

**Universidade Federal de Pernambuco
Centro de Filosofia e Ciências Humanas
Programa de Pós-Graduação em Psicologia Cognitiva**

Renato Guedes dos Santos

**ECCOs 4/10:
do papel ao teste adaptativo computadorizado**

**Recife
2015**

**UNIVERSIDADE FEDERAL DE PERNAMBUCO
PROGRAMA DