificar a qualidade da Educação, é comum pensar em mensurar o aprendizado dos alunos por meio do desempenho deles nas avaliações. Este desempenho é a característica medida em cada aluno, mesmo sabendo que seus valores variam de aluno para aluno. Segundo Pereira (2004), apesar de na Educação a variável de interesse ser de entendimento intuitivo de todos, na maioria das vezes, esta não é observável diretamente. Talvez por isso, exista tanta dificuldade na compreensão de instrumentos que se proponham a fazer esta medição, gerando várias discussões sobre a validade de seus resultados.

Geralmente, nas ciências humanas e da saúde as variáveis que não são observáveis e nem manipuláveis diretamente são chamadas de variáveis latentes (CURADO et al., 2014). Nas avaliações educacionais em larga escala,

¹⁶ Duas categorias de respostas. Por exemplo: (Sim, Não).

entende-se como traço latente a habilidade do aluno. Ou seja, a proficiência, o aprendizado, o saber.

De acordo com Pasquali (2009), a Psicometria é o ramo da Psicologia, especialmente aplicada à Educação, responsável por explicar o sentido entre as respostas dadas pelos sujeitos a determinadas tarefas e propor técnicas de medida dos processos mentais. Duas dessas técnicas de medidas são: a Teoria Clássica dos Testes (TCT) e a Teoria da Resposta ao Item (TRI). De acordo com a pesquisa realizada com os professores (Apêndice A), a TCT é mais prática, objetiva, simples e é clara em relação à forma de mensurar o desempenho do aluno, pois se sabe claramente a quantidade de acertos e erros. Já a TRI, segundo os mesmos, é uma metodologia complexa e de difícil compreensão. Mas, as justificativas apontam que os professores acham que a TRI é melhor por avaliar as habilidades dos alunos, permitindo, por meio do resultado e de posse da escala, verificar o que o aluno domina e quais habilidades precisarão ser melhor trabalhadas.

Dessa forma, a ênfase e o destaque neste trabalho serão dados à TRI, uma vez que esta teoria é utilizada nas principais avaliações externas em larga escala no país, com o intuito de apresentar a teoria de modo didático, para facilitar a compreensão dos docentes da escola básica.

A TRI é uma teoria que tem por objetivo analisar especificamente cada item de uma avaliação, a fim de saber qual a probabilidade de um aluno responder corretamente um item específico dado determinado nível de habilidade.

Atualmente, nas avaliações de larga escala, utiliza-se a TRI sob duas hipóteses básicas que devem ser respeitadas para sua aplicação (PASQUALI, 2013):

1. Unidimensionalidade: esta hipótese supõe que apenas uma aptidão ou habilidade seja medida. Ou seja, o conjunto de itens deve estar medindo um único traço latente, que no caso, é a habilidade do aluno, a proficiência. É claro que para execução de uma tarefa exige-se uma série de habilidades, mas para satisfazer esta hipótese é suficiente admitir que exista uma habilidade dominante, responsável pelo conjunto de itens. E é

esta habilidade dominante a considerada como o traço latente medido na avaliação.

2. Independência Local: supõe que para uma determinada habilidade a resposta dada a um item independe da outra, ou seja, as respostas dadas aos itens são independentes entre si. Segundo Hambleton (1991, apud PEREIRA, 2004, p. 34), teoricamente a unidimensionalidade implica na independência local e assim apenas uma e não duas hipóteses devem ser verificadas.

Diante de um item, a probabilidade de um aluno respondê-lo corretamente depende de sua habilidade (PEREIRA, 2004). Rasch, ao propor o seu modelo, enunciou um postulado simples e direto:

“uma pessoa que tem uma habilidade mais desenvolvida do que outra pessoa deve ter maior probabilidade de resolver qualquer item de um tipo apresentado, e da mesma forma, um item mais difícil do que o outro implica que para qualquer pessoa a probabilidade de resolver o item mais difícil é menor do que a de resolver o item mais fácil”. (RASCH, 1960, p. 117, apud BOND & FOX, 2010, p. 10 – Tradução livre do autor)

O item também é conhecido como cada uma das questões que compõem a prova, a avaliação ou o teste. Porém, como o foco é a avaliação em larga escala, o Guia do CAEd (JUIZ DE FORA, 2008) destaca que o item se diferencia das questões que compõem as avaliações comuns devido ao item avaliar apenas uma habilidade e abordar uma única dimensão do conhecimento. Entenda-se habilidade como a aptidão desenvolvida que permite realizar com certa facilidade determinadas atividades (HAMBLETON et al., 1991) e dimensão do conhecimento como os conhecimentos efetivo, conceitual, procedural e metacognitivo (SILVA, 2013) que os alunos devem dominar.

O modelo da TRI, utilizado nas avaliações externas, leva em conta a noção teórica de que o desenvolvimento do aluno é um processo cumulativo. De acordo com Todd & Wolpin (2003 e 2007), a produção de habilidades cognitivas é um

processo cumulativo e que depende da história dos insumos relacionados ao conhecimento fornecidos pelas famílias e escolas, bem como sobre dotes herdados das crianças. Deste modo, quanto maior a habilidade ou proficiência do aluno, maior sua probabilidade de responder corretamente ao item.

Em suma, nesse modelo, a habilidade dos alunos pode ser representada por meio do comportamento das respostas dadas ao conjunto de itens selecionados. Segundo Pasquali (2013), é a partir destas respostas que se pode inferir sobre o traço latente do aluno, criando relações hipotéticas entre as respostas observadas e o nível do traço latente, de tal forma que as relações possam ser expressas por meio de uma equação matemática, chamada de modelo, que descreva a forma de função que estas relações possuem.

Capítulo 4

Abordagem Pedagógica sobre a Construção dos Modelos da TRI

Frente aos resultados da pesquisa (Apêndice A) realizada com professores da educação básica, a percepção é que ainda há muitas dúvidas acerca de como se traça a curva do item que possui apenas dois resultados possíveis: acerto ou erro. E ainda mais sobre a interpretação dessa curva, cujo gráfico traz uma relação entre a habilidade do aluno (no eixo das abscissas) e a probabilidade de acerto do item (no eixo das ordenadas), não ficando claro como foi obtida ou estimada.

Com objetivo de dirimir essas dúvidas e contribuir para um melhor entendimento sobre a TRI, o processo de construção do gráfico abordará o pensamento de desenvolvimento da metodologia. Será utilizada uma **abordagem pedagógica**, ou seja, os procedimentos realizados a seguir, por não serem usuais, visam apenas facilitar a compreensão da TRI (ver nota de rodapé nº 10, na Introdução).

4.1. Construção do Gráfico a Partir de Dados Empíricos

No contexto do Capítulo 3, em que a relação entre a habilidade dos alunos e as respostas dadas ao conjunto de itens selecionados podem ser expressas por meio de uma equação matemática, modelo é uma descrição da base de dados do desempenho dos alunos em termos de parâmetros (informações), ou seja, de informações contidas e de relacionamentos existentes na base de dados. A ideia do uso de modelo matemático em avaliações é, a partir de respostas dadas pelo aluno a determinados itens especificados, poder inferir sobre suas habilidades e competências.

A partir disso, a TRI assume dois princípios fundamentais (PASQUALI, 2013) que, para facilitar a compreensão, podem ser interpretados da seguinte forma:

1. A probabilidade de um aluno acertar um item especificado pode ser predita por meio de um conjunto de habilidades, em que uma é dominante. Nesse sentido, esta habilidade dominante é a causa e o desempenho é a consequência ou efeito dessa habilidade.
2. Quanto maior a habilidade do aluno, maior será a probabilidade de ele acertar o item e o inverso também é verdadeiro.

A partir desta breve descrição dos princípios, pode-se considerar que níveis maiores de habilidade conduzem a valores mais altos de probabilidade de responder corretamente ao item, apresentando um comportamento sempre crescente (de 0% a 100%).

Pela Figura A.12, foi visto que os professores acham que saber sobre a construção do gráfico poderá sim trazer algum tipo de benefício para o ensino em sala de aula e contribuir para melhoria de sua prática pedagógica (93%). Considerando os pressupostos descritos acima e visando facilitar o entendimento sobre a TRI, será apresentada uma sequência de passos até a construção do gráfico característico da *curva da probabilidade* de responder corretamente aos itens, utilizando-se de procedimentos que apenas possuem fins pedagógicos à compreensão da aplicação prática da teoria. Sendo assim, suponha que estamos na etapa de calibração dos itens. No exemplo a ser explorado, foi considerada uma avaliação de matemática composta inicialmente por 14 itens, respondidos por 15 alunos de um mesmo ano da educação básica, para a construção inicial do modelo que, como já dissemos, será o modelo com um único parâmetro (dificuldade do item), conhecido como modelo de Rasch.

Passo 1: Construir uma tabela em que cada linha representa um aluno e cada coluna um item respondido. Dado que os itens da avaliação já foram respondidos pelos alunos, identificar os itens corretos e incorretos para cada aluno, indicando com o número “1” (um) os itens acertados e com o “0” (zero) os itens errados. Mesmo que o item possua mais de duas alternativas, o que é comum nas avaliações em larga escala, o que é considerado aqui para cálculo dos resultados é apenas se o aluno acertou ou errou o item. Após identificar os

acertos e os erros, deve-se totaliza-los para cada aluno, conforme a representação feita na Tabela 4.1.

Tabela 4.1: Tabela de acertos e erros por alunos e itens.

Alunos	Itens														Nº de Acertos do Aluno	Nº de Erros do Aluno
	a	b	c	d	e	f	g	h	i	j	k	l	m	n		
Amanda	1	1	1	1	1	1	1	1	1	1	1	1	0	0	12	2
Bruno	0	0	0	1	0	1	0	0	0	0	0	0	0	0	2	12
Carlos	0	1	0	1	0	1	1	0	1	1	0	0	0	0	6	8
Débora	1	1	0	1	0	1	1	1	1	1	1	1	0	1	11	3
Enzo	0	1	1	1	1	1	1	0	1	1	1	0	0	1	10	4
Fábio	0	1	1	1	1	1	1	1	1	1	0	1	0	1	11	3
Gilson	0	1	0	1	0	1	1	0	0	1	0	0	0	0	5	9
Hélen	0	0	0	1	0	1	1	0	0	0	0	0	0	0	3	11
Ivis	0	1	0	1	1	1	1	0	0	1	1	1	0	1	9	5
Joice	0	0	0	1	1	1	1	0	1	1	0	1	0	0	7	7
Kelly	0	1	0	1	0	1	0	0	0	0	0	0	0	0	3	11
Lais	0	0	0	1	1	1	1	0	1	1	1	1	0	0	8	6
Maria	0	1	1	0	0	1	1	0	1	1	1	1	0	1	9	5
Natan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
Ophir	1	1	1	1	1	1	1	1	1	1	1	1	1	1	14	0
Nº de Acertos no Item	3	10	5	13	7	14	12	4	9	11	7	8	1	6		
Nº de Erros no Item	12	5	10	2	8	1	3	11	6	4	8	7	14	9		

Passo 2: Ordenar os alunos de forma decrescente do total de acertos e os itens de forma crescente de acertos (o que corresponde, para os itens, a uma ordem decrescente de dificuldade, ou seja, o número de acertos pequeno sugere que o item seja difícil). Este procedimento, apresentado na Tabela 4.2, permite algumas análises preliminares. Por exemplo, verificar os alunos com melhor desempenho (ou pior) e os itens mais difíceis (ou fáceis). Repare que na tabela pode-se perceber a divisão entre os “uns” e os “zeros”, onde a partir da diagonal decrescente do lado esquerdo para o direito e acima da tabela estão os alunos que mais acertaram (Ophir, Amanda e Débora) e os itens que tiveram mais acertos (item “f”, item “d” e item “g”), e abaixo da mesma diagonal exatamente o oposto.

Note também que há alguns erros (zeros), ou seja, itens respondidos incorretamente, não esperados na tabela. Por exemplo, veja o caso da Maria: o item “d” foi respondido corretamente por quase todos os alunos, mas, a Maria o respondeu incorretamente. Porém, ela acertou os itens “c” e “n” que a maioria errou. Caberia uma análise pontual em casos como esses para verificar o que aconteceu. Mas, é importante lembrar que a TRI leva em consideração a coerência dos acertos que pode ser vista no Capítulo 5, quando se discute o cálculo da habilidade do respondente.

Tabela 4.2: Tabela ordenada pelo total de acertos dos alunos e nos itens.

Alunos	Itens														Nº de Acertos do Aluno	Nº de Erros do Aluno
	m	a	h	c	n	e	k	l	i	b	j	g	d	f		
Ophir	1	1	1	1	1	1	1	1	1	1	1	1	1	1	14	0
Amanda	0	1	1	1	0	1	1	1	1	1	1	1	1	1	12	2
Débora	0	1	1	0	1	0	1	1	1	1	1	1	1	1	11	3
Fábio	0	0	1	1	1	1	0	1	1	1	1	1	1	1	11	3
Enzo	0	0	0	1	1	1	1	0	1	1	1	1	1	1	10	4
Ivis	0	0	0	0	1	1	1	1	0	1	1	1	1	1	9	5
Maria	0	0	0	1	1	0	1	1	1	1	1	1	0	1	9	5
Lais	0	0	0	0	0	1	1	1	1	0	1	1	1	1	8	6
Joice	0	0	0	0	0	1	0	1	1	0	1	1	1	1	7	7
Carlos	0	0	0	0	0	0	0	0	1	1	1	1	1	1	6	8
Gilson	0	0	0	0	0	0	0	0	0	1	1	1	1	1	5	9
Hélen	0	0	0	0	0	0	0	0	0	0	0	1	1	1	3	11
Kelly	0	0	0	0	0	0	0	0	0	1	0	0	1	1	3	11
Bruno	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	12
Natan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
Nº de Acertos no Item	1	3	4	5	6	7	7	8	9	10	11	12	13	14		
Nº de Erros no Item	14	12	11	10	9	8	8	7	6	5	4	3	2	1		

Passo 3: Calcular os percentuais de acertos e erros para cada aluno e cada item, conforme a Tabela 4.3. O percentual de erros é o complementar do percentual de acertos, ou seja, o percentual de erros mais o percentual de acertos é igual a 100%.

Tabela 4.3: Tabela ordenada com os percentuais de acertos dos alunos e nos itens.

Alunos	Itens														Nº de Acertos do Aluno	Nº de Erros do Aluno	% Acertos	% Erros
	m	a	h	c	n	e	k	l	i	b	j	g	d	f				
Ophir	1	1	1	1	1	1	1	1	1	1	1	1	1	1	14	0	100%	0
Amanda	0	1	1	1	0	1	1	1	1	1	1	1	1	1	12	2	86%	14%
Débora	0	1	1	0	1	0	1	1	1	1	1	1	1	1	11	3	79%	21%
Fábio	0	0	1	1	1	1	0	1	1	1	1	1	1	1	11	3	79%	21%
Enzo	0	0	0	1	1	1	1	0	1	1	1	1	1	1	10	4	71%	29%
Ivis	0	0	0	0	1	1	1	1	0	1	1	1	1	1	9	5	64%	36%
Maria	0	0	0	1	1	0	1	1	1	1	1	1	0	1	9	5	64%	36%
Lais	0	0	0	0	0	1	1	1	1	0	1	1	1	1	8	6	57%	43%
Joice	0	0	0	0	0	1	0	1	1	0	1	1	1	1	7	7	50%	50%
Carlos	0	0	0	0	0	0	0	0	1	1	1	1	1	1	6	8	43%	57%
Gilson	0	0	0	0	0	0	0	0	0	1	1	1	1	1	5	9	36%	64%
Hélen	0	0	0	0	0	0	0	0	0	0	0	1	1	1	3	11	21%	79%
Kelly	0	0	0	0	0	0	0	0	0	1	0	0	1	1	3	11	21%	79%
Bruno	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	12	14%	86%
Natan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0%	100%
Nº de Acertos no Item	1	3	4	5	6	7	7	8	9	10	11	12	13	14				
Nº de Erros no Item	14	12	11	10	9	8	8	7	6	5	4	3	2	1				
% Acertos	7%	20%	27%	33%	40%	47%	47%	53%	60%	67%	73%	80%	87%	93%				
% Erros	93%	80%	73%	67%	60%	53%	53%	47%	40%	33%	27%	20%	13%	7%				

Passo 4: Criar um gráfico com os acertos dos alunos na escala de pontos percentuais. A Figura 4.1 apresenta a execução deste passo para os alunos da Tabela 4.3.

Figura 4.1: Distribuição do desempenho dos alunos em termos percentuais.

Apesar de ser usual trabalhar com escala em pontos percentuais, principalmente tratando-se do desempenho de alunos, cuidados devem ser tomados. A escala

percentual é muito sensível à sua base comparativa e não deve ser usada quando se trabalha com poucos dados (BOND & FOX, 2010). Por exemplo, para um grupo de alunos de um mesmo ano do ensino fundamental, suponha uma prova que possua apenas 4 questões e que um aluno tenha respondido corretamente duas questões dessas quatro. Ora, o seu desempenho foi de $2/4$ (dois quartos) ou 0,5 ou $0,5 \times 100\% = 50\%$. Se este mesmo aluno, ao invés de ter errado 2 questões, tivesse errado mais uma questão, o seu desempenho seria de $1/4$ ou 25% (1 acerto em 4 questões). Neste exemplo, o erro de um item fez o desempenho do aluno variar 25 pontos percentuais. Agora considere uma prova com 100 questões, em que o desempenho de um aluno foi de 50% (acertou 50 questões em 100). Caso este aluno tivesse errado mais uma questão (obtendo ao todo 51 questões erradas em 100), ele teria um desempenho de 49% (49 acertos em 100 questões). Neste exemplo, o erro de um item fez o desempenho do aluno variar apenas 1 ponto percentual. Devido a sensibilidade da escala em percentual, cuidados são necessários ao se comparar resultados nessa escala, pois a comparação pode não ser aplicável.

Além disso, as distâncias na escala em percentual não possuem os mesmos significados quando se compara o centro com as extremidades da escala. Segundo Bond & Fox (2010), ao usar esta escala para pontuações brutas, ela tende a agrupar os alunos em torno dos valores médios, sem contrastar adequadamente os alunos menos capazes e os mais capazes. Assim, considerando ainda um grupo de alunos de um mesmo ano do ensino fundamental, os alunos posicionados no inferior da escala podem ter maior dificuldade de avançar do que os demais que já possuem uma base consolidada. E neste caso, qualquer ponto percentual adquirido poderia ser significativo. Isto também vale para os alunos classificados na parte superior da escala, pois eles já dominam tantos assuntos que elevar esse domínio, com conteúdo mais complexo, poderia ser mais difícil para eles do que os demais alunos posicionados em pontos inferiores. Em ambos os casos extremos, qualquer ponto percentual poderia ser significativo. Enfim, é como se o ponto percentual não tivesse o mesmo grau de equivalência ao longo da escala.

Guskey (2013) discute o uso e interpretação da escala em percentual, destacando no início de seu artigo que “*é hora de abandonar escalas de*

classificação que distorcem a precisão, objetividade e confiabilidade das notas dos alunos". Ele coloca que a porcentagem de itens corretos, dentre o total respondido, é utilizada para refletir sobre a porcentagem do conteúdo que um aluno aprendeu. Porém, os graus percentuais derivados apenas da porcentagem correta, sem um exame cuidadoso dos itens ou tarefas realizadas pelos alunos, não trazem toda a informação. Ou seja, os erros podem trazer outras informações a respeito do desempenho dos alunos que podem justificar o seu desempenho. Daí a ideia de utilizar acertos e erros em conjunto.

Ao criar uma relação entre acertos e erros, estabelece-se uma medida comparativa e então diminui a interferência (ou dependência) da quantidade total de itens, obtendo-se assim, a quantidade de acertos para cada erro. A discussão de Guskey trazida acima, tem por objetivo introduzir a ideia de chances, também conhecida como "odds"¹⁷.

No Brasil, a palavra chance é utilizada em várias definições, dentre as quais destacamos:

1. Acaso, sorte, probabilidade¹⁸;
2. Possibilidade de algo acontecer¹⁹;
3. Razão entre as probabilidades de ocorrência e não ocorrência do mesmo evento²⁰.

Em especial na educação básica, o termo chance tem sido utilizado no sentido das duas primeiras definições acima. Já em alguns ambientes da probabilidade e estatística, o termo chance está relacionado com a terceira definição. A probabilidade é uma medida de uma incerteza ou de uma expectativa de

¹⁷ Geralmente, o tema "odds" (chance) é abordado dentro de outro mais difundido: "odds ratio" ou razão de chances. Para mais informações, consulte: Agresti (2002), Merchan-Hamann et al. (2000) e Hosmer & Lemeshow (2000).

¹⁸ Definição dada pelo Dicionário Webster's (Inglês-Português)

¹⁹ Definição dada pelo Dicionário Priberam da Língua Portuguesa (DPLP), que é um dicionário de português contemporâneo e que permite a consulta de acordo com a norma do português europeu ou do Brasil. Ele pode ser consultado pela web no site: <http://www.priberam.pt/dlpo/>.

²⁰ Na Probabilidade e Estatística, a chance de um evento acontecer é definida como a razão entre as probabilidades de ocorrência e não ocorrência do mesmo evento (TRIOLA, 2005).

ocorrência acerca de um evento. E a razão entre probabilidades exprime a relação de ocorrência entre os eventos, cujo resultado informa se (dentre os dois) há algum com ocorrência mais provável do que o outro ou não.

Na língua inglesa, o termo *chance* possui as definições 1 e 2 acima. E o termo *Odds* a definição 3. Ou seja, não há confusão dentre as definições. Neste texto, será considerada a nomenclatura *chances (odds)* para referir-se à divisão do número de acertos pelo número de erros. Ou seja, quantos acertos obtidos para cada erro. Assim, a base comparativa deixa de ser o total de itens e passa a ser, no caso dos alunos, a quantidade de erros.

Passo 5: Calcular as chances (*odds*) entre os percentuais de acertos e erros para cada aluno, conforme a Tabela 4.4.

Por exemplo, considere os alunos da Tabela 4.3 e veja os cálculos executados abaixo:

- **Amanda:**

Ela acertou 11 dentre os 14 itens, então o percentual de acertos foi de $\frac{11}{14} \times 100\% = 78,57\%$. Já o percentual de erros foi de $\frac{3}{14} \times 100\% = 21,43\%$. As chances (*odds*) da Amanda são:

$$\frac{78,57\%}{21,43\%} = \frac{\frac{11}{14}}{\frac{3}{14}} = \frac{11}{14} \times \frac{14}{3} = \frac{11}{3} = 3,67.$$

Note que o número total de itens (14) é cancelado, eliminando sua interferência. Interpretando as chances (*odds*) da Amanda, temos que ela teve quase 4 acertos para cada erro (ela mais acertou do que errou). De forma análoga, foram calculadas as chances (*odds*) dos demais alunos (Tabela 4.4), com exceção do Ophir, cujo cálculo das chances (*odds*) não está definido devido à divisão por zero, já que ele acertou 14 dos 14 itens, obtendo o percentual de erros, portanto, igual a 0%.

Assim, para seguir com o procedimento, **o Ophir será excluído das considerações a partir deste ponto.**

Tabela 4.4: Tabela ordenada com as chances (*odds*) dos alunos.

Alunos	Itens														Nº de Acertos do Aluno (a)	Nº de Erros do Aluno (b)	% Acertos (c)	% Erros (d)	Chances ou Odds (a/b ou c/d)
	m	a	h	c	n	e	k	l	i	b	j	g	d	f					
Amanda	0	1	1	1	0	1	1	1	1	1	1	1	1	1	12	2	86%	14%	6,0
Débora	0	1	1	0	1	0	1	1	1	1	1	1	1	1	11	3	79%	21%	3,7
Fábio	0	0	1	1	1	1	0	1	1	1	1	1	1	1	11	3	79%	21%	3,7
Enzo	0	0	0	1	1	1	1	0	1	1	1	1	1	1	10	4	71%	29%	2,5
Ivis	0	0	0	0	1	1	1	1	0	1	1	1	1	1	9	5	64%	36%	1,8
Maria	0	0	0	1	1	0	1	1	1	1	1	1	0	1	9	5	64%	36%	1,8
Lais	0	0	0	0	0	1	1	1	1	0	1	1	1	1	8	6	57%	43%	1,3
Joice	0	0	0	0	0	1	0	1	1	0	1	1	1	1	7	7	50%	50%	1,0
Carlos	0	0	0	0	0	0	0	0	1	1	1	1	1	1	6	8	43%	57%	0,8
Gilson	0	0	0	0	0	0	0	0	0	1	1	1	1	1	5	9	36%	64%	0,6
Hélen	0	0	0	0	0	0	0	0	0	0	0	1	1	1	3	11	21%	79%	0,3
Kelly	0	0	0	0	0	0	0	0	0	1	0	0	1	1	3	11	21%	79%	0,3
Bruno	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	12	14%	86%	0,2
Natan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0%	100%	0,0
Nº de Acertos no Item	0	2	3	4	5	6	6	7	8	9	10	11	12	13					
Nº de Erros no Item	14	12	11	10	9	8	8	7	6	5	4	3	2	1					
% Acertos	0%	14%	21%	29%	36%	43%	43%	50%	57%	64%	71%	79%	86%	93%					
% Erros	100%	86%	79%	71%	64%	57%	57%	50%	43%	36%	29%	21%	14%	7%					

A partir da Tabela 4.4, criou-se um gráfico de pontos das chances (*odds*) dos alunos, conforme Figura 4.2.

Figura 4.2: Distribuição das chances (*odds*) (acertos/erros) dos alunos.

Observe que ao trocar a escala percentual pela escala de chances (*odds*), houve uma alteração na distância entre os alunos (mas não da ordem).

Enquanto os alunos estão mais espalhados ao longo da escala percentual (Figura 4.1), observa-se uma concentração de alunos no lado esquerdo da escala de chances (*odds*) (Figura 4.2), no intervalo de 0 a 2. Note também que o espaçamento (diferença entre duas chances | *odds*) entre os alunos tende a ficar menor à medida que as chances (*odds*) são menores.

Pode ser observado que a escala de chances (*odds*) é limitada inferiormente por 0 (zero), mas sem limite superior. Ou seja, se a escala em percentual era limitada aos valores 0% (limite inferior) e 100% (limite superior), a escala de chances (*odds*) varia de 0 (zero) ao *mais infinito* ($+\infty$).

Uma saída para resolver o problema da concentração de pontos na parte inferior da escala, é aplicação de uma transformação matemática a fim de distribuir as chances (*odds*) de uma forma mais simétrica, em torno de um valor que seria tomado como referência. A transformação usualmente utilizada é aplicação do Logaritmo Natural²¹ nas chances (*odds*), obtendo-se uma escala logarítmica que se aproxima de uma escala intervalar, com o qual o valor da escala é mantido em qualquer posição ao longo desta escala (BOND & FOX, 2010). A partir de uma escala intervalar, pode-se começar a dizer o quanto um aluno é mais capaz do que outro, em vez de simplesmente dizer que um é mais capaz do que o outro.

Passo 6: Calcular o Logaritmo Natural das Chances (*odds*) obtidas para cada aluno no passo anterior (Tabela 4.5). Com a aplicação da transformação logarítmica nas chances (*odds*), notou-se que há um problema no cálculo do logaritmo da nota do Natan, que é 0 (zero). Isto ocorreu porque a função logarítmica não está definida no valor 0 (zero)²², o que implica na impossibilidade de aplicação da transformação na nota do Natan. Desta forma, **o Natan será excluído das considerações a partir deste passo**. A eliminação de itens e/ou alunos foi realizada neste texto devido a abordagem deste trabalho ser de cunho

²¹ Sobre Logaritmo Natural consulte Lima (2009).

²² Para saber mais sobre o intervalo de definição da função logarítmica, consulte Matos (2014). Sua dissertação se apoia em referências como Geraldo Ávila e Elon Lages Lima.

pedagógico. Porém, atualmente, casos como os de Ophir e Natan são tratados pelas metodologias usuais da TRI e nenhuma exclusão é realizada. Esses tratamentos não serão aprofundados neste texto.

Tabela 4.5: Tabela ordenada com o cálculo do logaritmo natural aplicado às chances (*odds*) dos alunos.

Alunos	Itens														Nº de Acertos do Aluno (a)	Nº de Erros do Aluno (b)	% Acertos (c)	% Erros (d)	Chances ou Odds (a/b ou c/d)	LN das Chances
	m	a	h	c	n	e	k	l	i	b	j	g	d	f						
Amanda	0	1	1	1	0	1	1	1	1	1	1	1	1	1	12	2	86%	14%	6,0	1,79
Débora	0	1	1	0	1	0	1	1	1	1	1	1	1	1	11	3	79%	21%	3,7	1,30
Fábio	0	0	1	1	1	1	0	1	1	1	1	1	1	1	11	3	79%	21%	3,7	1,30
Enzo	0	0	0	0	1	1	1	0	1	1	1	1	1	1	10	4	71%	29%	2,5	0,92
Ivis	0	0	0	0	1	1	1	1	0	1	1	1	1	1	9	5	64%	36%	1,8	0,59
Maria	0	0	0	1	1	0	1	1	1	1	1	1	0	1	9	5	64%	36%	1,8	0,59
Lais	0	0	0	0	0	1	1	1	1	0	1	1	1	1	8	6	57%	43%	1,3	0,29
Joice	0	0	0	0	0	1	0	1	1	0	1	1	1	1	7	7	50%	50%	1,0	0,00
Carlos	0	0	0	0	0	0	0	0	0	1	1	1	1	1	6	8	43%	57%	0,8	-0,29
Gilson	0	0	0	0	0	0	0	0	0	0	1	1	1	1	5	9	36%	64%	0,6	-0,59
Hélen	0	0	0	0	0	0	0	0	0	0	0	0	1	1	3	11	21%	79%	0,3	-1,30
Kelly	0	0	0	0	0	0	0	0	0	0	1	0	0	1	3	11	21%	79%	0,3	-1,30
Bruno	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	12	14%	86%	0,2	-1,79
Nº de Acertos no Item	0	2	3	4	5	6	6	7	8	9	10	11	12	13						
Nº de Erros no Item	13	11	10	9	8	7	7	6	5	4	3	2	1	0						
% Acertos	0%	15%	23%	31%	38%	46%	46%	54%	62%	69%	77%	85%	92%	100%						
% Erros	100%	85%	77%	69%	62%	54%	54%	46%	38%	31%	23%	15%	8%	0%						

A Figura 4.3 apresenta um gráfico de pontos com o logaritmo natural aplicado nas chances (*odds*) dos alunos. Observe como as características (concentração de pontos na parte inferior da escala de chances | *odds*) apresentadas na Figura 4.2 se desfizeram quando se aplicou o logaritmo nas chances (*odds*). O logaritmo das chances (*odds*) estão distribuídos de uma forma mais simétrica em torno do valor 0 (zero), assemelhando-se visualmente à distribuição observada na escala em percentual (Figura 4.1). Note que a escala do logaritmo das chances (*odds*) não possui limite inferior e nem superior. Ou seja, a escala do logaritmo das chances (*odds*) varia do infinito dos números negativos, também chamado de *menos infinito* ($-\infty$), ao infinito dos números positivos ou *mais infinito* ($+\infty$).

Figura 4.3: Distribuição do logaritmo natural das chances (*odds*) (acertos/erros) dos alunos.

Apenas para efeito comparativo entre as distribuições do percentual de acertos e o logaritmo natural das chances (*odds*) dos Alunos, observe a Figura 4.4. Note que a aluna Joice que acertou 50% dos itens da avaliação está centrada no valor 0 (zero) do logaritmo natural das chances (*odds*). Observe também, que o logaritmo natural das chances (*odds*) separou um pouco os alunos nos extremos da escala, que estavam mais concentrados em relação ao percentual de acertos, permitindo assim, uma melhor distinção entre os desempenhos dos alunos. Por exemplo, a distância²³ entre o Bruno (14% de acertos) e a Hélen (21% de acertos) em relação ao percentual de acertos é de 7 pontos percentuais (p.p.), que é a mesma distância entre o Carlos (43% de acertos) e a Joice (50% de acertos). Já em relação ao logaritmo natural das chances (*odds*), a distância entre o Bruno e a Hélen é de 0,49, valor maior que a distância entre o Carlos e a Joice que é de 0,29.

É imediato perceber que não há uma correspondência entre a magnitude da distância entre dois alunos em relação ao percentual de acertos e o logaritmo das chances (*odds*). Mas a ordem entre os alunos é mantida tanto nas escalas percentual e das chances (*odds*) quanto no logaritmo das chances (*odds*).

²³ Módulo da diferença entre dois pontos. Stewart, James B. Calculus: concepts and contexts. Belmont: Brooks Cole, 4th ed., 2009. 1152p.

Figura 4.4: Comparação entre as distribuições do percentual de acertos e o logaritmo natural das chances | *odds* - (acertos/erros) dos alunos.

Passo 7: Repetir os passos 4 e 5, mas agora para os itens. No caso dos itens, há uma inversão no cálculo das chances (*odds*), considerando a divisão do número de erros pelo número de acertos. Ou seja, quantos erros obtidos para cada acerto. Esta inversão se deve à interpretação da escala, para que valores positivos sejam interpretados como maior dificuldade e valores negativos como menor dificuldade. Abaixo segue o exemplo de cálculo para o logaritmo da divisão do número de erros pelo número de acertos no item:

- Item “a”: $\text{LN}\left(\frac{11}{2}\right) = 1,7$.
- Item “d”: $\text{LN}\left(\frac{1}{12}\right) = -2,5$.

Note que devido à inversão no cálculo das chances (*odds*), considerando a divisão do número de erros pelo número de acertos, como o item “a” possui um

grau de dificuldade maior que o item “d”, ele também possui o logaritmo das chances (*odds*) maior. O que faz mais sentido.

Observe que a execução dos passos 4 e 5 (para os itens) leva à necessidade de exclusão dos itens “m” e “f”, pelos mesmos motivos já apresentados nos cálculos das chances (*odds*) e do logaritmo para os alunos: problemas na divisão por zero e no cálculo do logaritmo no valor zero. Observe na Tabela 4.5 que o item “m” não possui acertos e o item “f” não possui erros. Diante da exclusão destes dois itens, também se faz necessário um ajuste no número de acertos e erros dos alunos, pois se antes eram considerados 14 itens, agora serão 12. A Tabela 4.6 traz o logaritmo natural das chances (*odds*) obtida para cada item e aluno, mediante exclusões realizadas (tanto de itens, como de alunos).

Tabela 4.6: Tabela ordenada com o cálculo dos logaritmos naturais das chances (*odds*) dos alunos e itens.

Alunos													Nº de Acertos do Aluno (a)	Nº de Erros do Aluno (b)	% Acertos (c)	% Erros (d)	Chances ou Odds (a/b ou c/d)	LN das Chances
	a	h	c	n	e	k	l	i	b	j	g	d						
Amanda	1	1	1	0	1	1	1	1	1	1	1	1	11	1	92%	8%	11,0	2,40
Débora	1	1	0	1	0	1	1	1	1	1	1	1	10	2	83%	17%	5,0	1,61
Fábio	0	1	1	1	1	0	1	1	1	1	1	1	10	2	83%	17%	5,0	1,61
Enzo	0	0	1	1	1	1	0	1	1	1	1	1	9	3	75%	25%	3,0	1,10
Ivis	0	0	0	1	1	1	1	0	1	1	1	1	8	4	67%	33%	2,0	0,69
Maria	0	0	1	1	0	1	1	1	1	1	1	0	8	4	67%	33%	2,0	0,69
Lais	0	0	0	0	1	1	1	1	0	1	1	1	7	5	58%	42%	1,4	0,34
Joice	0	0	0	0	1	0	1	1	0	1	1	1	6	6	50%	50%	1,0	0,00
Carlos	0	0	0	0	0	0	0	1	1	1	1	1	5	7	42%	58%	0,7	-0,34
Gilson	0	0	0	0	0	0	0	0	1	1	1	1	4	8	33%	67%	0,5	-0,69
Hélen	0	0	0	0	0	0	0	0	0	0	1	1	2	10	17%	83%	0,2	-1,61
Kelly	0	0	0	0	0	0	0	0	1	0	0	1	2	10	17%	83%	0,2	-1,61
Bruno	0	0	0	0	0	0	0	0	0	0	0	1	1	11	8%	92%	0,1	-2,40
Nº de Acertos no Item (e)	2	3	4	5	6	6	7	8	9	10	11	12						
Nº de Erros no Item (f)	11	10	9	8	7	7	6	5	4	3	2	1						
% Acertos (g)	15%	23%	31%	38%	46%	46%	54%	62%	69%	77%	85%	92%						
% Erros (h)	85%	77%	69%	62%	54%	54%	46%	38%	31%	23%	15%	8%						
Chances ou Odds (f/e ou h/g)	5,5	3,3	2,3	1,6	1,2	1,2	0,9	0,6	0,4	0,3	0,2	0,1						
LN das Chances	1,70	1,20	0,81	0,47	0,15	0,15	-0,15	-0,47	-0,81	-1,20	-1,70	-2,48						

Para que o Leitor possa visualizar os resultados obtidos após a execução dos passos 3, 4 e 5 para os itens, a Figura 4.5 traz os gráficos obtidos ao final de cada passo, já considerando as exclusões dos itens “m” e “f”.

Figura 4.5: Apresentação das distribuições do percentual de acertos, das chances (*odds*) (erros/acertos) e do logaritmo natural das chances (*odds*) dos itens da avaliação.

Até este ponto, foram abordados dois aspectos importantes: o desempenho do aluno e a dificuldade do item. O desempenho do aluno foi mensurado inicialmente pelo percentual de acertos em relação ao total de itens. E a dificuldade do item pelo percentual de erros em relação ao total de respondentes.

Porém, com a transformação para a escala do logaritmo das chances (*odds*), o referencial foi alterado e passou-se a comparar o número de acertos em relação ao número de erros para o aluno, e o número de erros em relação ao número de acertos para o item. Diante disso, a escala logarítmica das chances (*odds*) passa a ser interpretada como habilidade para o aluno e permanece como grau de dificuldade para o item. A escala de medida obtida com a transformação logarítmica é a escala (0,1), em que 0 (zero) é a média e 1 é o desvio padrão. A transformação para a escala de proficiência usual será abordada na Seção 6.1.

Como a habilidade do aluno e a dificuldade do item estão dispostas em uma mesma escala, ambas podem ser apresentadas de forma comparativa em um mesmo gráfico. Por meio da Figura 4.6, observe que:

- Os itens estão ordenados pelo grau de dificuldade e os alunos pela sua habilidade;
- Os maiores valores positivos implicam em alto grau de dificuldade para o item ou maior habilidade para o aluno;
- Os menores valores negativos implicam em baixo grau de dificuldade para o item ou menor habilidade para o aluno;
- Em torno do valor 0 (zero) estão os valores intermediários, tanto para a habilidade do aluno quanto para o grau de dificuldade do item.

Ao colocar os itens e os alunos distribuídos conjuntamente na escala do logaritmo da chance (*odds*), podemos compará-los. Por exemplo, observa-se que a habilidade da Amanda supera um pouco o grau de dificuldade do item “a”, que possui o maior grau de dificuldade entre os itens, e supera muito o grau de dificuldade do item “d”, que é o mais fácil (Figura 4.6). Assim, espera-se que Amanda responda corretamente o item “a” e que talvez a Débora não.

Figura 4.6: Distribuição do logaritmo natural das chances (*odds*) dos itens e dos alunos.

Como os itens e os alunos estão em uma mesma escala, parece razoável compará-los por meio da distância. Ou seja, fazer a diferença entre a habilidade do aluno e o grau de dificuldade do item, analisando se esta diferença é a favor ou contra o aluno, se ela favorece ou não o aluno.

Fazendo uma analogia, considere um exemplo citado em Bond & Fox (2010), relacionado à competição de salto com vara, em que competem 6 participantes. De forma muito simplificada, suponha que a dificuldade dos saltos está relacionada somente com a variável altura da barra acima do solo, embora em uma competição real exista outras variáveis que podem exercer influência no salto, tais como a superfície, a temperatura do ar, o vento predominante (direção e força), a chuva, a iluminação (brilho, natural ou artificial), o apoio (ou não) do público, a saúde do atleta (lesões recentes, maleabilidade, etc), dentre outras. Considere também que a melhor marca desses 6 atletas em metros (m) foram 2.01, 1.98, 1.97, 1.95, 1.95 e 1.91, respectivamente, do 1 ao 6. Se for colocada uma barra a uma altura de 1.96m do solo, quem possui a menor probabilidade de passar por cima da barra? Podemos ter interesse em determinar a probabilidade de sucesso (vencer o obstáculo) em cada caso.

Os autores supõem que a melhor marca registrada de cada atleta possa representar sua habilidade. Então a altura da barra acima do solo representa a dificuldade do salto. Repare que tanto a habilidade do atleta quanto a dificuldade

do salto estão em uma mesma escala, são medidas da mesma forma, ou seja, em metros. Portanto, seria razoável compará-las por meio da distância. Dado que a barra está a 1.96m do solo, para o atleta 1, a distância de sua melhor marca a altura da barra é $2.01\text{m} - 1.96\text{m} = 0.05\text{m}$. Para os demais atletas, de 2 a 6, as distâncias em metros são, respectivamente, 0.02, 0.01, -0.01, -0.01 e -0.05. Logo, o atleta 1 possui maior probabilidade de passar por cima da barra, pois a sua melhor marca o favorece, e o atleta 6 possui a menor probabilidade. Já os atletas 4 e 5 possuem a mesma probabilidade, porém suas melhores marcas não favorecem a passagem sobre a barra.

Voltando à situação anterior, em que os itens são considerados dicotômicos, ou seja, se um item for respondido corretamente recebe o valor 1 e se respondido incorretamente recebe o valor 0, considere que a probabilidade de um respondente acertar este item seja p , em que p é um valor entre 0 e 1, e, conseqüentemente, a probabilidade de erra-lo seja $1-p$. Assim, o valor $\text{LN}(p/(1-p)) = \text{LN}(\text{chances} \mid \text{odds})$, já utilizado anteriormente, recebe o nome de logito.

Para efeito de comparabilidade, é definido o **logito como a unidade de medida da escala do logaritmo das chances (odds)** (BOND & FOX, 2010, p. 38). Esta unidade de medida é utilizada tanto para a habilidade dos alunos quanto para a dificuldade dos itens, pois, conforme já dito, ambas estão na mesma escala. A Tabela 4.6 apresenta em sua última coluna o logito calculado para cada aluno e em sua última linha o logito para cada item.

Em continuidade ao exemplo anterior, a diferença entre a habilidade da Amanda e a dificuldade do item “a” é de +0,70 logito. Esta diferença positiva (maior do que zero) implica que a probabilidade de responder corretamente ao item “a” está em favor da Amanda. Já a diferença entre a habilidade do Bruno e a dificuldade do item “a” é de -4,10 logitos, implicando exatamente o contrário da Amanda, ou seja, o Bruno possui probabilidade alta de responder incorretamente o item “a”. Porém, observando a habilidade do Bruno em relação à dificuldade do item “d” (+0,08 logito), nota-se que ele possui maior probabilidade de responder corretamente este item do que respondê-lo incorretamente. Já a habilidade do Bruno em relação à dificuldade do item “g”, a diferença é de -0,70, indicando que o Bruno possui maior probabilidade de responder incorretamente

este item do que respondê-lo corretamente. Note que, qualquer outro item que não seja o item “d”, o Bruno possui maior probabilidade de errar do que de acertar.

Esta relação entre a habilidade do aluno e a dificuldade do item, expressa em logito, pode ser verificada para cada aluno e para cada item. Basta construir uma matriz com estas diferenças o que facilitará uma análise conjunta.

Passo 8: Construir uma matriz diferença entre a habilidade aluno e a dificuldade do item. Esta diferença também é chamada de habilidade do aluno, mas é uma habilidade ajustada, pois ela leva em conta tanto a habilidade do aluno quanto a dificuldade do item. Uma diferença positiva, indica que a habilidade do aluno pode favorecê-lo em responder corretamente ao(s) item(ns), ou seja, quanto maior a diferença positiva, maior a probabilidade de o aluno acertar o item. Já uma diferença negativa, indica que a habilidade do aluno pode não favorecê-lo em responder corretamente ao(s) item(ns), o que implica que quanto maior a diferença negativa, maior a probabilidade de o aluno errar o item. E uma diferença igual a 0 (zero) indica que a habilidade do aluno pode nem favorecer e nem desfavorecer, ou seja, a probabilidade de o aluno acertar ou errar é a mesma (50%).

A Tabela 4.7 apresenta um exemplo das habilidades ajustadas para os alunos em relação à dificuldade do item. Os ajustes foram construídos pela diferença entre o logito da habilidade aluno e o logito da dificuldade do item. As cores dispostas na Tabela 4.7 servem para facilitar a compreensão da habilidade do aluno, indo do vermelho (menor habilidade), passando pelo amarelo (habilidade intermediária), e seguindo para o verde (maior habilidade).

Tabela 4.7: Tabela da habilidade ajustada do aluno (diferença entre a habilidade do aluno e a dificuldade do item).

Alunos	Itens												
	LN()	a	h	c	n	e	k	l	i	b	j	g	d
		1,70	1,20	0,81	0,47	0,15	0,15	-0,15	-0,47	-0,81	-1,20	-1,70	-2,48
Amanda	2,40	0,70	1,20	1,59	1,93	2,25	2,25	2,55	2,87	3,21	3,60	4,10	4,88
Débora	1,61	-0,09	0,41	0,80	1,14	1,46	1,46	1,76	2,08	2,42	2,81	3,31	4,09
Fábio	1,61	-0,09	0,41	0,80	1,14	1,46	1,46	1,76	2,08	2,42	2,81	3,31	4,09
Enzo	1,10	-0,60	-0,10	0,29	0,63	0,95	0,95	1,25	1,57	1,91	2,30	2,80	3,58
Ivis	0,69	-1,01	-0,51	-0,12	0,22	0,54	0,54	0,84	1,16	1,50	1,89	2,39	3,17
Maria	0,69	-1,01	-0,51	-0,12	0,22	0,54	0,54	0,84	1,16	1,50	1,89	2,39	3,17
Laís	0,34	-1,36	-0,86	-0,47	-0,13	0,19	0,19	0,49	0,81	1,15	1,54	2,04	2,82
Joice	0,00	-1,70	-1,20	-0,81	-0,47	-0,15	-0,15	0,15	0,47	0,81	1,20	1,70	2,48
Carlos	-0,34	-2,04	-1,54	-1,15	-0,81	-0,49	-0,49	-0,19	0,13	0,47	0,86	1,36	2,14
Gilson	-0,69	-2,39	-1,89	-1,50	-1,16	-0,84	-0,84	-0,54	-0,22	0,12	0,51	1,01	1,79
Hélen	-1,61	-3,31	-2,81	-2,42	-2,08	-1,76	-1,76	-1,46	-1,14	-0,80	-0,41	0,09	0,87
Kelly	-1,61	-3,31	-2,81	-2,42	-2,08	-1,76	-1,76	-1,46	-1,14	-0,80	-0,41	0,09	0,87
Bruno	-2,40	-4,10	-3,60	-3,21	-2,87	-2,55	-2,55	-2,25	-1,93	-1,59	-1,20	-0,70	0,08

Retomando o exemplo já discutido anteriormente, considere a Amanda que possui habilidade estimada em 2,40 logitos. Em relação ao item “d”, que possui baixo grau de dificuldade estimado em -2,48 logitos, a habilidade da Amanda a favorece e sua habilidade para responder o item “a” passa a ser de 4,88 logitos. Ou seja, a Amanda possui alta probabilidade de acertar ao item “d” que possui baixo grau de dificuldade e, conforme tabelas anteriores, ela acertou. Mas a Amanda poderia perfeitamente errar o item “d”, e caberia aqui uma investigação para saber o porquê a aluna com ótimo desempenho errou um item considerado fácil e verificar a necessidade de um plano de ação.

Sendo assim, a Tabela 4.7 obedece à seguinte relação: a habilidade ajustada de um aluno em um determinado item é igual à sua habilidade menos a dificuldade desse item. Ou seja, para o indivíduo n e o item “k”,

$$\text{Habilidade Ajustada}_{nk} = \text{Habilidade}_n - \text{Dificuldade}_k$$

Visando a compreensão do Leitor e uma vez que os professores estão acostumados a fazer uma análise do desempenho do aluno em relação ao total de itens da avaliação (em que o percentual de acertos é calculado com base no total de itens) e não em relação a um determinado item, os passos seguintes serão executados em termos do resultado geral da avaliação. Mesmo sabendo

que isto não é usual, o autor adaptou o raciocínio anterior, com o intuito de facilitar o entendimento do Leitor sobre a construção do gráfico. Isso também permitirá prosseguir os passos seguintes com o mesmo número de alunos.

Porém, como o usual é realizá-los por item, isto será feito na próxima seção, quando serão levados em conta os passos 9 e 10, já que o baixo número de alunos respondentes por item considerados até aqui poderia dificultar a compreensão das etapas a seguir.

Passo 9: Construir uma tabela com os dados da habilidade de cada aluno e o seu respectivo percentual de acerto, conforme apresentado na Tabela 4.8.

Tabela 4.8: Percentual de acertos por habilidade do aluno.

Alunos	Habilidade	Número de Acertos na Avaliação	Total de Itens da Avaliação	% de Acertos
Amanda	2,40	11	12	91,7%
Débora	1,61	10	12	83,3%
Fábio	1,61	10	12	83,3%
Enzo	1,10	9	12	75,0%
Ivis	0,69	8	12	66,7%
Maria	0,69	8	12	66,7%
Laís	0,34	7	12	58,3%
Joice	0,00	6	12	50,0%
Carlos	-0,34	5	12	41,7%
Gilson	-0,69	4	12	33,3%
Hélen	-1,61	2	12	16,7%
Kelly	-1,61	2	12	16,7%
Bruno	-2,40	1	12	8,3%

Passo 10: Construir um gráfico de dispersão entre a habilidade do aluno (na abscissa) e o percentual de acertos na avaliação (na ordenada). A Figura 4.7 sugere um formato de curva a partir da disposição dos pontos.

Figura 4.7: Percentual de acertos na avaliação devido à habilidade do aluno.

A Figura 4.7 refere-se ao percentual de acertos de cada aluno segundo a sua habilidade. Este gráfico de pontos já sugere uma tendência de crescimento ao longo da escala e dois limites: um inferior (tendendo a 0% pela esquerda) e outro superior (tendendo a 100% pela direita). Considerando o conhecimento cumulativo (TODD & WOLPIN, 2003 e 2007), isto sugere que um aluno com certa habilidade domina os assuntos de seu nível e os assuntos dos níveis inferiores ao seu, conforme interpretação cumulativa da Matriz de Referência, organizada por ano da Educação Básica (SÃO PAULO, 2009).

A Figura 4.7 apresentou os dados empíricos dos alunos que realizaram esta avaliação de matemática, representados por pontos. Para obter uma curva que possa ser aplicável a outros alunos que irão realizar esta prova, é preciso descrever uma curva contínua, que passe por estes pontos (ou próximo), e que esteja definida para todo intervalo da escala de habilidades. É aqui que entram os modelos matemáticos, que serão ajustados com base nos dados, e trarão a probabilidade de acerto para qualquer ponto da escala de habilidades. Apenas como exemplo, a Figura 4.8 apresenta uma curva ajustada²⁴ aos dados do percentual de acertos dos alunos nessa mesma avaliação de matemática.

²⁴ A curva foi ajustada pelo modelo de Rasch que será apresentado à frente.

Figura 4.8: Curva ajustada ao percentual de acertos na avaliação devido a habilidade do aluno.

A partir de um ajuste de curva aos percentuais de acertos dos alunos que realizaram a avaliação, pode-se prever o desempenho geral de um novo aluno, desde que fornecida previamente a sua habilidade. A maneira de fazer este ajuste de curva será abordada na Seção 4.3.1.

Dada a habilidade, pode-se encontrar o desempenho previsto desse aluno por meio da Figura 4.9. Por exemplo: o desempenho esperado desse aluno com habilidade 2,00 logitos é estimado em 88% de acertos nesta avaliação de matemática.

Figura 4.9: Previsão de acertos na avaliação com base na curva ajustada.

4.2. Curva Característica do Item (CCI)

Na seção anterior, a ênfase foi dada na avaliação como um todo. Ou seja, privilegiou-se o desempenho geral dos alunos, ainda obtido de forma clássica: número de acertos em relação ao total. Porém, a TRI vai além do desempenho geral, tornando o item a unidade básica de análise. Se as formas clássicas de análise dos resultados por item pautam-se no percentual de respondentes que acertaram o item e nos percentuais destes por alternativa desse item, com a TRI tornou-se possível verificar a probabilidade de acerto do item mediante a habilidade do respondente.

Deste modo, os procedimentos executados para a avaliação serão realizados para o item, obtendo-se assim, uma curva para cada um desses itens. Esta curva é conhecida como Curva Característica do Item (CCI), semelhante à Figura 4.9, e, segundo Pereira (2004), é a base fundamental da TRI, pois todas as outras construções dependem dela.

Para ajustar essa curva para cada item de uma avaliação recomenda-se que cada um deles seja respondido por pelo menos 100 pessoas, pois, de acordo com Bond & Fox (2010), este seria o **valor mínimo** para se conseguir um ajuste para o modelo de Rasch. Porém, o tamanho da amostra depende do modelo e

geralmente é bem maior do que 100. Segundo Nunes & Primill (2005), que fizeram um estudo sobre o impacto do tamanho da amostra na calibração dos itens, quanto maior o tamanho da amostra maior a confiabilidade das estimativas dos parâmetros dos itens e das capacidades dos sujeitos. Isto se deve porque a variabilidade dos dados está associada ao erro da estimativa e é inversamente proporcional ao tamanho da amostra. Ou seja, quanto maior o número de respondentes, menor será o erro de ajuste da curva característica do item. Seguindo essa sugestão, além dos 13 alunos considerados na curva da avaliação, serão considerados mais 87 alunos, totalizando a cota mínima recomendada por Bond & Fox (2010).

Desta forma, suponha que a avaliação de matemática considerada inicialmente tenha sido aplicada a 100 alunos e que todos responderam a todos os itens. Considere também que todos os alunos tiveram pelo menos um acerto, ou seja, que eles acertaram pelo menos um item da avaliação, e que ninguém tenha acertado todos os itens, mas que o item “m” não teve acertos e o item “f” foi respondido corretamente por todos. Assim, nenhum aluno será retirado da análise a seguir, mas os itens “m” e “f” serão excluídos²⁵.

Como o objetivo desta seção é construir a curva característica para um determinado item é necessário obter a habilidade de cada aluno, uma vez que o modelo depende dessas informações. Realizados os passos de 1 a 6, ou seja, calculado em logitos para cada aluno a sua habilidade e para cada item a sua dificuldade, suponha que a dificuldade de cada item foi a mesma encontrada na Tabela 4.7 e que a Figura 4.10 apresenta uma distribuição dos alunos, segundo suas habilidades estimadas. Neste gráfico, cada aluno é representado por X (identificando apenas os 13 alunos até aqui considerados) e estão agrupados em intervalos de comprimento 0,4 logito, alocados no limite inferior de cada intervalo.

²⁵ Conforme já dito, as metodologias usuais da TRI fazem o tratamento de itens e alunos, de tal forma que nenhuma exclusão tenha que ser realizada.

Figura 4.10: Distribuição dos alunos segundo sua habilidade estimada.

Seja o item “k” o item cuja curva característica deseja-se construir. Se for calculado o percentual de acertos do item “k” para cada habilidade encontrada, certamente será obtido algo semelhante à Figura 4.11.

Figura 4.11: Percentual de acertos no item “k” por habilidade do aluno.

Por exemplo, considere que apenas um aluno possui a habilidade de -2,75 logitos. Se ele acertar o item “k”, o percentual de acertos para essa habilidade

no item será de 100%. Caso ele erre, será de 0%. Considere agora que existam três alunos com a mesma habilidade 0 logito. Se um dentre os três alunos acertar o item “k”, o percentual de acertos para essa habilidade no item será de 33%, se dois acertarem será de 67% e se os três acertarem será de 100%.

Diante dessas considerações, note que na Figura 4.11 há alunos com habilidades baixas que acertaram e alunos com habilidades altas que erraram o item “k”. Porém, a concentração de erros está dentre as habilidades mais baixas e a concentração de acertos está dentre as habilidades mais altas. Nas habilidades intermediárias observe que parece existir uma transição. A partir da habilidade 0 (zero) logito o percentual de acertos começa a subir e a partir da habilidade 2,5 logitos o percentual tende a ficar próxima de 100%. Mesmo assim, quando se observa o todo nota-se que à medida que se eleva a habilidade do aluno, os percentuais de acertos alternam-se de forma brusca, rápida, sem uma continuidade. Há habilidades muito próximas com percentuais de acertos no item “k” muito distintos. Por exemplo, há um aluno com habilidade -2 logitos que acertou o item “k” e outro com habilidade -1,95 logito que errou o item “k”. Os percentuais de acertos desses alunos foram 100% e 0%, respectivamente. Ou seja, eles possuem habilidades bem próximas com desempenho no item completamente opostos um do outro. Como mensurar a probabilidade de acerto de alunos com habilidades próximas às deles, em torno de -2,15 logitos, no item “k”? A probabilidade de acerto para a habilidade -2,15 logitos é alta ou baixa?

A fim de obter essas respostas, os alunos serão agrupados de acordo com a proximidade de suas habilidades e, então para cada grupo de alunos, será computado o número de acertos do item “k”. O intervalo de agrupamento não possui regras definidas, mas deve-se levar em conta a quantidade de alunos em cada intervalo, pois baixas quantidades podem fazer os percentuais variarem muito diante de um acerto ou erro. Conforme já dito, os intervalos aqui considerados terão o comprimento de tamanho 0,4 logito.

Para tanto, será executado o Passo 8, obtendo-se os percentuais de acertos do item “k” para cada faixa de comprimento 0,4 logito da escala de habilidade, conforme a Tabela 4.9.

Tabela 4.9: Percentual de acertos no item “k” por intervalo de habilidades.

Intervalo de Habilidades	Qtde de Alunos que responderam o Item "k"	Número de Acertos no Item "k"	% de Acertos do Item "k"
-4 ---- -3,6	0		
-3,6 ---- -3,2	6	0	0,00%
-3,2 ---- -2,8	8	1	12,50%
-2,8 ---- -2,4	0		
-2,4 ---- -2	4	1	25,00%
-2 ---- -1,6	5	1	20,00%
-1,6 ---- -1,2	7	2	28,57%
-1,2 ---- -0,8	3	1	33,33%
-0,8 ---- -0,4	7	2	28,57%
-0,4 ---- 0	8	3	37,50%
0 ---- 0,4	5	2	40,00%
0,4 ---- 0,8	4	2	50,00%
0,8 ---- 1,2	5	3	60,00%
1,2 ---- 1,6	3	2	66,67%
1,6 ---- 2	4	3	75,00%
2 ---- 2,4	5	3	60,00%
2,4 ---- 2,8	5	4	80,00%
2,8 ---- 3,2	4	4	100,00%
3,2 ---- 3,6	7	6	85,71%
3,6 ---- 4	10	10	100,00%

Agora basta executar o Passo 9 para obter um gráfico de dispersão entre os intervalos de habilidades dos alunos e os percentuais de acertos nesses intervalos da escala de habilidade do item “k”. A partir desses dados, obtém-se a curva que será característica do item “k” (ver Figura 4.12).

Figura 4.12: Curva característica ajustada ao item “k”.

A curva da Figura 4.12 permite verificar a probabilidade de acerto do item “k” em função da habilidade do aluno. Pela curva, note que alunos com habilidade abaixo de 0 (zero) possuem no máximo 45% de probabilidade de acerto e alunos acima dessa habilidade possuem valor da probabilidade maior do que 45%.

Por meio desta curva, pode-se inferir que o aluno com habilidade de 2 logitos, frente ao item “k”, possui cerca de 85% de probabilidade de acertar este item. Este valor está um pouco distante do observado, que para o intervalo que contém a habilidade 2 foi de 60% de acerto dentre os respondentes. Deve-se levar em conta que o intervalo “2 |--- 2,4” contém apenas 5 alunos e que uma resposta correta deles representa 20% de acertos. Se mais um aluno tivesse respondido corretamente, então o valor teria sido de 80%. À medida que o item “k” for sendo respondido por outros alunos, ele poderá ser reajustado, obtendo um melhor ajuste da curva característica.

Sobre a curva ajustada na Figura 4.12, note que ela não passa sobre todos os pontos, mas está próxima a eles. Portanto, esta curva pode ser usada como uma estimativa da probabilidade de acerto do item “k”, dada a habilidade do aluno, e, por ser uma estimativa, há uma margem de erro associada.

Teoricamente, a margem de erro é o valor adicionado e subtraído do valor da estimativa pontual obtida por meio de uma amostragem. Note que esse

procedimento cria um intervalo em torno da estimativa pontual que infere, com certa confiança, a variabilidade dessa estimativa (daí o nome intervalo de confiança). A construção da margem de erro se baseia em 3 componentes: tamanho da amostra, confiança e precisão (que é o tamanho da margem de erro). Eles estão interligados, de modo que, fixando dois deles, o terceiro estará definido. Por exemplo, se a confiança e a precisão forem pré-definidas, o tamanho da amostra necessário fica determinado.

Assim, apesar de estimar pontualmente que o aluno com habilidade de 2 logitos frente ao item “k” possui 85% de probabilidade de acertar este item, na realidade ele pode ter 85% mais ou menos uma margem de erro em pontos percentuais, que é o intervalo de confiança (BOND & FOX, 2010; HOSMER & LEMESHOW, 2000).

Apesar dos passos terem sido executados em específico para o item “k”, eles podem ser feitos para os demais itens. E como todos estão em uma mesma escala, eles podem ser colocados em um mesmo gráfico para comparação.

A Figura 4.13 traz um exemplo hipotético em que todos os itens são colocados juntos para comparação. Como apresentado na Tabela 4.7, a Figura 4.13 mostra o item “a” como o item de maior dificuldade e o item “d” como o de menor dificuldade. Observa-se ainda que para responder corretamente ao item “a” a maior probabilidade está entre os alunos com habilidades altas, já para o item “d” a probabilidade já é alta mesmo para os alunos com habilidades consideradas baixas. Note também, que a curva do item “k” coincide com a do item “e”, já que os dois itens possuem o mesmo grau de dificuldade.

Figura 4.13: Comparativo entre as curvas características ajustadas para cada item.

4.3. Ajuste da Curva Logística

Antes de iniciar o ajuste da curva logística sob uma abordagem pedagógica, vale a pena ressaltar que os modelos da TRI a serem apresentados a seguir consideram itens dicotômicos (onde os possíveis resultados são acerto ou erro). Apesar de se trabalhar nas avaliações em larga escala com itens de múltipla escolha que possuem mais de duas alternativas de respostas, estes itens podem ser tratados como dicotômicos, desde que se considere apenas se o respondente acertou ou errou o item.

Os gráficos apresentados sobre os itens nas seções anteriores levaram em conta o percentual de acertos do item dentre os respondentes que estão em um mesmo intervalo de habilidades, conforme foi apresentado nas Figuras 4.12 e 4.13. Quando se faz a análise de um item, deseja-se também tirar conclusões para um respondente em específico. E interpretar os gráficos dos itens apresentados até aqui para um único respondente, em uma escala que retrata o percentual de acertos de um item, seria no mínimo, estranho. Por exemplo, considerando o item “k” (Figura 4.12), um respondente com habilidade 1 possui 70% de acertos deste item. Mas o item é um só, então o respondente acerta ou

erra o item “k” sem existir a possibilidade de ele acertar 70% do item. Assim, se faz necessário uma compreensão sobre o que representa o cálculo do percentual de acertos por faixa de habilidade.

Pois bem, este percentual de acertos nada mais é do que a razão entre o número de respostas corretas dos respondentes que se encontram em um determinado intervalo de habilidades e o total de respondentes deste mesmo intervalo.

A Tabela 4.9 apresentou o cálculo realizado para o item “k”, em cada intervalo de habilidade, realizado da seguinte forma:

% Acertos no Item “k” (Em determinado intervalo de habilidades) =

$$\frac{\text{Número de respostas corretas do Item dentre os respondentes desse intervalo}}{\text{Número total de respondentes desse intervalo}} * 100\%$$

Essa proporção de respostas corretas dentre os respondentes de um mesmo intervalo de habilidades corresponde à **frequência relativa, a qual pode ser utilizada como estimativa da probabilidade de acerto do item “k”** de um respondente pertencente a este intervalo de habilidades (MAGALHÃES & LIMA, 2013).

Por exemplo, considere o intervalo de habilidade de 0,8 (inclusive) a 1,2 (exclusive) apresentado na Tabela 4.9. A estimativa da probabilidade de acerto do item “k” para um respondente com habilidade neste intervalo é:

$$P(\text{Acerto do Item "k"})_{\text{Hab}[0,8; 1,2[} =$$

$$\frac{\text{Nº de Acertos do item "k" entre os respondentes com habilidades em } [0,8; 1,2[}{\text{Nº Total de respondentes do Item "k" com habilidades em } [0,8; 1,2[} =$$

$$\frac{3}{5} = 0,6 \text{ ou } 60\%$$

Portanto, como o percentual de acertos é uma estimativa da probabilidade de acerto do item, a relação a ser apresentada para interpretação gráfica e ajuste do modelo logístico é a probabilidade de acerto do item em função da habilidade do respondente. Diante disso, faz sentido interpretar que um respondente com

habilidade 1 possui 70% de **probabilidade de acerto** do item “k”. Por este motivo, a probabilidade de acerto do item será a variável resposta da curva a ser ajustada em função da habilidade.

Conforme dito anteriormente, a construção de uma curva contínua **por item** e que cubra toda a escala de habilidades se fez necessária para ser aplicável a qualquer habilidade e para qualquer respondente, sendo desejável que a curva passe o mais próximo possível dos pontos. Assim, será introduzido o modelo matemático associado à Teoria da Resposta ao Item (TRI), cujas raízes encontram-se na década de 1930 (PASQUALI, 2013).

Os gráficos de dispersão apresentados nas figuras 4.7 e 4.12 sugerem um tipo de curva com algumas características especiais, como por exemplo:

- Formato de “S”;
- Crescente em todo intervalo de habilidade, ou seja, uma curva que represente que quanto maior a habilidade do aluno, maior a probabilidade de ele acertar o item;
- Limitada superiormente e inferiormente, pois a probabilidade é um valor entre 0 e 1. Ou seja, à medida que a habilidade do aluno aumenta, a probabilidade de ele acertar o item deve crescer (até 1 ou 100%) e à medida que a habilidade do aluno diminui, a probabilidade de ele acertar o item deve diminuir (até 0 ou 0%).

Levando em consideração essas características, estudiosos do assunto começaram a tentar ajustar aos dados curvas cujas expressões matemáticas já eram bem conhecidas. Segundo Andriola (2009), foi em 1936 que Marion Webster Richardson tentou ajustar a curva da distribuição normal acumulada, conhecida como ogiva normal, às respostas de itens, mas coube a Frederic Mather Lord, em 1952 nos Estados Unidos, também utilizando a ogiva normal, a formulação mais sistemática dos principais conceitos e modelos da TRI. A Figura 4.14 apresenta o gráfico da ogiva normal.

Figura 4.14: Curva da distribuição normal acumulada - ogiva normal.

Ao observar o gráfico da ogiva normal, nota-se que ela possui as características apresentadas nos gráficos exibidos nas figuras 4.7 e 4.12. Porém, a expressão matemática da ogiva normal possui cálculos sofisticados, envolvendo integrais, o que dificultaria a sua aplicação prática. Foi então que, em 1957, Birnbaum deu uma contribuição extremamente importante ao sugerir a substituição da função ogiva normal pela função logística. A expressão da função logística (com comportamento similar) possibilitou um tratamento matemático mais simples e gerou novos procedimentos que permitiram o seu uso na prática (ANDRIOLA, 2009).

Em 1960, na Dinamarca, Georg Rasch aplicou a TRI em testes de desempenho e aptidão (PASQUALI, 2013). O modelo utilizado por Rasch leva em consideração a dificuldade do item (modelo de um parâmetro), análogo ao já desenvolvido no passo a passo da compreensão da abordagem da TRI na Seção 4.1. Apesar da sugestão de Birnbaum ser anterior à aplicação da TRI por Rasch, o modelo de um parâmetro foi transcrito para o modelo logístico somente em 1977, por Benjamin D. Wright (PASQUALI, 2013). As três seções seguintes tratarão do ajuste da função logística.

4.3.1. Modelo Logístico de 1 Parâmetro (Modelo de Rasch)

Já foi dito anteriormente que o modelo logístico possui sua representação gráfica em formato de “S” e é conhecido como curva logística (MARQUEZ & BARROS, 2015). A função dessa curva é bem conhecida e uma das expressões mais simples que descreve a curva logística pode ser dada pela equação:

$$f(x) = \frac{1}{1 + e^{-x}}$$

onde “e” é o conhecido número de Euler (aproximadamente 2,72). A representação gráfica dessa curva encontra-se na Figura 4.15.

Figura 4.15: Curva logística.

A partir das considerações expostas, será proposto o ajuste do modelo logístico de um parâmetro (dificuldade do item) aos dados do item “k”. Este modelo também é conhecido, conforme mencionado na Introdução, como modelo de Rasch. Como o objetivo do ajuste é traçar a curva logística de modo que ela descreva a relação entre a probabilidade de acerto de determinado item com a habilidade do respondente, as funções de probabilidade e logística serão relacionadas de tal forma que $P(\text{Acertar determinado item, dada a habilidade do respondente}) = f(X)$.

Foi visto no Passo 8 da Seção 4.1 que:

- A probabilidade de acerto do item depende não só da habilidade do aluno, mas também da dificuldade deste item;
- A relação entre a habilidade do aluno e a dificuldade do item é dada pela diferença desses dois parâmetros e foi definida como habilidade ajustada.

No Passo 7 da mesma seção, para facilitar a interpretação da diferença entre a habilidade do aluno e o grau de dificuldade do item, foi apresentado um exemplo que fez uma analogia ao esporte salto com vara. Neste exemplo, o maior salto em metros que cada atleta já teve representou sua habilidade e a altura da barra em metros acima do solo a dificuldade do salto. Desta forma, subtraindo da habilidade do atleta a dificuldade do salto, para valores negativos podia-se concluir que o salto não favoreceria ao atleta, mas para valores positivos sim. Ou seja, $P(\text{sucesso}) = f(\text{habilidade} - \text{dificuldade})$.

Levando isso em conta, no exemplo dado no Passo 8 (Seção 4.1), a habilidade da Amanda foi estimada em 2,40 logitos, mas em relação ao item “d”, que possui baixo grau de dificuldade estimado em -2,48 logitos, a habilidade da Amanda passou a ser de 4,88 logitos (Tabela 4.7). Ou seja, a habilidade dela foi ajustada à dificuldade do item.

Na literatura, por exemplo, na área de Psicometria (PASQUALI, 2013; LAVEAULT & GREGOIRE, 2002), costuma-se representar a habilidade do respondente pela letra grega “ θ ” (teta) e a dificuldade do item pela letra “b”. Assim, a habilidade ajustada do respondente é representada por “ $\theta - b$ ”.

Da mesma forma que a variável X, a habilidade ajustada do respondente também pode assumir um valor no intervalo $] - \infty; +\infty[$, favorecendo a substituição, tal que $P(\text{Acertar determinado item, dada a habilidade } (\theta)) = f(\theta - b)$. Sendo assim, a probabilidade de acertar o item, dada a habilidade do respondente, será modelada por:

$$P(\text{Acertar o item} \mid \theta) = f(X) \xrightarrow{X = \theta - b} P(\text{Acertar o item} \mid \theta) = f(\theta - b)$$

$$P(\text{Acertar o item} \mid \theta) = \frac{1}{1 + e^{-(\theta - b)}}$$

Este é o modelo de 1 parâmetro, o chamado modelo de Rasch. Considerando a notação usual, seja Y_i o item dicotômico em questão, onde o valor de Y_i será 1 se o item “i” for respondido corretamente e 0 (zero) caso contrário. A forma geral do modelo pode ser expressa por:

$$P(Y_i = 1 | \theta) = \frac{1}{1 + e^{-(\theta - b_i)}}$$

Onde:

$i = 1, 2, \dots, n$;

n : é o número de itens no teste;

b_i : é o parâmetro de dificuldade do item i ;

e : base dos logaritmos naturais (número de Euler cujo valor é aproximadamente 2,72);

$P(Y_i = 1 | \theta)$: é a probabilidade de o respondente com habilidade (θ) responder corretamente o item i .

Observe-se que, quando a habilidade (θ) for igual ao parâmetro de dificuldade (b) do item, a expressão do modelo corresponderá a $P(Y_i = 1 | \theta) = \frac{1}{2}$. E apesar do parâmetro b , em teoria, poder assumir qualquer valor no intervalo $]-\infty; +\infty[$, na prática os valores observados de b , que de acordo com Hambleton et al. (1991) ficavam no intervalo $[-2; 2]$, mais recentemente, são considerados por Pasquali & Primi (2003) entre -3 e 3.

Na Figura 4.16 são apresentadas as curvas características de cinco itens cujos parâmetros de dificuldade (“ b ”) são: item “1”, $b_1 = 0$; item “2”, $b_2 = -1$; item “3”, $b_3 = 1$; item “4”, $b_4 = 2$; e item “5”, $b_5 = -1,5$. Observe que as curvas destes cinco itens se diferenciam apenas em relação ao deslocamento na escala de habilidades. Isto quer dizer que nesse modelo a dificuldade do item é a única característica considerada como influência no desempenho do aluno. As curvas cujos valores do parâmetro “ b ” são maiores do que 0 (zero) estão à direita do

item “1” (que possui $b_1 = 0$) e as curvas com parâmetro “b” menores que zero estão à esquerda, levando a entender que, quando comparados dois ou mais itens, os localizados à esquerda são os itens “mais fáceis”.

Figura 4.16: Curvas características do modelo de Rasch.

Para visualizar genericamente o comportamento da função referente ao modelo de Rasch, foram realizadas simulações para alguns valores de dificuldade do item, que estão representadas nas Figuras 4.17 e 4.18. A referência base destas simulações é a dificuldade igual a zero logito ($b = 0$), representada pela curva mais escura. Tanto a Figura 4.17 quanto a Figura 4.18 apresentam as curvas da probabilidade de acerto em função da habilidade do respondente para alguns valores de dificuldade do item. Observa-se na Figura 4.17 que a curva tende a se deslocar para a direita à medida que se aumenta a dificuldade do item em 0,2 logito. Ou seja, item mais difícil exige maior habilidade para ser respondido corretamente. Já a Figura 4.18 apresenta o comportamento oposto, em que a curva tende a se deslocar para a esquerda à medida que se diminui a dificuldade do item em -0,2 logito, exigindo menor habilidade para respondê-lo corretamente.

Figura 4.17: Curvas da probabilidade de acerto do item, à medida que se eleva a dificuldade em 0,2 logito.

Figura 4.18: Curvas da probabilidade de acerto do item, à medida que se diminui a dificuldade em -0,2 logito.

Na Seção 4.2, foi apresentado na Figura 4.12 o gráfico dos percentuais de acerto do item “k” em função da habilidade do respondente, juntamente com uma curva ajustada. A construção desse gráfico pode ser realizada a partir da Tabela 4.9, acrescentado uma coluna para o cálculo da função $P(Y_{\text{item "k"}} = 1 \mid \theta)$ no ponto

inferior²⁶ de cada faixa de habilidade, levando em consideração a dificuldade $b = +0,15$ logito do item “k” (Tabela 4.7), conforme apresentado na Tabela 4.10. Os resultados desta coluna foram considerados em termos percentuais para que se possa fazer as comparações.

Tabela 4.10: Estimativa da probabilidade de acerto do item “k” pelo modelo de Rasch.

Intervalo de Habilidades (θ)	Qtde de Alunos que responderam o Item “k”	Número de Acertos no Item “k”	% de Acertos do Item “k”	Estimativa - Modelo de Rasch: $P(Y_{\text{item "k"}} = 1 \theta) = \frac{1}{1 + e^{-(\theta - 0,15)}}$
-4,0 ---- -3,6	0			1,55%
-3,6 ---- -3,2	6	0	0,00%	2,30%
-3,2 ---- -2,8	8	1	12,50%	3,39%
-2,8 ---- -2,4	0			4,97%
-2,4 ---- -2,0	4	1	25,00%	7,24%
-2,0 ---- -1,6	5	1	20,00%	10,43%
-1,6 ---- -1,2	7	2	28,57%	14,80%
-1,2 ---- -0,8	3	1	33,33%	20,59%
-0,8 ---- -0,4	7	2	28,57%	27,89%
-0,4 ---- 0,0	8	3	37,50%	36,59%
0,0 ---- 0,4	5	2	40,00%	46,26%
0,4 ---- 0,8	4	2	50,00%	56,22%
0,8 ---- 1,2	5	3	60,00%	65,70%
1,2 ---- 1,6	3	2	66,67%	74,08%
1,6 ---- 2,0	4	3	75,00%	81,00%
2,0 ---- 2,4	5	3	60,00%	86,41%
2,4 ---- 2,8	5	4	80,00%	90,47%
2,8 ---- 3,2	4	4	100,00%	93,40%
3,2 ---- 3,6	7	6	85,71%	95,48%
3,6 ---- 4,0	10	10	100,00%	96,92%

Na tabela acima, pode-se comparar as estimativas obtidas pelo modelo de Rasch com os dados observados. Por exemplo, no ponto em que a habilidade é igual a 0 (zero), a probabilidade de acerto do item “k” é estimada por:

²⁶ Optou-se por utilizar o ponto inferior da faixa, pois ele fornece a probabilidade mínima da faixa. Poderia ter sido utilizado outros pontos de referência, como por exemplo, o ponto médio ou o ponto máximo da faixa.

- 40% de acordo com os dados observados; e
- 46,26% de acordo com o modelo de Rasch.

Existem técnicas para verificar a qualidade do ajuste do modelo que não serão abordadas no presente texto. Apesar disso, as principais delas estão citadas na Seção 4.3.3 que apresenta o modelo de 3 parâmetros.

A fim de observar o quanto o modelo de Rasch se aproxima dos dados observados, a Figura 4.19 apresenta um gráfico com os dados observados das probabilidades de acerto do item “k”, juntamente com o ajuste obtido por meio do modelo de Rasch. Aplicando este modelo para todo θ , obtém-se a curva característica do item “k”, ajustada pelo modelo de 1 parâmetro.

Figura 4.19: Curva característica do item “k” ajustada às estimativas da probabilidade de acerto.

Dada a curva característica de um item, também se pode obter uma estimativa do parâmetro de dificuldade desse item. Por definição, no modelo de Rasch o parâmetro de dificuldade (b) do item corresponde ao valor da habilidade cuja probabilidade de acerto deste item seja de 50% (PASQUALI & PRIMI, 2003). A

Figura 4.20 mostra a forma de obter esta estimativa da dificuldade do item “k” por meio do gráfico.

Figura 4.20: Representação da dificuldade do item “k” por meio gráfico.

Como último ponto desta seção, será apresentada a relação entre o logaritmo das chances (logito) e a expressão do modelo de Rasch. Esta relação demonstrará a abordagem utilizada no Passo 8 (Seção 4.1), quando se discutiu sobre a habilidade do aluno favorecê-lo ou não a responder corretamente ao item. Para isso, considere $P(Y_i = 1 | \theta) = p$ a probabilidade de responder corretamente ao item i , θ a habilidade do respondente e b_i a dificuldade do item i . Levando em conta a expressão do modelo de Rasch, tem-se que:

$$P(Y_i = 1 | \theta) = \frac{1}{1 + e^{-(\theta - b_i)}} \rightarrow p = \frac{1}{1 + e^{-(\theta - b_i)}} \rightarrow p \cdot (1 + e^{-(\theta - b_i)}) = 1 \rightarrow$$

$$p + p \cdot e^{-(\theta - b_i)} = 1 \rightarrow p \cdot e^{-(\theta - b_i)} = 1 - p \rightarrow e^{-(\theta - b_i)} = \frac{1 - p}{p} \rightarrow$$

$$\ln(e^{-(\theta - b_i)}) = \ln\left(\frac{1 - p}{p}\right) \rightarrow -(\theta - b_i) = \ln\left(\frac{1 - p}{p}\right) \rightarrow$$

$$\theta - b_i = -\ln \left(\frac{1-p}{p} \right) \rightarrow \theta - b_i = \ln \left(\left(\frac{1-p}{p} \right)^{-1} \right)$$

$$\therefore \theta - b_i = \ln \left(\frac{p}{1-p} \right)$$

Na última expressão obtida, note que p é a probabilidade de o indivíduo responder o item i corretamente. Então $(1 - p)$ é a probabilidade de o respondente errar. Observe também, que (habilidade – dificuldade) do item será zero quando $p = 0,5$ ou 50%. Neste caso, a habilidade do aluno não o favorece e nem o desfavorece. A habilidade do aluno só irá favorecê-lo ao acerto do item para valores de p acima de 0,5 ou 50%, e para valores de p abaixo de 0,5 ou 50% a habilidade não o favorecerá.

Outro ponto a observar é a razão $p/(1 - p)$, que representa as chances (odds) do indivíduo responder corretamente ao item i , discutido nos passos 4 e 5 (Seção 4.1), em que a razão entre as probabilidades de acerto e erro exprime a relação de ocorrência entre os eventos, cujo resultado informa se dentre os dois há algum com ocorrência mais provável ou não.

Assim, note que o modelo carrega as informações obtidas a partir dos dados empíricos, mantendo as relações e suposições inicialmente definidas.

A partir do modelo de Rasch, serão introduzidos os modelos de dois e três parâmetros nas seções 4.3.2 e 4.3.3, respectivamente, a fim de chegar aos modelos utilizados nas avaliações externas.

4.3.2. Modelo Logístico de 2 Parâmetros

Antes de introduzir o modelo de 2 parâmetros, é importante ressaltar que o que será feito a seguir não é usual e tem finalidade apenas pedagógica. Sendo assim, foi visto que o modelo de Rasch possui como parâmetro a dificuldade do item (“ b ”) e que esta medida desloca a curva para a direita (quando aumenta o valor de “ b ”) ou para a esquerda (quando diminui o valor de “ b ”). As curvas do modelo de Rasch, com parâmetros de dificuldades distintos, não se cruzam, ou

seja, não possuem intersecção. Isto quer dizer que as curvas descritas pelo modelo de Rasch possuem a mesma inclinação, ou o equivalente a dizer que estes itens possuem a mesma discriminação (mesma capacidade de diferenciar os respondentes).

O modelo de 2 parâmetros, além de considerar a dificuldade, introduz o parâmetro conhecido como discriminação do item. Este parâmetro nada mais é do que a medida do quanto o item pode diferenciar os respondentes em função da sua habilidade, sendo representado pela inclinação da curva no ponto em que a habilidade θ é igual à dificuldade “b” do item. Ou seja, a discriminação (ou o parâmetro “a”) é proporcional à inclinação da reta tangente à curva característica do item no ponto $\theta = b$.

No intuito de explorar o comportamento do parâmetro de discriminação do item, serão apresentados alguns casos hipotéticos de “curvas” para se observar a relação da inclinação da curva com o parâmetro de discriminação. Para isso, suponha que, para um item com dificuldade $b = 0$ (zero) logito, a curva característica correspondente tivesse a representação gráfica conforme apresentado pela Figura 4.21.

Figura 4.21: Gráfico característico do item, discriminação ótima.

Fonte: Baseado em Pasquali & Primi (2003).

No ponto $\theta = b = 0$, em relação ao eixo das abscissas, a inclinação do segmento de reta é máxima. Este seria um caso de discriminação de item ótima, perfeita (PASQUALI & PRIMI, 2003), pois um item que possui uma curva desse tipo deixa claro que os respondentes com habilidades abaixo de zero logito têm probabilidade zero de acertá-lo e respondentes com habilidades acima de zero logito têm probabilidade 1 de acertar. Já em relação aos que possuem habilidade zero logito, nada se pode afirmar com certeza sobre acertar ou errar.

Outro caso é a inclinação do segmento de reta para direita, como apresentado pela Figura 4.22. Note que a “curva” característica do item exibida nessa figura não discrimina tão bem, em determinado intervalo, quanto a “curva” da Figura 4.21.

Figura 4.22: Gráfico característico do item, discriminação boa.

Fonte: Baseado em PASQUALI & PRIMI (2003).

A discriminação é boa para habilidades menores que -0,8 logito, pois os respondentes possuem probabilidade 0 (zero) de acertar o item, e para habilidades iguais ou superiores a +0,8 logito, por terem probabilidade 1 de acertar o item. Para habilidades X entre $[-0,8; 0,8[$ a “curva” característica do item apresenta valores de probabilidade dados pela seguinte equação:

$$\left(\frac{X}{1,6} + \frac{1}{2}\right) * 100\%$$

Esta equação foi obtida a partir da equação da reta (Fig. 4.22):

$$(100\% - 50\%) = m(0,8 - 0)$$

Onde “m” é o coeficiente angular da reta. Assim,

$$m = \frac{0,5}{0,8} = \frac{1}{1,6}.$$

Obtido o coeficiente angular da reta e de posse de um ponto por onde essa reta passa, obtém-se a equação:

$$(Y - 50\%) = \frac{1}{1,6}(X - 0) \rightarrow Y = \left(\frac{X}{1,6} + \frac{1}{2}\right) * 100\%.$$

Esta equação fornece uma probabilidade de acerto crescente no intervalo $[-0,8; 0,8[$, sugerindo que os respondentes com habilidades entre $[-0,8; 0[$ logito possuem menor probabilidade de acertar ($<50\%$) e os respondentes com habilidades $]0; 0,8[$ logito possuem maior probabilidade de acerto ($>50\%$). Já os respondentes com habilidade igual 0 (zero) logito possuem a mesma probabilidade de acertar ou errar (0,5 ou 50%). E neste caso, o 0 (zero) logito também será a dificuldade do item, pois é o ponto em que o respondente possui a mesma probabilidade de acerto ou erro.

A Figura 4.23 apresenta o caso em que a “curva” característica do item não possui inclinação, indicando que a curva não discrimina ninguém. Ou seja, a probabilidade de acerto independe da habilidade do respondente, sendo igual para todos, no caso, de 0,3 ou 30%. Este seria um tipo de caso em que o item não serve para diferenciar, distinguir, discriminar os respondentes, sendo considerado um item sem utilidade no sentido de diferenciá-los.

Não foi explorado nenhum caso com inclinação à esquerda porque isso indicaria que a probabilidade de o respondente acertar ao item diminui com o aumento da habilidade, o que viola o segundo princípio fundamental da TRI (Capítulo 3).

Figura 4.23: Gráfico característico do item, ausência de discriminação.

Fonte: Baseado em PASQUALI & PRIMI (2003).

Nesta seção, foi descrita a discriminação do item relacionando-a a inclinação da curva. Na literatura, costuma-se representar o parâmetro de discriminação do item pela letra “a” (PASQUALI, 2013; LAVEAULT & GREGOIRE, 2002). A Figura 4.24 traz as curvas características de 3 itens que possuem a mesma dificuldade $b_1 = b_2 = b_3 = 1$ logito, porém com parâmetros de discriminação distintos: $a_1 = 0,5$, $a_2 = 1$ e $a_3 = 1,8$. Observe que cada curva possui uma inclinação diferente, remetendo ao fato dos valores dos parâmetros de discriminação serem distintos.

Figura 4.24: Curvas características de itens do modelo de dois parâmetros para itens com a mesma dificuldade.

Observa-se também que, diferentemente do modelo de Rasch, as curvas do modelo de 2 parâmetros se cruzam, existindo um ponto de intersecção entre elas. Como todos os itens possuem a mesma dificuldade, todas as curvas se cruzam no mesmo ponto (Figura 4.24).

Quando a dificuldade é distinta, a curva de um item terá intersecção com as curvas de outros itens em diferentes pontos, conforme apresentado na Figura 4.25. Nesta figura, o item 4 possui parâmetros de dificuldade $b_4 = 1$ e discriminação $a_4 = 0,5$; o item 5 possui $b_5 = -1$ e $a_5 = 1$; e o item 6 possui $b_6 = 2$ e $a_6 = 1,8$. Ressalta-se, novamente, que o valor do parâmetro “a” não corresponde à medida do ângulo, mas, é proporcional à inclinação da reta tangente no ponto $\theta = b$. Nota-se que o ponto de intersecção da curva do item 4 com o item 5 é distinto do ponto de intersecção das curvas dos itens 4 e 6.

Figura 4.25: Curvas características de itens do modelo de dois parâmetros para itens com dificuldades distintas.

A partir da expressão já vista para o modelo de Rasch (Seção 4.3.1), inclui-se o parâmetro (“a”), que representa a discriminação do item, formando-se assim o modelo de dois parâmetros. Assim, a probabilidade de responder corretamente ao item, em função da habilidade θ do respondente, é modelada por:

$$P(\text{Acertar o item} \mid \theta) = \frac{1}{1 + e^{-a(\theta-b)}},$$

onde os parâmetros são “b” (dificuldade do item) e “a” (discriminação do item).

Apesar de o parâmetro “a” poder assumir qualquer valor, ao menos teoricamente, entre $]-\infty; +\infty[$, na prática os valores deste parâmetro, que de acordo com Hambleton et al. (1991) ficavam no intervalo $[0; 2]$, mais recentemente, são considerados por Pasquali & Primi (2003) entre 0 e 3. Valores altos do parâmetro “a” farão com que a curva característica do item *suba* mais rápido, deixando menor o intervalo de habilidades onde ocorre a transição de probabilidades próximas de zero para probabilidades próximas de 1. E valores menores de “a” farão com que a curva característica do item *suba* lentamente, elevando o intervalo de habilidades onde ocorre a transição de probabilidades próximas de zero para probabilidades próximas de 1. Isto pode ser percebido na Figura 4.24, em que o intervalo de transição do item 3 que possui $a = 1,8$ é, aproximadamente, $[-1,8; 3,2]$; do item 2 que possui $a = 1$ é $[-4; 4]$ e do item 1 com $a = 0,5$ o intervalo é maior do que o intervalo do item 2.

Desta forma, quanto maior for o valor do parâmetro “a” (ou quanto maior a inclinação da curva do item), melhor o item discriminará os respondentes. E, obviamente, quanto menor o valor de “a”, pior será a discriminação do item.

Considerando o item dicotômico e escrevendo o modelo de 2 parâmetros na notação usual, em que o valor de Y_i será 1 se o item i for respondido corretamente e 0 (zero) caso contrário, a forma geral do modelo de 2 parâmetros pode ser expressa por:

$$P(Y_i = 1 \mid \theta) = \frac{1}{1 + e^{-a_i(\theta-b_i)}}$$

Onde:

$i = 1, 2, \dots, n$;

n : é o número de itens no teste;

a_i : é o parâmetro de discriminação do item i ;

b_i : é o parâmetro de dificuldade do item i ;

e: base dos logaritmos naturais (número de Euler cujo valor é aproximadamente 2,72);

$P(Y_i = 1 | \theta)$: é a probabilidade do respondente com habilidade (θ) responder corretamente o item i .

Hambleton et al. (1991) trazem uma abordagem pedagógica para estimação dos parâmetros que será descrita a seguir. Sendo assim, considere os dados empíricos do item “k” (% de acertos do item) apresentados na Tabela 4.11. A partir do modelo de 2 parâmetros, tem-se que:

Se $P(\text{Acertar o item} | \theta) = p$, então $p = \frac{1}{1 + e^{-a(\theta - b)}}$ e $p \cdot (1 + e^{-a(\theta - b)}) = 1 \rightarrow$

$$\frac{1-p}{p} = e^{-a(\theta - b)} \rightarrow \frac{p}{1-p} = e^{a(\theta - b)}.$$

Aplicando o logaritmo natural em ambos os lados da equação, segue que:

$$\ln\left(\frac{p}{1-p}\right) = a(\theta - b)$$

Para determinar os parâmetros “a” e “b” do modelo, escolhe-se arbitrariamente dois pontos na escala de habilidade. A partir dos dados empíricos (Tabela 4.11), foram selecionados, como exemplo, os pontos $\theta = -1,80$ e $\theta = 1,80$. Nestes pontos, os valores da probabilidade “p” de acerto do item “k” são, respectivamente, 0,20 e 0,75.

Então as equações a serem resolvidas são da seguinte forma:

$$\ln\left(\frac{0,20}{0,80}\right) = a(-1,80 - b) \text{ e } \ln\left(\frac{0,75}{0,25}\right) = a(1,80 - b)$$

As equações acima formam um sistema de equações. Subtraindo a primeira da segunda, tem-se que:

$$\ln\left(\frac{0,75}{0,25}\right) - \ln\left(\frac{0,20}{0,80}\right) = 1,80a - (-1,80a).$$

Resolvendo a equação, obtém-se:

$$a = 0,69.$$

E substituindo na segunda equação, tem-se:

$$b = 0,21.$$

Note que o valor do parâmetro “b” é próximo do valor utilizado no modelo de Rasch ($b = 0,15$ logito). A diferença em módulo é $|0,06|$ logito.

Portanto, uma estimativa da curva do modelo de 2 parâmetros será:

$$P(\text{Acertar o item} \mid \theta) = \frac{1}{1 + e^{-0,69(\theta - 0,21)}}.$$

Conforme já dito, esta foi apenas uma abordagem pedagógica. A estimação usual dos parâmetros do modelo exige tópicos especiais de cálculo avançado que não serão abordados aqui. O Leitor interessado poderá se aprofundar no modo de obter as equações de estimativas dos parâmetros, consultando a bibliografia existente sobre a TRI, como por exemplo, Hambleton et al. (1991) e Andrade et al. (2000).

Em continuidade a abordagem utilizada, a construção da curva do modelo de 2 parâmetros pode ser realizada da mesma forma que foi feita para o modelo de Rasch, também apresentado na Tabela 4.11. A última coluna desta tabela leva em consideração a dificuldade $b = +0,21$ logito do item “k” (Tabela 4.7) e a discriminação $a = 0,69$.

Para ter uma referência comparativa, as estimativas, referentes aos modelos de Rasch e de 2 parâmetros apresentados na Tabela 4.11, foram calculadas no ponto inferior da faixa de habilidades. Desta forma, a probabilidade de acerto do item “k” para a habilidade $-3,2$ logitos é:

- 12,50% de acordo com os dados empíricos;
- 3,39% de acordo com o modelo de Rasch;
- E 8,68% de acordo com o modelo logístico de 2 parâmetros.

Tabela 4.11: Estimativa da probabilidade de acerto do item “k” pelo modelo de 2 parâmetros.

Intervalo de Habilidades (θ)	Qtde de Alunos que responderam o Item “k”	Número de Acertos no Item “k”	% de Acertos do Item “k”	Estimativa - Modelo de Rasch: $P(Y_{\text{item "k"}} = 1 \theta) = \frac{1}{1 + e^{-(\theta - 0,15)}}$	Estimativa - Modelo de 2 Parâmetros: $P(Y_{\text{item "k"}} = 1 \theta) = \frac{1}{1 + e^{-0,69(\theta - 0,21)}}$
-4,0 ---- -3,6	0			1,55%	5,19%
-3,6 ---- -3,2	6	0	0,00%	2,30%	6,73%
-3,2 ---- -2,8	8	1	12,50%	3,39%	8,68%
-2,8 ---- -2,4	0			4,97%	11,14%
-2,4 ---- -2,0	4	1	25,00%	7,24%	14,17%
-2,0 ---- -1,6	5	1	20,00%	10,43%	17,87%
-1,6 ---- -1,2	7	2	28,57%	14,80%	22,29%
-1,2 ---- -0,8	3	1	33,33%	20,59%	27,43%
-0,8 ---- -0,4	7	2	28,57%	27,89%	33,25%
-0,4 ---- 0,0	8	3	37,50%	36,59%	39,63%
0,0 ---- 0,4	5	2	40,00%	46,26%	46,38%
0,4 ---- 0,8	4	2	50,00%	56,22%	53,27%
0,8 ---- 1,2	5	3	60,00%	65,70%	60,04%
1,2 ---- 1,6	3	2	66,67%	74,08%	66,44%
1,6 ---- 2,0	4	3	75,00%	81,00%	72,29%
2,0 ---- 2,4	5	3	60,00%	86,41%	77,47%
2,4 ---- 2,8	5	4	80,00%	90,47%	81,92%
2,8 ---- 3,2	4	4	100,00%	93,40%	85,66%
3,2 ---- 3,6	7	6	85,71%	95,48%	88,73%
3,6 ---- 4,0	10	10	100,00%	96,92%	91,21%

No ponto -3,2 logitos da escala de habilidade, a diferença entre a estimativa do modelo de Rasch e o dado empírico é de 9,11 pontos percentuais. Já entre o modelo de dois parâmetros e o dado empírico é de 3,82 pontos percentuais. Portanto, neste ponto, o modelo logístico de 2 parâmetros se aproximou mais dos dados empíricos. Observando os demais pontos, é possível perceber que, na maioria das faixas, o modelo de 2 parâmetros se aproxima mais do que o modelo de Rasch. Da mesma forma que foi dito para o modelo de Rasch, existem técnicas para verificar a qualidade do ajuste do modelo. Estas técnicas não serão abordadas no presente texto, mas as principais delas estão citadas na Seção 4.3.3 que apresenta o modelo de 3 parâmetros.

A fim de facilitar esta comparação, a Figura 4.26 apresenta os valores observados das probabilidades de acerto do item “k”, juntamente com os ajustes obtidos por meio do modelo de Rasch e pelo modelo de 2 parâmetros.

Observe na Figura 4.26 que a introdução do parâmetro de discriminação do item no modelo fez com que, visualmente, a curva se aproximasse mais dos dados empíricos. Da mesma forma que a análise sobre a qualidade do ajuste não será abordada neste texto, não será apresentado o método usual de estimativa dos parâmetros dos modelos da TRI, podendo ser obtido em Hambleton et al. (1991), Andrade et al. (2000) ou Bond & Fox (2010), sendo este último apenas para os modelos de Rasch.

Figura 4.26: Curvas ajustadas às estimativas da probabilidade de acerto do item “k” pelos modelos de Rasch e de 2 parâmetros.

4.3.3. Modelo Logístico de 3 Parâmetros

Da mesma forma que foi ressaltada na seção anterior, a abordagem a seguir possui fim apenas pedagógico. A estimação usual dos parâmetros dos modelos da TRI exige tópicos especiais de cálculo avançado e poderá ser encontrada nas diversas bibliografias existente sobre a TRI, como por exemplo, Hambleton et al. (1991) e Andrade et al. (2000).

Tanto o modelo de Rasch quanto o modelo de 2 parâmetros não consideram na equação do modelo o acerto ao acaso, pois levam em conta somente a dificuldade (no caso do modelo de Rasch) ou a dificuldade e a capacidade de discriminação (no caso do modelo de 2 parâmetros), dada a habilidade do respondente. Segundo Hambleton et al. (1991), a suposição de não haver acerto ao acaso é plausível nos casos em que o item é aberto (a resposta é livre), ou então quando o teste de múltipla escolha não é muito difícil para os respondentes.

Nos testes usuais de múltipla escolha, em que é comum apresentar um certo número de respostas dentre as quais somente uma é correta, o acerto ao acaso está presente e não deveria ser descartado. Assim, o modelo de três parâmetros, além de considerar a dificuldade e a discriminação do item, incorpora a probabilidade de respondentes com baixa habilidade acertarem o item no “chute”, o chamado acerto ao acaso. Por exemplo, os itens do SARESP apresentados na etapa de elaboração de itens (Seção 2.3) possuem 4 alternativas de resposta, mas somente uma é correta. Para cada um destes itens, a probabilidade de acerto de um respondente que escolhe uma alternativa ao acaso, considerando que ele desconheça a resposta correta, é 1 em 4 ou 1/4 (25%). Por isso, quando os itens são de múltipla escolha, todos os respondentes, independentemente de sua habilidade, já possuem uma probabilidade mínima de acerto ao acaso do item. Mesmo com itens de múltipla escolha, usam-se modelos com duas possibilidades: acertar o item ou errar o item (modelo dicotômico).

Devido ao modelo de três parâmetros possuir a influência simultânea da dificuldade, discriminação e do acerto ao acaso, as curvas características de cada item assumirão formatos distintos que dependerão da combinação dos parâmetros. Para ilustrar esta situação, a Figura 4.27 apresenta as curvas características de cinco itens cujos respectivos parâmetros estão disponibilizados na Tabela 4.12 (a = discriminação, b = dificuldade, c = acerto ao acaso).

Observe que os itens 2 e 4 estão deslocados para esquerda e os itens 1, 3 e 5 para a direita (Figura 4.27). O que provoca este efeito de deslocamento é o

parâmetro de dificuldade. Repare na Tabela 4.12 que os itens 2 e 4 possuem os menores valores do parâmetro “b”. Dessa forma, tendo como referência a habilidade 0 (zero), os itens mais fáceis sempre estarão deslocados à esquerda, enquanto itens mais difíceis estarão deslocados à direita.

Figura 4.27: Curvas características de itens do modelo de três parâmetros.

Tabela 4.12: Parâmetros dos itens apresentados na Figura 4.27.

Item	Parâmetro		
	a	b	c
"1"	0,5	1	0
"2"	1	-1	0
"3"	1,8	2	0,1
"4"	1,2	-1,5	0,2
"5"	1	1,5	0,25

Em relação à discriminação (inclinação da curva), nota-se que os itens 2 e 4 possuem os valores do parâmetro “a” muito próximos e por isso aparentam ter a “mesma” inclinação. Já os itens 1 e 3 possuem inclinações bem distintas, em que o parâmetro “a” do item 3 chega a ser 3,6 vezes maior do que o parâmetro “a” do item 1. Repare que isto implica numa subida rápida do item 3 em um curto

intervalo da escala de habilidade, enquanto que o item 1 sobe mais lentamente. Assim, uma análise da Figura 4.27 e da Tabela 4.12 permite dizer que os itens 2 e 4 diferenciam os respondentes de uma maneira semelhante e com grau de dificuldade próximos. O item 5 diferencia os respondentes de forma parecida com os itens 2 e 4, mas possui grau de dificuldade maior. O item 3 é o que melhor diferencia os respondentes e o item 1 o que pior diferencia.

Por fim, destaca-se o efeito do parâmetro de acerto ao acaso, representado na literatura pela letra “c” (PASQUALI, 2013; LAVEAULT & GREGOIRE, 2002), quando este é diferente de zero. Se nos modelos anteriores a probabilidade de acerto tendia a zero à medida que a habilidade (θ) diminuía, no modelo de três parâmetros quando a habilidade (θ) tende a $-\infty$ a probabilidade mínima de acerto do item tende ao parâmetro “c”. Por exemplo, para os itens 3, 4 e 5, que possuem o parâmetro de acerto ao acaso diferente de zero, note que as curvas características desses itens (Figura 4.27) tendem ao respectivo parâmetro “c” (Tabela 4.12) quando θ vai para $-\infty$.

O modelo de 3 parâmetros será apresentado a partir do modelo de 2 parâmetros, introduzindo o parâmetro “c” da probabilidade de acerto ao acaso do item. Uma proposta de introdução desse parâmetro foi apresentada por Lord (ARAÚJO, ANDRADE & BORTOLOTTI, 2009), com a sugestão de incluir o parâmetro de forma aditiva. Mas adicionar simplesmente uma quantidade ao modelo de dois parâmetros provocaria valores incompatíveis para probabilidade, pois levaria a existência de valores fora do intervalo [0; 1]. A título de exemplo, suponha um item que possua 10 alternativas e somente uma destas está correta. Neste caso, a probabilidade de acerto ao acaso seria 0,10. Se adicionar de forma simples esta quantidade no modelo, o intervalo de probabilidade será:

$$0,10 + \text{mínimo} \left(\frac{1}{1 + e^{-a(\theta - b)}} \right) = 0,10 + 0 = 0,10$$

e

$$0,10 + \text{máximo} \left(\frac{1}{1 + e^{-a(\theta - b)}} \right) = 0,10 + 1 = 1,10.$$

Portanto, o intervalo seria [0,10; 1,10], incompatível com os valores que a probabilidade pode assumir.

Para corrigir este problema, faz-se uma correção na expressão do modelo de dois parâmetros. Como a probabilidade de acerto do item deve estar no intervalo [0; 1] e já existe uma probabilidade de acerto ao acaso de 0,10, então o valor máximo que, neste caso, o modelo de dois parâmetros deve atingir é $1 - 0,10 = 0,90$, que nada mais é do que o complementar da probabilidade de acerto ao acaso. Desta forma, basta multiplicar a expressão do modelo de 2 parâmetros pelo complementar da probabilidade de acerto ao acaso, para que a nova expressão tenha valores no intervalo [0; 1]. Ou seja,

$$0,10 + (1 - 0,10) \times \text{mínimo} \left(\frac{1}{1 + e^{-a(\theta - b)}} \right) = 0,10 + 0,9 \times 0 = 0,10$$

e

$$0,10 + (1 - 0,10) \times \text{máximo} \left(\frac{1}{1 + e^{-a(\theta - b)}} \right) = 0,10 + 0,9 \times 1 = 1,$$

o que produz valores no intervalo [0,10; 1] que está contido no intervalo [0; 1].

Considerando o item dicotômico (acerto ou erro) e escrevendo o modelo de 3 parâmetros na notação usual, onde o valor de Y_i será 1 se o item i for respondido corretamente e 0 (zero) caso contrário, vem a forma geral do modelo de 3 parâmetros que pode ser expressa por:

$$P(Y_i = 1 | \theta) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta - b_i)}}$$

Onde:

$i = 1, 2, \dots, n$;

n : é o número de itens no teste;

a_i : é o parâmetro de discriminação do item i ;

b_i : é o parâmetro de dificuldade do item i ;

c_i : é o parâmetro de acerto ao acaso do item i ;

e: base dos logaritmos naturais (número de Euler cujo valor é aproximadamente 2,72);

$P(Y_i = 1 | \theta)$: é a probabilidade do respondente com habilidade (θ) responder corretamente o item i .

Este é o modelo conhecido como o modelo logístico de três parâmetros. Mas, Baker (2011) chama a atenção para o fato do modelo de 3 parâmetros não ser mais, tecnicamente, um modelo logístico. Isto porque a introdução do parâmetro de acerto ao acaso feita a partir do modelo de 2 parâmetros provocou a perda de algumas das boas propriedades matemáticas da função logística.

A introdução do acerto ao acaso também afetou a forma de se obter o valor do parâmetro “b” por meio do gráfico. Nos modelos de Rasch e de 2 parâmetros, por definição, o valor do parâmetro “b” é estabelecido como sendo a habilidade cuja probabilidade de acerto do item é de 0,5 ou 50%, pois é o ponto de inflexão (onde há alteração de concavidade) da curva do modelo. No modelo de 3 parâmetros, o valor do parâmetro “b” continua sendo a habilidade relacionada com o ponto de inflexão da curva do modelo. Porém, a probabilidade de acerto deste ponto não é mais fixa e dependerá do valor de “c”.

Na verdade, o valor do parâmetro “b” pode ser obtido (por meio do gráfico da curva) pela metade da soma entre o máximo e o mínimo do intervalo de probabilidades. Por exemplo, pegue o intervalo de probabilidades obtido quando foi considerado o acerto ao acaso igual a 0,10: [0,10; 1]. Observe o cálculo:

$$\frac{\text{máximo}[0,10;1] + \text{mínimo}[0,10;1]}{2} = \frac{1 + 0,10}{2} = 0,55.$$

A forma geral para encontrar a probabilidade que levará ao parâmetro “b”, qualquer que seja o valor do parâmetro “c”, é dada por:

$$c + (1 - c) \times 0,5 = \frac{(1 + c)}{2}.$$

Embora o parâmetro “c” represente o acerto ao acaso e que costuma ser quantificado pelo número de alternativas corretas sobre o total de alternativas do item, no modelo de 3 parâmetros ele é estimado em conjunto com os outros parâmetros (dificuldade e discriminação) e pode assumir valores menores ou

maiores do que a estimativa usual. Isto ocorre devido ao ajuste da curva aos dados empíricos, a fim de que esta curva se aproxime o máximo possível dos dados observados.

Para construir a curva do modelo de 3 parâmetros para o item “k”, será utilizada a mesma abordagem pedagógica da seção anterior, que foi baseada na abordagem de Hambleton et al. (1991). Desta forma, a partir do modelo de 3 parâmetros, tem-se que:

$$\text{Se } P(\text{Acertar o item} \mid \theta) = p, \text{ então } p = c + (1 - c) \frac{1}{1 + e^{-a(\theta - b)}} = \frac{(1 + e^{-a(\theta - b)})c + 1 - c}{1 + e^{-a(\theta - b)}} \rightarrow$$

$$p(1 + e^{-a(\theta - b)}) = 1 + ce^{-a(\theta - b)} \rightarrow (p - c)e^{-a(\theta - b)} = 1 - p \rightarrow$$

$$\frac{1 - p}{p - c} = e^{-a(\theta - b)} \rightarrow \frac{p - c}{1 - p} = e^{a(\theta - b)}.$$

Aplicando o logaritmo natural em ambos os lados da equação, segue que:

$$\ln\left(\frac{p - c}{1 - p}\right) = a(\theta - b)$$

Para determinar os parâmetros “a”, “b” e “c” do modelo, foram escolhidos três pontos na escala de habilidade. A fim de facilitar as contas, os pontos selecionados foram $\theta = \{-1,80; 0; 1,80\}$. Nestes pontos, os valores da probabilidade “p” de acerto do item “k” são, respectivamente, 0,20; 0,40 e 0,75 (Tabela 4.13).

Então as equações a serem resolvidas são da seguinte forma:

$$(I) \ln\left(\frac{0,20 - c}{0,80}\right) = a(-1,80 - b)$$

$$(II) \ln\left(\frac{0,40 - c}{0,60}\right) = a(0 - b)$$

$$(III) \ln\left(\frac{0,75 - c}{0,25}\right) = a(1,80 - b)$$

As equações I, II e III formam um sistema de equações. A partir de (II), tem-se que:

$$e^{-ab} = \left(\frac{0,40 - c}{0,60} \right) \cdot \text{(IV)}$$

De (III):

$$e^{1,8a} \cdot e^{-ab} = \left(\frac{0,75 - c}{0,25} \right) \cdot \text{(V)}$$

Substituindo (IV) em (V):

$$e^{1,8a} = \left(\frac{(0,75 - c)0,60}{0,25(0,40 - c)} \right) \rightarrow$$

$$e^{-1,8a} = \left(\frac{0,25(0,40 - c)}{(0,75 - c)0,60} \right) \cdot \text{(VI)}$$

De (I), segue:

$$e^{-1,8a} \cdot e^{-ab} = \left(\frac{0,20 - c}{0,80} \right) \cdot \text{(VII)}$$

Substituindo (IV) e (VI) em (VII), segue:

$$\left(\frac{0,25(0,40 - c)}{(0,75 - c)0,60} \right) \left(\frac{0,40 - c}{0,60} \right) = \left(\frac{0,20 - c}{0,80} \right) \cdot$$

Desenvolvendo esta equação, obtém-se uma equação do segundo grau:

$$0,16c^2 - 0,182c + 0,022 = 0$$

As raízes desta equação são: 1 e 0,1375. Como o valor 1 para o parâmetro “c” implica no respondente acertar o item com certeza no “chute”, o que para um item de múltipla escolha implica em todas as alternativas estarem corretas, este valor de probabilidade ao acaso não faz sentido. Portanto o valor de estimativa para o parâmetro “c” (acerto ao acaso) é 0,1375.

Substituindo $c = 0,1375$ em (VI), obtém-se a estimativa do parâmetro “a”:

$$e^{-1,8a} = \frac{0,25(0,40 - 0,1375)}{(0,75 - 0,1375)0,60} \rightarrow a = 0,96.$$

E para encontrar a estimativa de “b”, basta substituir $c = 0,1375$ e $a = 0,96$ em (III):

$$\ln\left(\frac{0,75 - 0,1375}{0,25}\right) = 0,96(1,80 - b) \rightarrow b = 0,87.$$

Portanto, uma estimativa da curva do modelo de 3 parâmetros para o item “k” será:

$$P(\text{Acertar o item} \mid \theta) = 0,1375 + (1 - 0,1375) \frac{1}{1 + e^{-0,96(\theta - 0,87)}}.$$

Dando continuidade na abordagem pedagógica utilizada, os valores estimados para a probabilidade de acerto do item “k”, obtidas pelo modelo de 3 parâmetros, são apresentados na última coluna da Tabela 4.13. Nesta tabela, também são apresentados os modelos de Rasch e de 2 parâmetros para comparação.

Tabela 4.13: Estimativa da probabilidade de acerto do item “k” pelo modelo de 3 parâmetros.

Intervalo de Habilidades (θ)	Qtde de Alunos que responderam o Item “k”	Número de Acertos no Item “k”	% de Acertos do Item “k”	Estimativa - Modelo de Rasch: $P(Y_{\text{item "k"}} = 1 \theta) = \frac{1}{1 + e^{-(\theta - 0,15)}}$	Estimativa - Modelo de 2 Parâmetros: $P(Y_{\text{item "k"}} = 1 \theta) = \frac{1}{1 + e^{-0,69(\theta - 0,21)}}$	Estimativa - Modelo de 3 Parâmetros: $P(Y_{\text{item "k"}} = 1 \theta) = \frac{1}{0,1375 + (1 - 0,1375) \frac{1}{1 + e^{-0,96(\theta - 0,87)}}$
-4,0 ---- -3,6	0			1,55%	5,19%	14,55%
-3,6 ---- -3,2	6	0	0,00%	2,30%	6,73%	14,91%
-3,2 ---- -2,8	8	1	12,50%	3,39%	8,68%	15,45%
-2,8 ---- -2,4	0			4,97%	11,14%	16,22%
-2,4 ---- -2,0	4	1	25,00%	7,24%	14,17%	17,33%
-2,0 ---- -1,6	5	1	20,00%	10,43%	17,87%	18,91%
-1,6 ---- -1,2	7	2	28,57%	14,80%	22,29%	21,12%
-1,2 ---- -0,8	3	1	33,33%	20,59%	27,43%	24,15%
-0,8 ---- -0,4	7	2	28,57%	27,89%	33,25%	28,20%
-0,4 ---- 0,0	8	3	37,50%	36,59%	39,63%	33,42%
0,0 ---- 0,4	5	2	40,00%	46,26%	46,38%	39,84%
0,4 ---- 0,8	4	2	50,00%	56,22%	53,27%	47,31%
0,8 ---- 1,2	5	3	60,00%	65,70%	60,04%	55,43%
1,2 ---- 1,6	3	2	66,67%	74,08%	66,44%	63,65%
1,6 ---- 2,0	4	3	75,00%	81,00%	72,29%	71,40%
2,0 ---- 2,4	5	3	60,00%	86,41%	77,47%	78,21%
2,4 ---- 2,8	5	4	80,00%	90,47%	81,92%	83,86%
2,8 ---- 3,2	4	4	100,00%	93,40%	85,66%	88,31%
3,2 ---- 3,6	7	6	85,71%	95,48%	88,73%	91,68%
3,6 ---- 4,0	10	10	100,00%	96,92%	91,21%	94,15%

Comparando os dados observados com os resultados obtidos pelos três modelos, a probabilidade de acerto do item “k” para a habilidade 0 (zero) logito é:

- 40% de acordo com os dados empíricos;
- 46,26% com o modelo de Rasch;
- 46,38% com o modelo logístico de 2 parâmetros e;
- 39,84% com o modelo logístico de 3 parâmetros.

Neste ponto da escala de habilidade, a menor diferença entre o dado empírico e as estimativas obtidas é dada pelo modelo de 3 parâmetros. Uma comparação visual, onde é possível verificar que o modelo de 3 parâmetros é o que mais se aproxima dos dados empíricos, é apresentada na Figura 4.29.

Não será abordado aqui o tópico relacionado a técnicas de verificação da qualidade do ajuste e nem técnicas de seleção de melhor modelo ajustado aos dados, pois isto fugiria dos objetivos deste texto. Porém, como este assunto é importante, pois quando o modelo não se encaixa nos dados as estimativas dos parâmetros do item são tendenciosas e podem levar a conclusões erradas, serão citados alguns testes utilizados na verificação do ajuste, para leitura do Leitor interessado no tópico.

Dentre os testes clássicos para modelos cumulativos unidimensionais da TRI, pode-se citar o teste Qui-Quadrado (em suas variações propostas por Bock (1972, apud LENHARD, 2013, p.11), Yen (1981, apud LENHARD, 2013, p.11) e Mckinley & Mils (1985, apud LENHARD, 2013, p.11)), que, verificadas as suposições do modelo, visa comparar os dados observados com os dados estimados pelo modelo. E dentre os testes alternativos, destaca-se:

- Estatística condicionada ao escore total do teste: compara a proporção de respostas corretas com a proporção esperada em subgrupos condicionados ao escore total (ORLANDO & THISSEN, 2000, apud LENHARD, 2013, p.14);
- Abordagem da regressão logística: verifica o ajuste logístico para respostas dicotômicas (MAIR et al., 2008; apud LENHARD, 2013, p.15);
- Análise gráfica dos resíduos: confronta a distribuição esperada com a observada, por meio dos resíduos (YIN, 2007, apud LENHARD, 2013, p.15);

- Raiz da Integral do Erro Quadrático: compara a CCI não paramétrica com a CCI paramétrica ajustada (DOUGLAS & COHEN, 2001, apud LENHARD, 2013, p.16).

O Leitor interessado em se aprofundar no assunto poderá consultar também Hambleton e Swaminathan (1985, apud HAMBLETON et al., 1991), que recomendam uma análise do ajuste do modelo com base em três tipos de evidência (validade dos pressupostos da TRI, avaliação das propriedades do modelo e precisão das previsões obtidas) e Maydeu-Olivares (2013), que traz uma visão geral sobre métodos de avaliação da qualidade do ajuste de modelos da TRI.

A Figura 4.28 apresenta os valores observados das probabilidades de acerto do item “k”, juntamente com o ajuste obtido por meio do modelo de 3 parâmetros.

Figura 4.28: Curva ajustada às estimativas da probabilidade de acerto do item “k” para o modelo de 3 parâmetros.

Já a Figura 4.29 apresenta os gráficos dos valores observados das probabilidades de acerto do item “k”, juntamente com os ajustes obtidos por meio do modelo de Rasch, do modelo de 2 parâmetros e do modelo de 3 parâmetros.

Figura 4.29: Curvas ajustadas às estimativas da probabilidade de acerto do item “k” para os modelos de Rasch, de 2 e de 3 parâmetros.

Note que a introdução do parâmetro de acerto ao acaso no modelo fez com que, visualmente, a curva se aproximasse melhor do que a curva do modelo de 2 parâmetros aos dados empíricos. Destaca-se novamente que **as fórmulas usuais para obter os parâmetros da TRI e a análise sobre a qualidade do ajuste não serão abordados neste texto**, podendo ser obtidas em Hambleton et al. (1991), Andrade et al. (2000) ou Bond & Fox (2010), sendo este último apenas para os modelos de Rasch.

Este modelo de 3 parâmetros é o modelo utilizado pelas avaliações em larga escala. Dependendo do tipo da avaliação em larga escala, serão divulgados todos os itens, por exemplo, o ENEM, ou somente alguns, como é o caso do SARESP. Porém, este último costuma apresentar os parâmetros dos itens divulgados em relatórios pedagógicos, para cada disciplina avaliada, com objetivo de serem analisados no contexto escolar.

4.4. Situações de Estimação dos Parâmetros dos Itens e das Habilidades

Apenas para conhecimento, serão apresentadas as três situações de estimação de parâmetros que a TRI possui. Conforme visto nas seções anteriores, a probabilidade de o respondente acertar um item depende de sua habilidade e dos parâmetros (dificuldade, discriminação e o acerto ao acaso) deste item. Portanto, para obter a probabilidade de acerto do item é imprescindível que se obtenha as estimativas desses parâmetros. Segundo Andrade et al. (2000), a estimação desses parâmetros, também conhecida como calibração, pode ser dividida em três situações:

- i. As habilidades dos respondentes são conhecidas e deseja-se estimar apenas os parâmetros dos itens;
- ii. Os parâmetros dos itens são conhecidos e deseja-se estimar apenas as habilidades;
- iii. Deseja-se estimar os parâmetros dos itens e as habilidades simultaneamente.

As duas primeiras situações são as mais fáceis, pois uma vez que se sabe as habilidades, basta estimar os parâmetros dos itens. E quando se tem os parâmetros dos itens já conhecidos, basta aplica-los para obter as habilidades.

Para a terceira situação existem duas abordagens: estimação conjunta (habilidades e parâmetros dos itens) e estimação em duas etapas. Na primeira, Birbaum (1968; apud ANDRADE et al.; 2000, p. 28) propôs um processo conhecido como “vai e volta”. Ele inicia o processo com estimativas “grosseiras” das habilidades, como por exemplo, os resultados padronizados²⁷, e então faz-se a estimação dos parâmetros dos itens considerando as habilidades conhecidas. A partir das estimativas obtidas dos parâmetros dos itens, faz-se as estimativas das habilidades dos respondentes. Este processo é realizado até que se obtenha a convergência dos parâmetros dos itens e das habilidades dos respondentes.

²⁷ A padronização é feita da seguinte forma: toma o desempenho do aluno, subtrai a média dos dados de todos os alunos e divide o resultado pelo desvio padrão destes mesmos dados.

Porém, a forma de se obter as estimativas dos parâmetros na primeira abordagem pode levar a um problema de inconsistência na estimação dos parâmetros dos itens e habilidades (ANDRADE et al., 2000). Por este motivo, Bock & Lieberman (1970; apud ANDRADE et al., 2000, p. 29) propuseram a estimação dos parâmetros em duas etapas: estima-se os parâmetros dos itens, com base na existência de uma distribuição associada à habilidade da população em estudo²⁸, e em seguida, de posse dos parâmetros dos itens, estima-se as habilidades dos respondentes de forma individual. Para se aprofundar mais sobre as abordagens da terceira situação, consultar Andrade et al. (2000).

²⁸ Construída a partir da frequência acumulada e aproximada por uma distribuição contínua. Para detalhes, ver Anderson (1980; apud ANDRADE, TAVARES & VALE; 2000, p. 51).

Capítulo 5

Cálculo da Habilidade de Um Respondente

Geralmente as avaliações em larga escala se utilizam da segunda situação, em que os parâmetros dos itens são conhecidos (“a”, “b” e “c”), ou seja, já se encontram calibrados, e deseja-se estimar apenas as habilidades dos respondentes. Isto porque o banco de itens vem sendo construído há algum tempo e periodicamente são realizadas as reposições de itens “queimados” (divulgados ao público em geral).

Assim, dado que os itens já estão calibrados, será iniciado o processo de obtenção das habilidades dos respondentes. Para isso, será retomado o exemplo da Seção 4.1, que abordou uma avaliação externa de matemática respondida por alunos de um mesmo ano da educação básica. Mas, para simplificar este processo, neste primeiro momento, considere que a prova dessa avaliação externa possua apenas **4 itens para serem respondidos: os itens “g”, “h”, “i” e “k”**.

As curvas características desses itens, ajustadas pelo modelo de 3 parâmetros, são apresentadas na Figura 5.1, juntamente com os parâmetros da TRI. Para facilitar o que será exposto a seguir, observe que os itens estão dispostos em ordem crescente da dificuldade. Ou seja, o item “g” é o mais fácil e o item “h” o mais difícil. Note também que os valores do parâmetro “c” são estimados pelo modelo e nem sempre as estimativas serão valores como 0,10 (1 correta dentre 10 alternativas); 0,20 (1 correta dentre cinco) ou 0,25 (1 correta dentre 4). Podem ser valores próximos.

Figura 5.1: Curvas características dos quatro itens considerados na composição da avaliação.

Uma vez que para obter a habilidade do respondente leva-se em conta seus acertos e erros, dada a curva que descreve a probabilidade de acerto do item em função da habilidade, o complementar desta curva descreverá a probabilidade de erro do item. Veja o exemplo realizado apenas para o item “k” na Figura 5.2. Observe que a soma das curvas corresponde a 100%.

Figura 5.2: Curvas das probabilidades de acerto e de erro para o item “k”.

De modo similar, a Figura 5.3 apresenta as curvas que descrevem a probabilidade de erro de cada um dos quatro itens considerados.

Figura 5.3: Curvas das probabilidades de erro para cada um dos quatro itens considerados na composição da avaliação.

A fim de compreender o cálculo da habilidade de um respondente, serão apresentados dois casos de desempenho para um único respondente no teste de avaliação com

quatro itens. Apenas o caso 1 será explorado em detalhes, sendo o caso 2 apresentado de forma consolidada.

Caso 1: o respondente acerta os itens “g” e “i”, mas erra os itens “k” e “h”.

- a) Se a habilidade desse respondente for igual a 2 logitos, qual a probabilidade dessa situação acontecer?

Considerando a habilidade do respondente, por meio das figuras acima, pode-se obter as probabilidades de acerto (Figura 5.1) e erro (Figura 5.3) para cada item da avaliação. Por exemplo, a probabilidade do respondente com habilidade de 2 logitos acertar o item “g” é 0,96, ou numa linguagem matemática $P(Y_{“g”} = 1 \mid \theta = 2) = 0,96$. E a probabilidade deste respondente errar o item “h” é 0,44, ou $P(Y_{“h”} = 0 \mid \theta = 2) = 0,44$. As probabilidades dos itens “i” e “k” estão postas abaixo:

- $P(Y_{“i”} = 1 \mid \theta = 2) = 0,94$
- $P(Y_{“k”} = 0 \mid \theta = 2) = 0,22$

Todas essas probabilidades descritas acima correspondem à probabilidade individual de cada item. Para saber a habilidade do respondente é necessário calcular a probabilidade conjunta, que nesse caso é a probabilidade de $P(Y_{“g”} = 1, Y_{“i”} = 1, Y_{“k”} = 0, Y_{“h”} = 0 \mid \theta = 2)$.

Para calcular essa probabilidade conjunta, será usada a hipótese de independência condicional entre os itens da TRI, apresentada no Capítulo 3. Desta forma, tem-se que:

$$\begin{aligned}
 &P(Y_{“g”} = 1, Y_{“i”} = 1, Y_{“k”} = 0, Y_{“h”} = 0 \mid \theta = 2) \stackrel{\text{Hip.de Independência}}{=} \\
 &P(Y_{“g”} = 1 \mid \theta = 2) * P(Y_{“i”} = 1 \mid \theta = 2) * P(Y_{“k”} = 0 \mid \theta = 2) * P(Y_{“h”} = 0 \mid \theta = 2) = \\
 &0,96 * 0,94 * 0,22 * 0,44 = 0,09
 \end{aligned}$$

- b) E se a habilidade desse respondente for igual a 0 logito?

Utilizando as figuras 5.1 e 5.3, tem-se que:

$$P(Y_{\text{"g"}} = 1, Y_{\text{"i"}} = 1, Y_{\text{"k"}} = 0, Y_{\text{"h"}} = 0 \mid \theta = 0) \quad \hat{=}$$

$$P(Y_{\text{"g"}} = 1 \mid \theta = 0) * P(Y_{\text{"i"}} = 1 \mid \theta = 0) * P(Y_{\text{"k"}} = 0 \mid \theta = 0) * P(Y_{\text{"h"}} = 0 \mid \theta = 0) =$$

$$0,85 * 0,55 * 0,60 * 0,83 = 0,23$$

Calculando, por meio de programas computacionais (por exemplo: R, BILOG, MULTILOG, XCALIBRE), a probabilidade de essa situação acontecer para todo θ e construindo um gráfico com os resultados, obtém-se algo muito similar ao apresentado pela Figura 5.4.

Note que a maior probabilidade de ocorrer o resultado, ou seja, de o respondente acertar os itens “g” e “i” e errar os itens “k” e “h”, é observada para a habilidade 0,4 logito. Então, esta é a habilidade estimada para o respondente com este padrão de resposta.

Figura 5.4: Curva da probabilidade conjunta do caso 1.

Caso 2: o respondente acerta os itens “g” e “k”, mas erra os itens “i” e “h”.

Calculando a probabilidade de essa situação acontecer para todo θ e comparando com os resultados obtidos no Caso 1 por meio gráfico (Figura 5.5), verifica-se que, apesar do respondente também ter obtido 2 acertos, a estimativa da habilidade

dele é menor do que a obtida no Caso 1. Isto se deve, por não haver uma “coerência” nas respostas no Caso 2 e então a TRI “penaliza” o respondente por ele ter errado um item mais fácil e acertado um item mais difícil. É como se o respondente tivesse acertado o item mais difícil por causa da probabilidade de acerto ao acaso e não porque ele sabia a resposta.

Figura 5.5: Comparativo entre as curvas da probabilidade conjunta dos casos 1 e 2.

Um resumo dos possíveis casos dessa avaliação com 4 itens é apresentado por meio do Quadro 5.1, onde os casos estão dispostos por ordem decrescente da estimativa da habilidade do respondente. O Leitor perceberá que o quadro contempla apenas 14 possibilidades e não 16, pois, conforme os comentários abaixo do Quadro 5.1, as possibilidades de acertar ou errar todos os itens não permitem estimar a habilidade do aluno.

Quadro 5.1: Casos possíveis de uma avaliação com 4 itens.

Possibilidade	Item "g"	Item "i"	Item "k"	Item "h"	Total de Acertos	Habilidade (θ)
1ª	Erra	Acerta	Acerta	Acerta	3	2,2
2ª	Acerta	Acerta	Erra	Acerta	3	2
3ª	Acerta	Acerta	Acerta	Erra	3	1,5
4ª	Acerta	Erra	Acerta	Acerta	3	1,2
5ª	Erra	Acerta	Acerta	Erra	2	0,5
6ª	Acerta	Acerta	Erra	Erra	2	0,4
7ª	Erra	Acerta	Erra	Acerta	2	-0,1
8ª	Acerta	Erra	Acerta	Erra	2	-0,3
9ª	Acerta	Erra	Erra	Acerta	2	-1,4
10ª	Acerta	Erra	Erra	Erra	1	-1,6
11ª	Erra	Erra	Acerta	Acerta	2	< -4
12ª	Erra	Acerta	Erra	Erra	1	< -4
13ª	Erra	Erra	Acerta	Erra	1	< -4
14ª	Erra	Erra	Erra	Acerta	1	< -4

Alguns comentários sobre o Quadro 5.1:

- Tanto a 1ª quanto a 3ª possibilidade possuem “coerência²⁹” nas respostas. Porém, como na 1ª possibilidade os acertos foram nos itens mais difíceis, então é mais provável que a habilidade do respondente seja maior;
- A 11ª, 12ª, 13ª e 14ª possibilidades estão com habilidade menor que -4 logitos devido a esses quatro itens não conseguirem discriminar habilidades muito baixas. Observe na Figura 5.1 que para habilidade igual ou menor que -2,5 logitos, as inclinações dos itens “i”, “k” e “h” são praticamente zero (não há discriminação). Se o objetivo for identificar respondentes com habilidades baixas, deve-se acrescentar itens mais fáceis na avaliação;
- As possibilidades de acertar ou errar todos os itens não foram contempladas no quadro, pois nada se sabe dizer sobre a habilidade de quem erra tudo ou acerta tudo. Isto foi abordado na Seção 4.1, que trata da construção do modelo a partir dos dados empíricos, quando foi necessário excluir os dados dos alunos Natan (errou todos os itens) e o Ophir (acertou todos os itens). Ou seja, para

²⁹ Diz-se *coerência* a um padrão de resposta, os acertos consecutivos de uma sequência de itens ordenados pela dificuldade. Ou seja, é coerente esperar que os respondentes que acertaram questões difíceis, também acertem questões fáceis. Por exemplo: se um indivíduo acerta um item que envolve equação do segundo grau, é esperado que ele acerte também um outro sobre equação do primeiro grau com uma incógnita.

estimar a habilidade de um respondente é necessário que ele acerte e erre itens. Para contornar esse problema, define-se uma regra de concessão de notas aos respondentes com esses padrões de repostas. No ENEM, por exemplo, como os itens da prova já estão calibrados e podem ser colocados em ordem crescente, a avaliação possui um valor mínimo e um valor máximo. Desta forma, o respondente que errar tudo ganhará o valor mínimo da avaliação e quem acertar tudo ganhará o valor máximo da avaliação (BRASÍLIA, 2012);

- Pela Teoria Clássica dos Testes (TCT), respondentes com padrão de respostas idêntico à 1ª possibilidade ou à 4ª possibilidade possuem o mesmo desempenho, pois o número de acertos é igual: 3 acertos. Ou seja, 3/4 ou 75% ou 7,5 em uma escala de 0 a 10. Já a TRI consegue diferenciá-los, pois leva em conta a coerência dos acertos do respondente.

Enfim, no modelo de 3 parâmetros existe uma correspondência entre o número de acertos e a habilidade do respondente. Porém, o percentual de itens acertados não possui relação biunívoca³⁰ com a escala de habilidade, pois, outros fatores também são considerados, tais como os parâmetros dos itens acertados e o padrão de resposta.

Apresentada a forma de estimar a habilidade dos respondentes a partir de itens já calibrados, é chegado o momento de calcular as estimativas das habilidades dos 15 alunos (Tabela 4.1) de um mesmo ano da educação básica que realizaram a avaliação externa de matemática e analisar estes resultados. Porém, como os itens “f” (acertado por todos) e “m” (errado por todos) não puderam ser calibrados, considerou-se a avaliação composta por 12 itens: “a”, “b”, “c”, “d”, “e”, “g”, “h”, “i”, “j”, “k”, “l” e “n”. As curvas características, desses 12 itens calibrados pelo modelo de três parâmetros, são apresentadas na Figura 5.6.

³⁰ Diz-se de uma correspondência entre os elementos de dois conjuntos tal que a cada elemento de um corresponda um e só um do outro. In Dicionário Priberam da Língua Portuguesa [em linha], 2008-2013, <https://www.priberam.pt/dlpo/biun%C3%ADvoca> [consultado em 18-01-2017].

Figura 5.6: Curva característica dos itens que compõem a avaliação de matemática.

Por meio da Figura 5.6 é possível notar a composição da avaliação em relação a sua dificuldade. Observe que quase 60% dos itens possuem dificuldade acima de 0 logito, implicando em uma avaliação que visa diferenciar habilidades altas. Apenas 4 itens diferenciam habilidades baixas.

Considerando o padrão de resposta de cada respondente (Tabela 4.1), referente aos 12 itens calibrados, calcula-se a probabilidade desse padrão de resposta acontecer para todo " θ ", da mesma forma que foi feito para os casos 1 e 2 anteriormente. A Figura 5.7 apresenta as curvas referentes a cada padrão de resposta dos participantes da avaliação. Observe que, por não ser uma curva acumulativa como é a curva do acerto do item, as curvas do padrão de respostas não possuem o mesmo máximo.

Assim, para garantir a visualização das curvas, o eixo das ordenadas está com valores de 0% a 10%.

Figura 5.7: Curva da probabilidade de ocorrência do padrão de resposta do respondente em relação à habilidade.

Note que as curvas referentes aos padrões de resposta do Ophir e do Natan não possuem um ponto de máximo, mas tendem a 100% (para habilidades altas no caso do Ophir e para habilidades baixas no caso do Natan). Isto ocorre porque o primeiro acertou tudo e o segundo errou tudo. Como as habilidades do Ophir e do Natan não puderam ser estimadas pela avaliação, então se adota como estimativa dessas habilidades, respectivamente, o máximo e o mínimo da avaliação. O máximo e o mínimo da avaliação não são, necessariamente, a maior e a menor habilidade obtida por um respondente, mas sim a maior e a menor habilidade encontrada dentre todos os possíveis padrões de resposta da avaliação. Por exemplo, no Quadro 5.1, onde

foram apresentados todos os casos possíveis de uma avaliação com 4 itens, o máximo dessa avaliação é 2,2 logitos e o mínimo é -1,6 logito.

O valor da estimativa da habilidade dos 15 respondentes encontra-se no Quadro 5.2. Observe que apesar do Ivis, a Laís e a Maria possuírem a mesma habilidade, a Laís acertou menos que os outros dois. Outro ponto para reflexão é a falta de equivalência entre as diferenças de dois respondentes em relação ao número de acertos e habilidades. Repare que o Fábio teve 10 acertos: um a menos que a Amanda e um a mais que o Enzo. Porém, a habilidade da Amanda é 1,6 logito maior e a habilidade do Enzo é apenas 0,4 logito menor que a do Fábio.

Quadro 5.2: Estimativa das habilidades dos respondentes da avaliação de matemática.

Respondente	Acertos	Estimativa da Habilidade
Amanda	11/12	3,6
Bruno	1/12	-3,2
Carlos	5/12	-0,4
Débora	10/12	2,4
Enzo	9/12	1,6
Fábio	10/12	2
Gilson	4/12	-0,8
Hélen	2/12	-2,4
Ivis	8/12	0,8
Joice	6/12	0,4
Kelly	2/12	-2,8
Laís	7/12	0,8
Maria	8/12	0,8
Natan	0/12	-3,2*
Ophir	12/12	3,6**

* Mínimo da Avaliação

** Máximo da Avaliação

Apenas para salientar, como as habilidades do Ophir e do Natan foram obtidas pelo mínimo e máximo da avaliação, caso seja necessário estimá-las deve-se aplicar uma avaliação com itens mais fáceis para o Natan e outra com itens mais difíceis para o Ophir.

Antes de transformar as habilidades estimadas em proficiência, interpretá-las e divulgá-las, o que será mostrado no próximo capítulo, faz-se uma validação dos resultados obtidos. Esta validação consiste em verificar se todos os respondentes tiveram sua habilidade estimada, se houve problema em algum item, se as regras de cálculo e divulgação de resultados foram cumpridas e no caso do SARESP, por exemplo, se todas as escolas tinham, no mínimo, 18 (dezoito) alunos por ano/série a serem avaliados (SÃO PAULO, 2014a).

Capítulo 6

Construção da Escala de Proficiência e Divulgação dos Resultados

6.1. Construção da Escala de Proficiência

Após o cômputo da habilidade do aluno, também tida como desempenho do aluno, é necessário entender o que significa essa habilidade. Comparar a habilidade do aluno apenas com a habilidade média da avaliação não é suficiente para dizer se o resultado foi bom ou ruim e muito menos inferir sobre o que o aluno já sabe e o que precisa aprender. A estimativa da habilidade por si serve apenas para classificar os alunos em ordem decrescente de desempenho, que pode ser o objetivo de algumas avaliações que são usadas em processos seletivos para vagas em faculdades e universidades. Mesmo assim, toda avaliação educacional, que tem como base a TRI, tem a necessidade de fornecer uma escala de medida, uma “régua”. É esta escala, chamada de escala de proficiência, que torna possível a interpretação pedagógica e facilita a compreensão dos resultados.

Numa escala de proficiência não existe um zero absoluto e nem uma unidade de medida absoluta, ou seja, um ponto que indique a ausência de conhecimento e uma quantidade específica de conhecimento que serve de padrão para eventuais comparações com outras medidas. Segundo Rabelo (2011), como na TRI a habilidade pode assumir qualquer valor entre $-\infty$ a $+\infty$, na definição de uma escala de medida é preciso estabelecer **uma origem** e **uma unidade de medida** que represente, respectivamente, **o valor médio** e **o desvio padrão** das habilidades dos respondentes. Segundo o mesmo autor, como o número de participantes nas avaliações em larga escala é muito grande, é costume fazer a suposição de que a distribuição das proficiências deles em relação à sua habilidade média é praticamente simétrica em relação à média.

A construção da escala, estabelecendo-se uma origem e uma unidade de medida, resolve outro problema chamado de *falta de identificabilidade do modelo*. Isto ocorre

quando mais de um conjunto de parâmetros produzem o mesmo valor de habilidade dada pelo modelo (ANDRADE et al. (2000) e HAMBLETON et al. (1991)). A definição de uma escala evita problemas desse tipo, permitindo colocar todos os anos avaliados em uma mesma escala, eliminando os problemas que se teria ao encontrar habilidades numericamente iguais e com significados diferentes, como por exemplo, o 0 logito do 7º ano com o 0 logito do 9º ano.

Até a seção anterior foi utilizada a escala (0,1), em que 0 (zero) é a média e 1 é o desvio padrão. Mesmo esta escala sendo ilimitada tanto inferiormente quanto superiormente, os valores do parâmetro θ (habilidade) estimados para os respondentes costumam ficar entre -5 e 5. Isto porque na prática, conforme a Figura 5.6, a partir de valores menores que -5 as probabilidades de acerto do item já estão muito próximas do seu respectivo parâmetro de acerto ao acaso ("c") e a partir de valores maiores que 5 estão bem próximas de 1.

Segundo Andrade et al. (2000), se não existir uma ligação entre os valores da escala (0,1) e os conteúdos contemplados na avaliação, pouca coisa se poderá dizer sobre um indivíduo que possui uma habilidade maior do que a do outro. Conforme os mesmos autores exemplificam, as únicas coisas que se pode dizer, sobre um respondente que obteve uma habilidade de, por exemplo, 1,80 na escala (0,1) frente a outro com habilidade -0,50, é que:

- O primeiro deve possuir um conhecimento muito maior do conteúdo avaliado do que o segundo;
- A habilidade do primeiro está 1,80 desvios-padrão acima da média da população avaliada e a habilidade do segundo está 0,50 desvios padrão abaixo da média dessa mesma população;
- Nada se pode afirmar a respeito do que o respondente com habilidade 1,80 sabe a mais do que aquele com habilidade -0,50.

Fatos como estes, e também visando facilitar a compreensão dos valores numéricos, levaram à construção da escala de proficiência, permitindo a interpretação pedagógica das habilidades estimadas. Os valores escolhidos para a escala de proficiência são arbitrários (SÃO PAULO, 2009). Por exemplo, o ENEM utiliza uma escala que varia

de 0 a 1.000, em que a média é 500 e o desvio padrão é 100. Já o SAEB possui uma escala de 0 a 500, comum a todos os anos da educação básica, com média igual a 250 e desvio padrão 50. Ambas as escalas estão referenciadas no desempenho dos alunos da 8ª série do ano de 1997 (BRASÍLIA, 2006). O SARESP utiliza a escala do SAEB desde 2007 (SÃO PAULO, 2009).

Mas, apesar da arbitrariedade na escolha dos números, a escala deve manter a relação de ordem existente entre os seus pontos (RABELO, 2011). Ou seja, se na escala (0, 1) um aluno estiver a dois desvios padrões acima da média, na nova escala ele também deve estar. A conversão dos valores originais para essa nova escala ocorre por meio de uma transformação linear, ou seja, por uma relação direta e pré-definida, conforme exemplo da Figura 6.1, realizada para o 9º ano do ensino fundamental que possui média referenciada em 250. Vale ressaltar que os valores da escala não possuem qualquer relação com as escalas usuais utilizadas em sala de aula, como por exemplo, notas de 0 (zero) a 10, ou com conceitos de A a E.

Figura 6.1: Representação da transformação linear entre escalas (0,1) e SAEB para o 9º ano do ensino fundamental.

Um exemplo de equação matemática que representa a transformação linear apresentada na Figura 6.1 é $F(\theta) = 250 + 50\theta$, onde θ é a habilidade do respondente na escala (0, 1), 250 é a média e 50 é o desvio padrão da proficiência dos alunos da 8ª série do ano de 1997, atualmente, equivalente ao 9º ano do ensino fundamental.

Supondo que os 15 respondentes que fizeram a avaliação externa de matemática são alunos do 9º ano do ensino fundamental, por meio da transformação linear acima, pode-se obter as proficiências deles na escala SAEB (Quadro 6.1).

Quadro 6.1: Proficiência dos respondentes da avaliação de matemática na escala SAEB.

Respondente	Acertos	Estimativa da Habilidade	Proficiência (Escala SAEB)
Amanda	2/12	3,6	430
Bruno	3/12	-3,2	90
Carlos	2/12	-0,4	230
Débora	8/12	2,4	370
Enzo	10/12	1,6	330
Fábio	4/12	2	350
Gilson	6/12	-0,8	210
Hélen	10/12	-2,4	130
Ivis	5/12	0,8	290
Joice	4/12	0,4	270
Kelly	12/12	-2,8	110
Laís	0/12	0,8	290
Maria	0/12	0,8	290
Natan	0/12	-3,2	90
Ophir	12/12	3,6	430

Mesmo com as proficiências obtidas para cada respondente, ainda não é possível falar sobre o que cada um sabe dentre os conteúdos de matemática. Então, alguns pontos ou níveis da escala foram escolhidos, também de forma arbitrária, para serem interpretados sob a ótica do que os alunos sabem e são capazes de fazer. A interpretação desses pontos tem como base as respostas dadas pelos respondentes aos itens, sendo feita por área de conhecimento ou disciplina, devendo conter o ponto 250 e distarem entre si de meio desvio padrão (25) (BRASÍLIA, 2006).

A escala do SAEB, apresentada em níveis para interpretação pedagógica, está baseada no método de ancoragem de itens (PERRY, 2009). Este método consiste na fixação de itens em pontos ou níveis escolhidos da escala para serem interpretados. A partir do momento em que o item é ancorado em determinado nível, sendo denominado de item-âncora, ele passa a integrar um conjunto de itens que irão caracterizar este nível da escala.

Conforme já mencionado, o SARESP utiliza a escala SAEB e sua ancoragem de itens segue a metodologia descrita por Beaton & Allen (1992, apud RABELO, 2011, p. 50). Esta metodologia ancora os seus itens sob os critérios apresentados em Andrade et al. (2000), em que um item é considerado âncora se e somente se as três condições abaixo forem consideradas satisfeitas simultaneamente:

- O percentual de acerto do item no nível considerado deve ser de pelo menos 65%;
- O percentual de acerto do item no nível imediatamente anterior ao considerado deve ser de no máximo 50%;
- A diferença entre a proporção de indivíduos que acertaram o item no nível considerado e o nível imediatamente anterior deve ser de pelo menos 30%.

Para compreender um processo de ancoragem, será dado um exemplo. Suponha uma escala de proficiência que varia de 0 a 500, cujos pontos escolhidos para serem interpretados são 100, 150, 200, 250, 300, 350 e 400. Considere também os itens “j” e “a” que compõem o exemplo da avaliação externa de matemática já citado. Será mostrado que estes dois itens podem ser considerados âncoras nessa avaliação, podendo ser utilizados para interpretar pedagogicamente a escala de proficiência. A Figura 6.2 traz as curvas características desses itens.

Figura 6.2: Exemplo de itens âncoras.

Fonte: Baseado em Andrade et al. (2000).

Para mostrar que os itens “j” e “a” são âncoras dessa avaliação, respectivamente, dos pontos 200 e 400, será verificado todos os critérios pré-estabelecidos para ancoragem de itens. Assim, tem-se que:

A. Item “j” é ancora do ponto 200:

1. $P(Y_{“j”} = 1 \mid \text{Proficiência} = 200) = 0,77 \geq 0,65$;
2. $P(Y_{“j”} = 1 \mid \text{Proficiência} = 150) = 0,43 < 0,50$;
3. $P(Y_{“j”} = 1 \mid \text{Proficiência} = 200) - P(Y_{“j”} = 1 \mid \text{Proficiência} = 150) = 0,77 - 0,43 = 0,34 \geq 0,30$.

B. Item “a” é ancora do ponto 400:

1. $P(Y_{“a”} = 1 \mid \text{Proficiência} = 400) = 0,76 \geq 0,65$;
2. $P(Y_{“a”} = 1 \mid \text{Proficiência} = 350) = 0,39 < 0,50$;
3. $P(Y_{“a”} = 1 \mid \text{Proficiência} = 400) - P(Y_{“a”} = 1 \mid \text{Proficiência} = 350) = 0,76 - 0,39 = 0,38 \geq 0,30$.

Uma vez que o item foi ancorado em um ponto ou nível, ele será utilizado para descrever de forma pedagógica este nível da escala. Por exemplo, considere o item

“a”, cuja habilidade explorada é a H31 (calcular áreas de polígonos de diferentes tipos, com destaque para os polígonos regulares), dado a seguir:

- a. Um terreno tem a forma de um trapézio retangular ABCD, conforme mostra a figura, e as seguintes dimensões: $AB = 34\text{m}$, $BC = 24\text{m}$ e $CD = 24\text{m}$.

O perímetro e a área desse terreno são, respectivamente:

- a) 100m e 700m^2 .
- b) 116m e 816m^2 .
- c) 108m e 696m^2 .
- d) 116m e 696m^2 .
- e) 108m e 816m^2 .

Como foi dado ao aluno um trapézio retangular (composto por um quadrado e um triângulo reto) e pedido que calculasse a área dessa figura, a ferramenta ensinada ao aluno para encontrar a hipotenusa do triângulo é o teorema de Pitágoras (mesmo que alguns resolvam de outra maneira, a maioria utilizará o que lhe foi ensinado).

Conforme visto anteriormente, o item “a” foi ancorado no nível 400. Assim, a interpretação pedagógica do nível 400 da escala de proficiência contemplará as informações sobre a habilidade e conteúdo explorados pelo item “a”. Por exemplo, os respondentes dessa avaliação com proficiência igual ou superior a 400:

- Calculam:
 - Perímetro de polígonos;
 - Área de polígonos como triângulos, quadrados e trapézios.
- Identificam:
 - O número de lados de polígonos apresentados em figuras;
 - Quadrados e triângulos formando um trapézio.
- Aplicam e resolvem:
 - O Teorema de Pitágoras.

Note que a descrição é feita para compreensão pedagógica do nível 400 e não somente para os respondentes que acertaram o item “a”. Pode ter respondentes com proficiência acima de 400 que erraram o item “a”. E mesmo errando este item, eles podem dominar essa habilidade porque outros itens, ancorados no nível 400 ou acima, podem avaliar outras habilidades e outros conteúdos que possuem como pré-requisito a habilidade explorada pelo item “a”.

Realizada a interpretação pedagógica da escala, a partir da proficiência do aluno, o professor consegue verificar as habilidades e competências por ele dominadas e os conteúdos aprendidos. Por exemplo, o Ophir e a Amanda obtiveram uma proficiência de 430, e dominam a habilidade H31. Já a Débora, que obteve proficiência igual a 370, ainda não possui esta habilidade consolidada.

A partir do desempenho obtido por cada aluno (ou da média obtida pela turma ou escola) e da interpretação pedagógica de cada nível da escala de proficiência, é possível obter as habilidades dominadas e as que ainda precisam ser desenvolvidas. No entanto, as avaliações que possuem como foco o sistema de ensino e usam a TRI divulgam apenas as médias do ano avaliado por escola, e não possuem como objetivo a interpretação para cada aluno. Na prática, após analisar o resultado da avaliação em conjunto com a escala de proficiência, o professor deverá identificar em sua turma quais habilidades esperadas ainda não foram consolidadas e elaborar atividades para desenvolvê-las.

A escala de proficiência de matemática do SARESP é comum ao 5º, 7º e 9º anos do Ensino Fundamental e 3ª série do Ensino Médio e deve ser interpretada de forma cumulativa. Ou seja:

“... os alunos que estão situados em um determinado ponto dominam não só as habilidades associadas a esse ponto, mas também as proficiências descritas nos pontos anteriores.” (SÃO PAULO, 2014b, p. 179).

Os pontos da escala de proficiência em matemática do SARESP, escolhidos para serem interpretados, são: 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 450 e 475. A descrição desses pontos tem como base os resultados de desempenho dos alunos na prova de matemática realizada e as habilidades detalhadas na Matriz de

Referência. Por este motivo, a escala de proficiência do SARESP é atualizada após cada aplicação da avaliação e sua descrição para interpretação pedagógica dos níveis da escala encontra-se como anexo do relatório pedagógico (SÃO PAULO, 2014b). A título de exemplo, o Quadro 6.2 traz a descrição da escala de matemática para o ponto 400.

Quadro 6.2: Descrição da escala de matemática para o ponto 400.

Neste ponto, os alunos do 9º ano do Ensino Fundamental e 3ª série do Ensino Médio, além das habilidades descritas no ponto anterior:

- **Calculam:**
 - a medida de um segmento de uma figura de um Tangran desenhado em um quadrado de 20 cm de lado, comparando medidas de lados das demais figuras desenhadas.
- **Identificam:**
 - as coordenadas do ponto de interseção de duas retas que definem um sistema de equações do 1º grau;
 - as coordenadas de pontos específicos utilizando o plano cartesiano.
- **Resolvem problema** envolvendo:
 - o cálculo do volume de um cilindro;
 - o cálculo das áreas de um quadrado e de um hexágono regular, dadas as medidas de seus lados;
 - metro cúbico e litro;
 - análise combinatória. (numero possível de placas de automóvel em um a determinada configuração) - contagem (arranjo).
- **Simplificam** expressão que envolve o quadrado da soma e o quadrado da diferença entre x e y .

Neste ponto ainda, os alunos de 3ª série do Ensino Médio também:

- **Calculam:**
 - medidas de comprimento de um triângulo, usando as relações de proporcionalidade identificadas na sua representação gráfica;
- **Identificam:**
 - a equação de uma reta apresentada em um plano cartesiano;
 - a equação da circunferência, dada a medida do seu raio;
 - a expressão matemática de uma função exponencial definida em linguagem corrente;
 - a inequação associada à região sombreada de um plano desenhado no sistema cartesiano, e vice-versa;
 - a representação gráfica em um sistema cartesiano, de uma circunferência, dada a sua equação;
 - o ângulo formado pelos meridianos que determinam dois fusos horários no Brasil;
 - no plano de Argand Gauss, o resultado da adição e da subtração de 2 números complexos;
 - o quadrante do afixo de um número complexo, (dada a definição de afixo).
- **Resolvem problema envolvendo:**
 - o lançamento de dois dados simultaneamente para cálculo de probabilidade;
 - comparação entre volume de cones;
 - a área superficial de uma pirâmide;
 - comparação da área superficial de cilindros;
 - as relações entre coeficientes e raízes de uma equação do 2º grau;
 - o cálculo da taxa de crescimento de uma variável que cresce exponencialmente de acordo com uma função dada;
 - relações entre coeficientes e raízes de uma equação do 3º grau, dadas estas relações para uma equação na forma genérica;

- o cálculo do perímetro de uma figura composta por um retângulo e dois semicírculos;
- razões trigonométricas no triângulo retângulo.

Fonte: Relatório Pedagógico de Matemática do SARESP (SÃO PAULO, 2014b).

Visto que o desempenho dos alunos da 8ª série do ano de 1997 (BRASÍLIA, 2006) é a referência da escala do SAEB, falta verificar como os outros anos foram posicionados na escala. Para que todos os anos estejam em uma mesma escala, realiza-se um processo chamado de equalização. Segundo Andrade et al. (2000), equalizar nada mais é do que tornar comparável. O que no caso da educação básica, equalizar significa colocar em uma mesma métrica os parâmetros de itens provenientes de respondentes de vários anos avaliados, tornando-os assim comparáveis. Isto é o que permitirá dizer que um aluno domina mais habilidades e quais em relação a outro aluno, seja do mesmo ano ou de qualquer outro ano do ensino básico.

Mas, para que os anos avaliados de cada disciplina estejam em uma mesma métrica, as provas devem ter itens em comum. Por exemplo, na prova do 7º ano do ensino fundamental deve existir itens do 5º ano e também itens do 9º ano. O uso de itens comuns entre provas dos anos avaliados da educação básica permite que todos os parâmetros estejam na mesma escala de habilidades ao final do processo de estimação (ANDRADE et al., 2000).

De acordo com Klein (2003), a estimação conjunta dos parâmetros dos itens em diferentes anos avaliados permite que se faça simultaneamente a equalização entre séries/anos. Este procedimento de estimação conjunta garante que se obtenha uma estimativa única dos parâmetros do modelo, não importando a série/ano, evitando a propagação de erros quando se estima os parâmetros série a série (ou ano a ano).

Para que a escala do SAEB mantenha as características dos itens e a referência de grupo continue sendo a 8ª série de 1997, os novos itens são sempre calibrados com itens antigos já posicionados nesta escala. Desta forma, garante-se que as estimações dos parâmetros dos itens novos estão na mesma escala dos antigos

(KLEIN, 2003). Uma ideia simples do que foi discutido acima pode ser obtida visualizando a Figura 6.3.

Figura 6.3: Esquema de ligação entre itens de anos avaliados.

Levando em conta que a escala SAEB é a mesma para todas as disciplinas e anos avaliados, após a calibração, os itens são organizados por disciplina e posicionados pelo grau de dificuldade, existindo uma ligação entre os anos avaliados. Afinal, um item que é considerado difícil para os respondentes de determinado ano (do ensino fundamental ou médio), pode ser um item fácil para os respondentes do ano posterior.

Em resumo, a escala de proficiência, além de facilitar a compreensão dos resultados, permite interpretar de forma pedagógica a habilidade estimada do aluno. Isto se deve aos itens classificados como âncoras que subsidiam a caracterização feita por especialistas em cada nível da escala de proficiência (para cada disciplina), composta pelas descrições das habilidades e competências dominadas e os conteúdos compreendidos. Já os itens comuns permitem colocar em uma mesma escala e tornar comparáveis os anos avaliados da educação básica (por exemplo, o 3º, 5º, 7º e 9º anos do fundamental e 3º ano do ensino médio) e as edições anteriores da avaliação.

6.2. Divulgação dos Resultados

Após os cálculos dos resultados, a entidade responsável pela avaliação externa elabora materiais para divulgar os resultados e, na maioria das vezes, também os deixa disponíveis para consulta em seu endereço na internet.

No SARESP, por exemplo, os resultados da avaliação são divulgados pela Secretaria da Educação do Estado de São Paulo (SEE-SP). Neles, estão contidos os desempenhos alcançados por cada rede de ensino e escolas, por disciplinas e anos avaliados. Os materiais divulgados pela SEE-SP sobre o SARESP compreendem (SÃO PAULO, 2015a):

- Boletins de resultados por escola: são disponibilizados no site da SEE-SP (www.educacao.sp.gov.br), sendo o acesso aos resultados das escolas particulares e técnicas restrito aos seus gestores;
- Sumário Executivo: traz uma análise geral da avaliação e dos principais resultados de desempenho obtidos por disciplina, segundo os anos avaliados em cada rede de ensino. Também podem ser obtidos no site da SEE-SP;
- Relatórios Pedagógicos: trazem uma análise geral do desempenho obtido em cada disciplina e ano avaliado, apresentando alguns itens da prova e fazendo uma relação com a habilidade explorada da Matriz de Referência.

Considerações Finais

A Teoria da Resposta ao Item (TRI) surgiu da busca por uma metodologia alternativa à Teoria Clássica dos Testes (TCT). Sua aplicação na área educacional favorece a comparação dos resultados das avaliações e o acompanhamento da qualidade do ensino ao longo do tempo. No Brasil, a TRI passou a ser utilizada em 1995 no Sistema de Avaliação da Educação Básica (SAEB), composto pela Prova Brasil, disseminando-se para outras avaliações, como por exemplo, as que visam os ensinos estaduais, municipais e seleção para o vestibular.

Porém, a “sofisticação” dessa teoria, com a utilização de métodos diferentes dos habituais utilizados pelos professores e equipe escolar, pode impactar na compreensão de pontos importantes da metodologia, como por exemplo, a construção do gráfico de probabilidade de acerto do item. Além disso, cria resistências ao uso de resultados apurados pela TRI, utilizados no desenvolvimento de planos de ação para melhoria da qualidade do ensino.

Foi na área da Educação que a TRI ganhou ampla divulgação e grande aplicação prática. De acordo com Castro et al. (2013) esta área foi o assunto mais abordado em relação a esta metodologia nos últimos tempos, conforme um levantamento exploratório da produção científica sobre a TRI. No entanto, as várias formas exploradas para apresentar a metodologia da TRI ou são bem teóricas, explorando principalmente os modelos logísticos com suas fórmulas, ou são descritivas, apresentando vantagens do uso do procedimento, sem nenhum aparato teórico.

O desafio de apresentar a TRI a partir da construção do pensamento da metodologia e explorá-la de forma pedagógica, principalmente a construção da curva de probabilidade de acerto do item, foi o que deu origem a este trabalho, também motivado pelo desejo manifestado pelos professores que disseram ter interesse em conhecer mais sobre a técnica, em uma pesquisa, aqui descrita, realizada pelo autor. Não houve aqui o objetivo de discutir se esta é a melhor metodologia avaliativa na área educacional.

Espera-se que a compreensão da TRI contribua para que o profissional da educação reflita sobre os resultados das avaliações externas tendo mais segurança sobre a

metodologia utilizada, aprimorando os planos de ação voltados para o desenvolvimento das habilidades e competências dos alunos.

De fato, os próprios professores alvo da pesquisa realizada disseram que entender mais sobre a TRI pode trazer benefícios para a prática pedagógica e também contribuir para interpretar com maior clareza os resultados, facilitar a verificação de quais habilidades são mais difíceis de serem desenvolvidas, melhorar estratégias de ensino em sala de aula, etc.

Em um estudo sobre o fenômeno da repetência escolar no ensino fundamental pela busca de uma pedagogia de qualidade, realizado em uma escola do Rio de Janeiro, Lüdke (1994) expressa-se sobre a prática avaliativa da seguinte forma:

"certamente o que o professor faz em avaliação é condicionado pela concepção que tem desse processo mais amplo de ensino e aprendizagem". (LÜDKE, 1994, p. 30)

Se a prática pedagógica e a avaliação do processo ensino aprendizagem na sala de aula se concretizam pela bagagem teórica e concepção que o professor traz em si, então conhecer a metodologia que avalia os seus alunos e que, por consequência, monitora a qualidade de seu ensino poderá contribuir de forma positiva para um ensino com foco no desenvolvimento de habilidades e competências.

Este trabalho visou tratar a TRI sob uma abordagem pedagógica a partir do modelo de Rasch, que contém somente o parâmetro “dificuldade do item”, utilizando-se do pensamento sobre a metodologia para facilitar um primeiro contato e apresentar a construção do gráfico da curva de probabilidade de acerto do item. Vale ressaltar, mais uma vez, que este trabalho não procurou fornecer ferramentas ou ensinar o professor a usar a TRI nas suas avaliações em sala de aula. Apenas buscou-se uma forma para ajudar o professor a compreender a metodologia TRI.

Espera-se que este trabalho contribua para subsidiar a compreensão do professor da educação básica sobre essa metodologia, oferecendo este primeiro contato com o pensamento subjacente à construção da TRI. Como passo futuro, pretende-se ir além desta dissertação, disponibilizando esta discussão em outros meios e oferecendo a informação aos atuais e futuros professores.

Alinhado aos objetivos do programa de Mestrado Profissional em Ensino de Matemática do IME-USP, que coloca a importância de autonomia do indivíduo no exercício da ação docente e incentiva o professor a ser investigador de sua prática, acredita-se que este trabalho possa contribuir na perspectiva de uma educação continuada aos professores da educação básica.

Apêndices

Apêndice A

A. Pesquisa Sobre a TRI Com Professores da Educação Básica

Visando verificar as impressões e sensações obtidas nos trabalhos realizados sobre avaliações externas com professores e gestores da educação básica, principalmente com o objetivo de se verificar o conhecimento sobre fundamentos da Teoria da Resposta ao Item (TRI), como por exemplo, a construção da curva do item, o autor elaborou uma pesquisa com professores de uma grande rede de ensino privado do estado de São Paulo.

Essa pesquisa foi realizada no mês de setembro/2016 por meio de questionário eletrônico e atingiu os professores da educação básica (educação infantil, ensino fundamental, ensino médio e educação de jovens e adultos), obtendo 860 respostas, o que corresponde a um retorno de 22,4% da base total. O formulário disponibilizado continha ao todo 15 questões para serem respondidas, sendo 8 de informações complementares sobre o respondente e 7 diretamente ligadas ao tema da pesquisa. Em todas as questões ligadas ao tema da pesquisa foram colocadas opções de respostas fechadas e um campo aberto, exigindo obrigatoriamente a seleção de uma opção e a resposta escrita, justificando a opção do professor. Isto permitiu ao professor apresentar sua opinião de maneira livre em cada questão.

As figuras de A.1 a A.6 apresentam os gráficos que descrevem o perfil dos professores que responderam à pesquisa. Em síntese, 81% dos professores é do sexo feminino e quase 50% destas professoras possuem graduação no curso de Pedagogia e somente 3% são da área de matemática. Quanto aos professores do sexo masculino, 48,8% se classificaram na categoria outra licenciatura (que exclui letras e matemática) e 24% possuem licenciatura em matemática. Quase 75% do total de professores possuem ou estão cursando uma pós-graduação, sendo a maior parte do tipo especialização (mínimo de 360h). Sobre a modalidade em que ministram aula, quase a totalidade, estão nos ensinos fundamental e médio, e possuem 16 anos em média na profissão de docente. Além disso, 27,5% desses professores também ministram aula na rede pública de ensino.

Figura A.1: Distribuição dos professores por sexo.

Figura A.2: Distribuição dos professores por sexo e formação.

Figura A.3: Formação em relação a pós-graduação.

Figura A.4: Modalidade da educação básica em que ministra aula na rede pesquisada.

Figura A.5: Anos completos na profissão de professor.

Figura A.6: Redes de ensino em que ministra aula, além da rede pesquisada.

As figuras de A.7 a A.12 tratam dos itens ligados diretamente à pesquisa. Estes itens serão apresentados a seguir, um a um, com exceção dos dois últimos, que foram analisados em conjunto devido aos professores entenderem que tratavam do mesmo assunto. Em cada gráfico está descrita a pergunta proposta.

A Figura A.7 apresenta o gráfico da primeira questão, trazendo a informação que para 97% dos professores, a avaliação externa possui uma importância relevante em seu ensino na sala de aula. Dentre as justificativas fornecidas pelos professores, as mais recorrentes foram:

- Avaliar/aprimorar a prática pedagógica (179);
- Direcionar o trabalho / Revisar conteúdo (55);
- Mensurar o nível de proficiência do aluno (48);
- Identificar/trabalhar as dificuldades/habilidades/conteúdos dos alunos (39);
- Permite identificar/priorizar/trabalhar habilidades e competências (38);
- Melhoria do processo ensino aprendizagem (31).

Figura A.7: Importância dos resultados da avaliação externa na sala de aula.

De acordo com as justificativas, percebe-se que pelo menos 20% (179/860) dos professores acreditam que a avaliação externa possa contribuir para a melhoria da prática pedagógica. As justificativas apresentadas sugerem que o professor dá

importância aos resultados das avaliações externas e entendem que estas estão voltadas a melhoria do ensino. Porém, há professores que entendem que as avaliações externas avaliam o aluno em si, não refletem a realidade da escola, é classificatória, não são justas com o professor, trazem resultados de turmas passadas e não das atuais, etc.

A Figura A.8 apresenta o gráfico da questão que abordou o conhecimento sobre a TRI, trazendo a informação que poucos professores (7%) declararam saber muita coisa sobre o assunto, mas o restante (93%) declarou não ter conhecimentos profundos, muito pouco ou quase nada a respeito da metodologia. Dentre os que declararam saber bastante, um professor disse trabalhar em uma Instituição responsável pela elaboração de avaliações externas de caráter estadual em algumas unidades federativas do país. Outros professores disseram ter participado de cursos para elaboração de itens que compõem avaliações com base na TRI ou mesmo que este assunto é tratado nas formações internas.

Figura A.8: Conhecimento sobre a metodologia TRI.

Grande parte dos professores acredita que, atualmente, a TRI é a melhor forma de mensurar o desempenho dos alunos (Figura A.9), pois, com esta metodologia, deixa-se de considerar apenas a quantidade de acerto e olha-se para a coerência das respostas, penalizando possíveis “chutes”.

Figura A.9: Opinião sobre a melhor forma de mensurar o desempenho do aluno em avaliações do sistema de ensino.

3. Antes as avaliações do sistema de ensino mensuravam o desempenho do aluno a partir do número de questões acertadas. Atualmente usam a TRI. Na sua opinião, qual forma é melhor e por quê? Se na sua opinião existir outra melhor, cite-a.

Em síntese, as justificativas apontam que os professores acham que a TRI é melhor por avaliar as habilidades dos alunos, permitindo, por meio do resultado e de posse da escala, verificar o que o aluno domina e quais habilidades precisarão ser melhor trabalhadas.

Os professores que optaram pela metodologia anterior à TRI, a TCT (Teoria Clássica dos Testes), justificam que a TCT é mais prática, objetiva, simples e é clara em relação à forma de mensurar o desempenho do aluno, pois se sabe claramente a quantidade de acertos e erros. Eles afirmam ainda, que a TRI é uma metodologia complexa e de difícil compreensão. Mas há professores (34% ou 20/58) que dizem optar pela TCT por desconhecer a TRI.

Dentre os professores que escolheram outra forma, 39% (32/82) disseram ter escolhido esta opção por não conhecer a TRI. O restante (50/82) sugere que a melhor forma é a avaliação contínua e formativa, pois acreditam que o raciocínio do aluno deve ser levado em conta e não somente o seu acerto ou erro.

Conforme observado na Figura A.10, a maioria dos professores declarou não possuir conhecimento sobre a construção do gráfico de probabilidade de acerto do item.

Figura A.10: Conhecimento sobre a construção do gráfico da probabilidade de acerto do item.

4. Você tem algum conhecimento sobre como é construído o gráfico da probabilidade de acerto do item em função da habilidade do aluno?

Dentre os professores que disseram “Sim”, poucos citaram métodos estatísticos que levam a construção da Curva Característica dos Itens (CCI).

A percepção, obtida a partir da leitura das justificativas desses professores, é que eles sabem o que significa a CCI, mas, por não terem dado evidências mínimas, não sabem como a curva foi construída. Por exemplo:

- 12% (34/276) disseram que a curva relaciona a probabilidade de acerto do item com a habilidade do aluno, de tal forma que quanto maior a habilidade do aluno maior será a probabilidade de ele acertar o item;
- 11% (29/276) afirmam ter estudado, mas não se aprofundaram a tal ponto;
- Outros apenas citam que estudam o tema em reuniões pedagógicas, leem os relatórios do SARESP, pesquisam pela Internet, viram o tema em cursos de formação de professores, citam que a rede estadual aborda muito esta questão, etc.

Nota-se também que não há um consenso dentre as justificativas apresentadas pelos professores que afirmaram que sabem. A fim de exemplificar, abaixo constam algumas justificativas extraídas da pesquisa sobre como o gráfico é construído:

- a) “Por uma representação simbólica de dados”;

- b) “É construída uma linha mediana que permeia o resultado das habilidades dos estudantes”;
- c) “É feita uma análise das possíveis respostas e suposições de respostas dadas pelos alunos”;
- d) “Pela elaboração de escalas de proficiência, competências e habilidades”;
- e) “Ele é baseado nas formas de raciocínio que o aluno pode ter para responder”;
- f) “Feita uma escala mediante os descritores, cujas propriedades estatísticas permitem situar o aluno a seu nível”;
- g) “Imagino que há um levantamento das hipóteses que o aluno teria formulado para chegar a alguns resultados”;
- h) “Imagino que seja a partir dos resultados anteriores”;
- i) “Número de acertos”;
- j) “O cálculo é baseado no nº de acertos e erros, tira uma média ponderada”;
- k) “O rendimento é calculado de acordo com o valor de cada questão”;
- l) “O rendimento total do gráfico apresenta a soma do valor de cada questão”;
- m) “O rendimento total do gráfico tem por base o valor de cada questão”;
- n) “Pela complexidade da questão proposta e pela mediana de acertos de todo o grupo realizador da prova”.

Desta forma, é possível que o percentual de professores que não sabem como a curva característica do item (CCI) foi construída seja bem superior aos 68% declarados na pesquisa (Figura A.10). Ao perguntar sobre o interesse de saber mais sobre a construção do gráfico de probabilidade de acerto do item, a grande maioria respondeu que acha interessante Figura A.11.

Figura A.11: Interesse em saber sobre a construção do gráfico da probabilidade de acerto do item.

5. Você acha interessante saber como se constrói o gráfico da probabilidade de acerto do item em função da habilidade do aluno?

Dentre os professores que responderam “Sim” na Figura A.11, ou seja, que declaram ter algum conhecimento sobre a construção da curva de probabilidade do item, observou-se que 30% (84/276) disseram ser interessante saber como se constrói esta curva. Nas justificativas destes professores consta que eles querem saber mais, aprender mais, sobre o assunto. Isto pode ser mais uma evidência de que o percentual de professores que não sabem como a CCI é construída seja superior aos 68%.

A Figura A.12 apresenta duas questões de forma conjunta, pois, os professores entenderam que estas questões tratavam do mesmo assunto. Os gráficos mostram que os professores acham que saber sobre a construção do gráfico poderá sim trazer algum tipo de benefício para o ensino em sala de aula e contribuir para melhoria de sua prática pedagógica.

Figura A.12: Contribuição da TRI para a prática pedagógica.

6. Em sua opinião, existe algum benefício em aprender sobre a construção do gráfico da probabilidade de acerto do item em função da habilidade do aluno? Em caso negativo, explique o porquê e em caso afirmativo, cite qual(is).

7. Você acredita que “saber mais” sobre a TRI poderia contribuir para melhorar sua prática pedagógica?

Dentre os benefícios citados pelos professores, destacam-se os seguintes:

- Apropriar-se de um conhecimento matemático que não possuem;
- Realizar uma análise melhor e interpretar com maior clareza os resultados;
- Extrair informações sobre as habilidades que os alunos não dominam;
- Permitir melhorar estratégias para explorar as habilidades avaliadas;
- Contribuir com a melhoria da prática docente no cotidiano;
- Obter interpretações mais rápidas e eficazes, sem precisar folhear vários documentos;
- Facilitar a verificação de quais habilidades são mais difíceis de serem desenvolvidas;
- Acompanhar o processo de desenvolvimento de habilidades e competências dos alunos na educação básica;
- Analisar melhor os resultados das avaliações externas dos alunos, direcionando o trabalho de uma maneira mais efetiva;
- Ajudar o aluno a compreender o seu resultado obtido na avaliação;
- Traçar/planejar o que será dado aos alunos durante o ano;
- Fazer uma análise mais qualitativa, saindo do quantitativo;
- Ser mais assertivo na elaboração do diagnóstico do ensino da escola e no plano de ação;

- Entender como é avaliada a habilidade do aluno em cada questão;
- Fazer um *feedback* individualizado, quando de posse dos resultados dos alunos.

Sobre a prática pedagógica, destaca-se:

- Planejar as atividades com base nas habilidades que devem ser desenvolvidas nos alunos;
- Nortear os trabalhos desenvolvidos em sala de aula;
- Maior segurança na elaboração dos planos de ação do ensino;
- Melhoria na qualidade das avaliações do conhecimento utilizadas no cotidiano;
- Aperfeiçoamento da prática em sala de aula;
- Auxílio no processo ensino-aprendizagem.

Apêndice B

B. Exemplo Para Compreensão da Metodologia BIB

Considerando uma avaliação que tem por objetivo avaliar o sistema de ensino, após a seleção dos itens que contemplam todas as habilidades das Matrizes de Referência, as provas poderão ser montadas.

A título de exemplo, será apresentada a seguir uma construção hipotética da prova da disciplina de Matemática para o 7º ano do ensino fundamental, com base na metodologia BIB. O objetivo do exemplo é encontrar a quantidade de cadernos de provas a partir do número ideal de blocos obtidos e, consequentemente de itens, que possam ser submetidos aos alunos em determinado tempo de prova.

Para isso, suponha que cada item possa ser representado por uma bolinha e que um conjunto de itens, referentes a todas as habilidades da Matriz de Referência da disciplina de matemática dos ensinos fundamental e médio, esteja **organizado de forma crescente pela respectiva dificuldade**, cuja métrica varia de 0 a 500, a mesma utilizada pelo SARESP. A Figura B.1 exibe uma representação desse banco de itens.

Figura B.1: Banco de itens que contemplam todas as habilidades das matrizes de referência de matemática da educação básica.

Uma vez que os itens foram ordenados de forma crescente pela dificuldade de acerto, é realizada uma seleção dos itens que irão compor a prova de avaliação. **A seleção não é feita de forma aleatória, pois esta deve cobrir toda a Matriz Referência e pertencer ao intervalo de proficiência que corresponde ao(s) ano(s) avaliado(s)**, que nesse caso, é o 7º ano do ensino fundamental.

Para facilitar a compreensão desse exemplo de seleção e montagem dos cadernos de provas, sob a suposição hipotética de que os itens selecionados cubram toda a Matriz de Referência da disciplina de matemática do 7º ano do ensino fundamental, foi selecionado um número baixo de itens (12) compostos por: 4 itens considerados fáceis (verdes), 4 médios (laranjas) e 4 difíceis (vermelhos), conforme apresentado na Figura B.2.

Figura B.2: Seleção de itens que irão compor a prova de matemática do 7º ano do ensino fundamental.

Considere que o tempo de prova destinado à disciplina de matemática seja de 1 hora e que cada questão leve em média 7 minutos para ser resolvida. Então, para resolver os 12 itens seria preciso 1 hora e 30 minutos. Logo, 12 itens é uma quantidade inadequada para o tempo da avaliação e, portanto, deve-se limitar a quantidade de itens submetida a cada aluno a um valor aceitável e adequado ao tempo da prova.

Poder-se-ia, simplesmente, selecionar uma quantidade menor aleatória de itens para cada aluno e aplicar. Mas, as proporções de respostas obtidas por item poderiam ser bem diferentes e causar um desbalanceamento. Uma saída é agrupar de forma aleatória, os 12 itens selecionados no banco em diversos blocos balanceados (com a mesma quantidade de itens), como por exemplo:

- Sem Repetição de itens no bloco: 2 blocos com 6 itens cada um, 3 blocos com 4 itens cada um, 4 blocos com 3 itens cada um;
- Com Repetição de itens no bloco: 4 blocos com 5 itens cada um, 5 blocos com 4 itens cada um, 6 blocos com 3 itens cada um, etc.

Para aplicar o BIB, geralmente, divide-se a quantidade total de itens em blocos que possuam a mesma quantidade de itens, sem repetições dentro dos blocos, em que cada bloco mantém a proporção de dificuldade do conjunto inicial (% de itens fáceis, % de médios e % de difíceis). Por exemplo, a divisão de 12 itens em 4 blocos resulta em 3 itens para cada bloco, que de acordo com a composição inicial (1/3 fácil (verdes), 1/3 médio (laranjas) e 1/3 difícil (vermelhos)), cada bloco deverá ter um item considerado fácil, outro médio e outro difícil.

Após a definição do número de blocos, que foi considerado no exemplo acima como sendo 4 blocos com 3 itens cada um, a metodologia BIB fornecerá o número ideal de cadernos de prova que devem ser confeccionados para a aplicação. Mas, para o exemplo em questão isto será feito de forma manual, por meio de combinações, levando em consideração que cada aluno deve responder o mesmo bloco uma única vez. Os resultados obtidos das combinações estão apresentados no Quadro B.1.

Quadro B.1: Quantidade de cadernos de provas, devido ao número de blocos a serem respondidos pelo aluno.

Caso	Número de Blocos a serem respondidos pelo aluno	Combinações que podem ser feitas com os Blocos, sem repetição e importância da ordem		Total de Cadernos
1º	4 Blocos	1	Qual: 1. <u>I</u> <u>II</u> <u>III</u> <u>IV</u>	1
2º	3 Blocos	4	Quais: 1. <u>I</u> <u>II</u> <u>III</u> 2. <u>I</u> <u>II</u> <u>IV</u> 3. <u>I</u> <u>III</u> <u>IV</u> 4. <u>II</u> <u>III</u> <u>IV</u>	4
3º	2 Blocos	6	Quais: 1. <u>I</u> <u>II</u> 2. <u>I</u> <u>III</u> 3. <u>I</u> <u>IV</u> 4. <u>II</u> <u>III</u> 5. <u>II</u> <u>IV</u> 6. <u>III</u> <u>IV</u>	6

4º	1 Bloco	4	Quais: 1. <u>I</u> 2. <u>II</u> 3. <u>III</u> 4. <u>IV</u>	4
----	---------	---	--	---

Como cada bloco possui 3 itens, o primeiro caso do Quadro B.1 é inviável, pois a quantidade de 12 itens para um aluno responder é inadequada ao tempo que se dispõe de prova. Por outro lado, o último caso permite cada aluno responder uma fração “muito pequena” do total de itens e pode sobrar tempo de prova. Então, restam o segundo e o terceiro casos em que responder 6 itens no tempo de prova seria ótimo e responder 9 itens neste mesmo tempo seria razoável.

A decisão de optar pelo segundo ou terceiro caso terá como base o fator de utilização. Ou seja, a proporção em que cada bloco é submetido ao número total de alunos, que é equivalente à proporção de blocos (por consequência, itens) respondidos por cada aluno. No segundo caso, o fator utilização é 3/4 ou 75%, e no terceiro é 2/4 ou 50%. Como no segundo caso os itens estarão sendo respondidos por uma quantidade maior de alunos, levando a melhores conclusões sobre o sistema de ensino, então o segundo caso é escolhido.

Dessa forma, obtém-se 4 cadernos, cada um composto por 3 dos 4 blocos em que foram agrupados os 12 itens. E isto é considerado suficiente pela metodologia BIB para garantir que os itens sejam respondidos por uma “boa quantidade de alunos”³¹ e obter resultados fidedignos, permitindo que se tire conclusões a respeito da disciplina de matemática do 7º Ano do ensino fundamental. A Figura B.3 apresenta um esquema que faz uma analogia sobre o agrupamento realizado dos itens selecionados a partir do exemplo, onde cada bloco é representado por uma cor (roxo, amarelo, verde, rosa) e possui um item fácil (bolinha verde), um médio (bolinha laranja) e outro difícil (bolinha vermelha).

³¹ Este termo se deve ao fato de que todos os itens devem ser respondidos por uma quantidade grande de alunos. Porque se um item for respondido por 1000 e outro por 10, por exemplo, há um desbalanceamento. E isto, pode levar a conclusões erradas sobre o sistema. O ideal é que os itens sejam respondidos por uma “mesma” quantidade (grande) de alunos.

Figura B.3: Agrupamento dos itens para composição das provas da avaliação.

A fórmula utilizada pela metodologia BIB leva em conta de uma única vez todas as considerações realizadas acima, fornecendo de uma forma mais rápida a quantidade ideal de cadernos de provas. Em Bekman (2001) são apresentados exemplos de como calcular a quantidade de cadernos por meio do BIB.

A Figura B.4 traz a ilustração dos cadernos de provas compostos por 3 blocos cada um, de acordo com a metodologia. O caderno A, por exemplo, é composto pelos blocos roxo, amarelo e rosa. Cada cor de bloco contém três itens, porém um é do tipo fácil (bolinha verde), outro médio (bolinha laranja) e outro difícil (bolinha vermelha). Assim o caderno A mantém a mesma proporção de dificuldade apresentada pelo conjunto inicial de 12 itens. Isso também vale para os cadernos B, C e D.

Figura B.4: Composição dos cadernos de provas da avaliação.

Em resumo, a aplicação da metodologia implicou na composição de 4 cadernos de provas, com cada um contendo 3 blocos dos 4 inicialmente criados a partir da distribuição dos 12 itens. Como cada bloco é composto por 3 itens, este método

permitiu que, ao invés de compor um único caderno de prova com 12 itens, fossem obtidos 4 cadernos de provas com 9 itens. Ou seja, para este caso, os alunos responderiam 9 dos 12 itens selecionados em 1 hora de prova, o que corresponde a uma redução de 25% nos itens que deveriam ser respondidos por cada aluno.

Anexo A

Questionário da Pesquisa Apresentada no Apêndice A

Informações Complementares

1. Qual a escola da rede em que Presta Serviços?
2. Qual o seu Sexo?
 - a) Masculino
 - b) Feminino
3. Qual o seu Nível de Escolaridade que lhe permite lecionar? Assinale o nível mais alto.
 - a) Ensino Médio (Magistério)
 - b) Pedagogia
 - c) Normal Superior
 - d) Licenciatura em Matemática.
 - e) Licenciatura em Letras.
 - f) Outra Licenciatura. Qual curso? _____
 - g) Outros. Qual? _____
4. Possui pós-graduação?
 - a) Sim, concluída.
 - b) Sim, cursando.
 - c) Não.
 - 4.1 Qual o tipo da pós-graduação?
 - a) Aperfeiçoamento (mínimo de 180 horas).
 - b) Especialização (mínimo de 360 horas).
 - c) Mestrado.
 - d) Doutorado.
5. Ministra aulas em outra rede de ensino?

- a) Municipal.
- b) Estadual.
- c) Federal.
- d) Privada.
- e) Não.

6. Em qual(is) ano(s) da educação básica na rede você ministra aulas?

Ensino Fundamental:

- a) 1º
- b) 2º
- c) 3º
- d) 4º
- e) 5º
- f) 6º
- g) 7º
- h) 8º
- i) 9º

Ensino Médio:

- a) 1º
- b) 2º
- c) 3º

7. Em qual(is) disciplina(s) ministra aulas na educação básica (seja nesta rede ou em outra)?

- a) Polivalente
- b) Matemática
- c) Língua Portuguesa
- d) Artes / Educação Artística
- e) Biologia (Ensino Médio)
- f) Ciências (Ensino Fundamental)
- g) Educação Física
- h) Filosofia / Sociologia
- i) Física (Ensino Médio)

- j) Geografia
- k) História
- l) Informática
- m) Língua Estrangeira (Inglês, Espanhol, etc...)
- n) Química (Ensino Médio)
- o) Vivência (Ciência e Tecnologia, Lúdica, Artes, etc.)
- p) Outra(s). _____

8. Qual o tempo em anos completos na profissão de professor?

- a) Menos de 3 anos.
- b) De 3 a 5 anos.
- c) De 6 a 10 anos.
- d) De 11 a 15 anos.
- e) De 16 a 20 anos.
- f) De 21 a 25 anos.
- g) De 26 a 30 anos.
- h) Acima de 30 anos.

Itens sobre a Pesquisa:

9. Qual a importância do uso dos resultados de uma avaliação externa para melhoria do seu ensino em sala de aula?

- a) Alta.
- b) Média.
- c) Baixa.
- d) Não sei.

Justifique.

--

10. O que você conhece sobre a metodologia TRI (Teoria da Resposta ao Item) utilizada para mensurar a proficiência do aluno nas avaliações externas?

- a) Muita coisa, já estudei bastante sobre o assunto.
- b) Eu tenho conhecimento sobre a metodologia, mas nada tão profundo.
- c) Eu sei muito pouco sobre o assunto.
- d) Não sei nada a respeito.

Justifique.

11. Antes as avaliações do sistema de ensino mensuravam o desempenho do aluno a partir do número de questões acertadas. Atualmente usam a TRI. Na sua opinião, qual forma é melhor e por quê? Se na sua opinião existir outra melhor, cite-a.

- a) A anterior a TRI.
- b) A TRI.
- c) Outra. (Cite-a na no campo de justificativa)

Justifique.

12. Você tem algum conhecimento sobre como é construído o gráfico da probabilidade de acerto do item em função da habilidade do aluno?

- a) Sim.
- b) Não.

Justifique.

13. Você acha interessante saber como se constrói o gráfico da probabilidade de acerto do item em função da habilidade do aluno?

- a) Sim.
- b) Não.
- c) Não sei

Justifique.

14. Em sua opinião, existe algum benefício em aprender sobre a construção do gráfico da probabilidade de acerto do item em função da habilidade do aluno? Em caso negativo, explique o porquê e em caso afirmativo, cite qual(is).

- ☐ Sim.
- ☐ Não.
- ☐ Não sei

Justifique.

--

15. Você acredita que “saber mais” sobre a TRI poderia contribuir para melhorar sua prática pedagógica?

- ☐ Sim.
- ☐ Não.
- ☐ Não sei

Justifique.

--

Referências Bibliográficas

AGRESTI, Alan. **Categorical Data Analysis**. New Jersey: John Wiley & Sons, Inc., Hoboken, 2nd ed., 2002. 710p.

ALAVARSE, Ocimar M.; BRAVO, Maria Helena; MACHADO, Cristiane. **Avaliações Externas e Qualidade na Educação Básica: Articulações e Tendências**. Estudos em Avaliação Educacional, v. 24, n. 54, p. 12-31. São Paulo, FCC: 2013. Disponível em: <<http://www.fcc.org.br/pesquisa/publicacoes/eae/arquivos/1783/1783.pdf>>. Acesso em: 11/03/2017.

ANDRADE, Dalton Francisco de; VALLE, Raquel da Cunha. **Introdução à Teoria da Resposta ao Item: conceitos e aplicações**. Estudos em Avaliação Educacional, nº 18. São Paulo, FCC: 1998. Disponível em: <<http://www.fcc.org.br/pesquisa/publicacoes/eae/arquivos/1116/1116.pdf>>. Acesso em: 03/09/2016.

ANDRADE, Dalton Francisco de; TAVARES, Heliton Ribeiro; VALLE, Raquel da Cunha. **Teoria da Resposta ao Item: Conceitos e Aplicações**. São Paulo: ABE, 14^o SINAPE, 2000.

ANDRADE, Josemberg Moura de; LAROS, Jacob Arie; GOUVEIA, Valdiney Veloso. **O uso da teoria de resposta ao item em avaliações educacionais: diretrizes para pesquisadores**. Avaliação Psicológica: v.9, n.3, p. 421-435. Porto Alegre: Dez/2010. Disponível em: <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712010000300009>. Acesso em: 07/11/2016.

ARAÚJO, Eulália Aparecida Candido de; ANDRADE, Dalton Francisco de; BORTOLOTTI, Silvana Ligia Vincenzi. **Teoria da Resposta ao Item**. Revista da Escola de enfermagem da USP, versão on-line ISSN 0080-6234, v.43, p. 1000-1008. São Paulo: Dez/2009. Disponível em: <<http://www.scielo.br/pdf/reeusp/v43nspe/a03v43ns.pdf>>. Acesso em: 07/11/2016.

ANDRIOLA, Wagner B. **Utilização da Teoria da Resposta ao Item (TRI) para a organização de um banco de itens destinados a avaliação do raciocínio verbal**.

Psicologia: Reflexão e Crítica, on-line version ISSN 1678-7153, vol.11, n.2. Porto Alegre, 1998. Disponível em: < http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-79721998000200009&lng=pt&nrm=iso&tlng=pt>.

Acesso em: 07/05/2016.

ANDRIOLA, Wagner B. **Psicometria Moderna: Características e Tendências**. Estudos em Avaliação Educacional, v. 20, n. 43, maio/ago. São Paulo, FCC: 2009. Disponível em: <<http://www.fcc.org.br/pesquisa/publicacoes/eae/arquivos/1496/1496.pdf>>. Acesso em: 14/11/2015.

BAKER, Frank B. **The Basics of Item Response**. 2ª ed. Madison: ERIC Clearinghouse on Assessment and Evaluation, 2011.

BEKMAN, Roberto. **Aplicação dos Blocos Incompletos Balanceados na Teoria de Resposta ao Item**. Estudos em Avaliação Educacional, nº 24, jul/dez. São Paulo, FCC: 2001. Disponível em: < <http://www.fcc.org.br/pesquisa/publicacoes/eae/arquivos/990/990.pdf>>. Acesso em: 07/11/2015.

BOND, Trevor G.; FOX, Christine M. **Applying the Rasch Model: Fundamental Measurement in the Human Sciences**. 2ª ed. Routledge, Nova York, EUA: Routledge, 2010.

BORTOLOTTI, S. L. V; et al. **Relevance and advantages of using the item response theory**. Quality & Quantity: International Journal of Methodology, March, 2012.

BRASÍLIA (Distrito Federal), INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Relatório Nacional Saeb 2003**. Brasília: O Instituto, 2006. 266 p.

BRASÍLIA (Distrito Federal), INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Guia do Participante: Entenda a sua Nota do ENEM**. Brasília: 2012. Disponível em: <http://download.inep.gov.br/educacao_basica/enem/guia_participante/2013/guia_do_participante_notas.pdf>. Acesso em: 02/01/2017.

BRASÍLIA (Distrito Federal), INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Caderno Pedagógico: Prova Brasil – Avaliação do Rendimento Escolar**. Brasília: INEP, versão atualizada, 2016. Disponível em: <http://download.inep.gov.br/educacao_basica/prova_brasil_saeb/resultados/2013/caderno2013_v2016.pdf>. Acesso em: 15/04/2017.

CASTRO, Rafael Bernardo de; et al. **Teoria da Resposta ao Item: levantamento exploratório da produção científica**. In: III CONBRATRI - Congresso Brasileiro de Teoria da Resposta ao ITEM. Anais n. 1. Belém: 2013. p. 129-142. Disponível em: <<http://abave.com.br/ojs/index.php/Conbratri/article/view/61/70>>. Acesso em: 26/03/2017.

CURADO, Maria A. Santos; TELES, Júlia; MAROCO, João. **Análise de variáveis não diretamente observáveis: influência na tomada de decisão durante o processo de investigação**. Revista da Escola de Enfermagem, São Paulo, USP, vol.48, n.1, Fev. 2014. Disponível em: <http://www.scielo.br/scielo.php?pid=S0080-62342014000100146&script=sci_arttext&tlng=pt#B06>. Acesso em: 07/05/2016.

D'AMORE, Bruno. **Elementos de Didática da Matemática**. 1ª ed. São Paulo: Livraria da Física, 2007.

GOMES, Cristiano M. A. **Feuerstein e a construção mediada do conhecimento**. Porto Alegre. Artmed, 2002.

GREGOIRE, Janes; LAVEAULT, Dany. **Introdução às Teorias dos Testes em Ciências Humanas**. Tradução: Botelho, Annete P. R. et al. 1ª ed. Porto, Portugal: Porto Editora, 2002.

GUSKEY, Thomas R. **The Case Against Percentage Grades**. Educational Leadership: Resiliense and Learning, v.71, n. 1, p. 68-72, Sep/2013. Disponível em: <<https://next-generation-leadership.wikispaces.com/file/view/GUSKEY+EL13+Percentage+Grades.pdf>> Acesso em: 16/04/2016.

HAMBLETON, Ronald K.; JONES, Russel W. **Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development**. Educational Measurement: Issues and Practice, ITEMS Modules, p. 38-47,

Philadelphia, 1993. Disponível em: <

HAMBLETON, Ronald K.; SWAMINATHAN, H.; ROGERS, H. Jane. **Fundamentals of Item Response Theory**. 1ª ed. London: Sage Publications, Inc: 1991. 174 p.

HOSMER, David W; LEMESHOW, Stanley. **Applied Logistic Regression**. New York: John Wiley & Sons, Inc., 2nd ed., 2000. 375p.

JUIZ DE FORA (Cidade), Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de. **Guia de Elaboração de Itens: Matemática**. Juiz de Fora: CAed/UFJF, 2008. Disponível em: <http://www.portalavaliacao.caeduffj.net/wp-content/uploads/2012/02/Guia_De_-Elabora%C3%A7%C3%A3o_De_Itens_MT.pdf>. Acesso em: 07/11/2015.

KLEIN, Ruben; FONTANIVE, Nilma. **Avaliação em Larga Escala: uma proposta inovadora**. Brasília, ano 15, n.66, abr./jun. 1995.

KLEIN, Ruben. **Utilização da Teoria de Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica (SAEB)**. Rio de Janeiro: Revista Ensaio, v. 11, n. 40, p. 283-296, jul./set. 2003. Disponível em: < <http://revistas.cesgranrio.org.br/index.php/metaavaliacao/article/view/38/17>>. Acesso em: 04/10/2016.

LENHARD, Thiago Henrique. **Métodos de Verificação das Suposições e da Qualidade de Ajuste dos Modelos TRI Cumulativos Unidimensionais**. Porto Alegre: UFRGS, 2013. Disponível em: < <https://www.lume.ufrgs.br/bitstream/handle/10183/94507/000914961.pdf?sequence=1>>. Acesso em: 11/01/2017.

LIMA, Elon Lages. **Logaritmos**. Rio de Janeiro: Sociedade Brasileira de Matemática, 2009. 113p.

LÜDKE, Menga; MEDIANO, Zélia D. (Coord.). **Avaliação na escola de 1º grau**. 2. ed. Campinas: Papirus, 1994.

MAYDEU-OLIVARES, Alberto. **Goodness-of-Fit Assessment of Item Response Theory Models**. Barcelona: Routledge, Measurement, V.11, p. 71-101, 2013. Disponível em: <http://www.ub.edu/gdne/documents/mirp13_focus_article.pdf>. Acesso em: 10/01/2017.

MAGALHÃES, Marcos Nascimento; LIMA, Antonio Carlos Pedroso de. **Noções de Probabilidade de Estatística**. 7ª Ed. São Paulo: Editora da Universidade de São Paulo, 2013.

MARQUES, Eunaihara Ligia Lira. **Principais Considerações Sobre Medida em Psicologia**. PORTAL EDUCAÇÃO, 2012. Disponível em: <<http://www.portaleducacao.com.br/educacao/artigos/19811/principais-consideracoes-sobre-medida-em-psicologia#ixzz3lqPp3Ead>>. Acesso em: 15/09/2015.

MARQUEZ, R. M. G.; BARROS; J. D. B. de. **Crescimento Populacional de Micro-organismos: Motivação para o Estudo de Funções Exponenciais e Logarítmicas**. Rio de Janeiro: UERJ, 2015. Disponível em: <http://www.editorarealize.com.br/revistas/ceduce/trabalhos/TRABALHO_EV047_MD1_SA7_ID1312_05052015223023.pdf> Acesso em: 17/11/2015.

MATOS, Marcos Pinheiro. **Funções Exponenciais e Logarítmicas**. Campo Grande: UFMS, 2014. Disponível em: <<http://repositorio.cbc.ufms.br:8080/jspui/bitstream/123456789/2153/1/MARCOS%20PINHEIRO%20MATOS.pdf>>. Acesso em: 15/04/2017.

MERCHAN-HAMANN, Edgar; TAUIL, Pedro Luiz; COSTA, Marisa Pacini. **Terminologia das medidas e indicadores em epidemiologia: subsídios para uma possível padronização da nomenclatura**. Inf. Epidemiol. Sus, Brasília, v. 9, n. 4, p. 276-284, dez. 2000. Disponível em: <http://scielo.iec.pa.gov.br/scielo.php?script=sci_arttext&pid=S0104-16732000000400006&lng=pt&nrm=iso>. Acesso em: 14/02/2017.

NUNES, Carlos; PRIMILL, Ricardo. **Impacto do tamanho da amostra na calibração de itens e estimativa de escores por teoria de resposta ao item**. Avaliação Psicológica: versão On-line ISSN 2175-3431, v.4, n.2. Porto Alegre, 2005. Disponível em: <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712005000200006> Acesso em: 05/06/2016.

PASQUALI, Luiz. **Psicometria: Teoria dos Testes na Psicologia e na Educação**. 5ª ed. Petrópolis, RJ: Vozes, 2013.

PASQUALI, Luiz; PRIMI, Ricardo. **Fundamentos da Teoria da Resposta ao Item – TRI**. Avaliação Psicológica, v.2, n.2. Porto Alegre: 2003. Disponível em: <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712003000200002>. Acesso em: 17/09/2015.

PASQUALI, Luiz. **Psicometria**. Revista da Escola de Enfermagem da USP, v. 43, n. SPE, p. 992-999. São Paulo: 2009. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0080-62342009000500002>. Acesso em: 18/09/2016.

PEREIRA, Vinicius R. **Métodos Alternativos no Critério Brasil para Construção de Indicadores Sócio-Econômicos: Teoria da Resposta ao Item**. Rio de Janeiro: PUC-RIO, 2004. Disponível em: <http://www.maxwell.vrac.puc-rio.br/Busca_etds.php?strSecao=resultado&nrSeq=5253@1>. Acesso em: 14/11/2015.

PERRY, Flavia Almeida. **Escalas de proficiência: diferentes abordagens de interpretação na avaliação educacional em larga escala**. Juiz de Fora: UFJF, 2009. Disponível em: <<http://www.ufjf.br/ppge/files/2010/07/Disserta%C3%A7%C3%A3o-flavia-perry.pdf>>. Acesso em: 03/10/2016.

PIMENTA, Cláudia O. **Avaliações Externas e o Trabalho de Coordenadores Pedagógicos: Estudo em uma Rede Municipal Paulista**. São Paulo, FE-USP, 2012. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/48/48134/tde-30052012-101850/pt-br.php>> Acesso em: 04/06/2016.

RABELO, Mauro Luiz. **Análise Comparativa dos Processos de Avaliação**. Brasília: II Colóquio de Matemática da Região Centro-Oeste, UNB, 2011. Disponível em: <<http://www.emis.de/journals/em/docs/coloquios/CO-2.02.pdf>>. Acesso em: 03/09/2016.

RIO DE JANEIRO (Estado), Secretaria de Estado de Educação. **Avaliação Interna da Aprendizagem**. Rio de Janeiro: SEE, 2013. Disponível em: <[http://download.rj.gov.br/documentos/10112/157759/DLFE-63748.pdf/MANUAL ORIENTACOESOPERACIONALIZACAOPORTARIASEEDUCSUGENN419docversaorevisadaefinalizada_certo.pdf](http://download.rj.gov.br/documentos/10112/157759/DLFE-63748.pdf/MANUAL_ORIENTACOESOPERACIONALIZACAOPORTARIASEEDUCSUGENN419docversaorevisadaefinalizada_certo.pdf)>. Acesso em: 13/12/2015.

SADOVSKY, Patrícia. **O Ensino de Matemática Hoje**. 1ª ed. São Paulo: Editora Ática, 2007. 111 p.

SÃO PAULO (Estado), Secretária da Educação. **Matrizes de Referência para Avaliação SARESP: Documento Básico/Secretária da Educação**. São Paulo, SEE: 2009. 174p. V1. Disponível em: <http://file.fde.sp.gov.br/saresp/saresp2015/Arquivos/MATRIZ_REFERENCIA_SARESP.pdf>. Acesso em: 07/11/2015.

SÃO PAULO (Estado), Secretaria da Educação. **Currículo do Estado de São Paulo: Matemática e suas tecnologias / Secretaria da Educação**. São Paulo: SEE, 2011. Disponível em: <<http://www.educacao.sp.gov.br/a2sitebox/arquivos/documentos/238.pdf>>. Acesso em: 07/11/2015.

SÃO PAULO (Estado), Secretaria da Educação. **Agenda de Trabalho**. São Paulo, SEE: 2014a. Disponível em: <https://sis.vunesp.com.br/Login/PegaArquivoPublica.aspx?arquivo=007538_20140929171003.pdf>. Acesso em: 18/02/2017.

SÃO PAULO (Estado), Secretaria da Educação. **Relatório Pedagógico 2014: Matemática**. São Paulo: SEE, 2014b. Disponível em: <http://file.fde.sp.gov.br/saresp/saresp2014/Arquivos/RELATORIO_PEDAGOGICO_MATEMATICA.pdf>. Acesso em: 04/02/2016.

SÃO PAULO (Estado), Secretaria da Educação. **Manual do Apoio Regional e Agente Vunesp**. São Paulo: SEE, 2015a. Disponível em:

<https://sis.vunesp.com.br/Logon/PegaArquivoPublica.aspx?arquivo=015933_20151110171652.pdf>. Acesso em: 17/02/2017.

SÃO PAULO (Estado), Secretaria da Educação. **Sumário Executivo 2014**. São Paulo: SEE, 2015b. Disponível em: <http://file.fde.sp.gov.br/saresp/saresp2014/Arquivos/SUMARIO_EXECUTIVO.pdf>. Acesso em: 16/07/2016.

SARTES, L. M. A. & Souza-Formigoni, M. L. O. **Avanços na Psicometria: Da Teoria Clássica dos Testes à Teoria de Resposta ao Item**. São Paulo: 2013. Disponível em: <<http://www.scielo.br/pdf/prc/v26n2/04.pdf>>. Acesso em: 20/08/2015.

SILVA, Vailton Afonso da. **Questões do ENEM nos livros didáticos de física recomendados pelo PNLD em 2012**. Belo Horizonte: PUC-MG, 2013. Disponível em: <http://www.biblioteca.pucminas.br/teses/EnCiMat_SilvaVA_1.pdf.pdf>. Acesso em: 20/07/2015.

TODD, P. E.; WOLPIN, K. I. **On the specification and estimation of the production function for cognitive achievement**. *Economic Journal*, v. 113, p. 3-33, 2003.

TODD, P. E.; WOLPIN, K. I. **The production of cognitive achievement in children: home, school and racial test score gaps**. *Journal of Human Capital*, v. 1, n. 1, p. 91-136, 2007.

TRIOLA, M. F. **Introdução à estatística**. Rio de Janeiro: LTC, 2005.

WERNECK, Vera R. **Sobre o processo de construção do conhecimento: O papel do ensino e da pesquisa**. *Ensaio: Avaliação de Políticas Públicas. Educação*, Rio de Janeiro, v.14, n.51, p. 173-196, abr./jun. 2006. Disponível em: <<http://www.scielo.br/pdf/ensaio/v14n51/a03v1451.pdf>>. Acesso em: 12/11/2015.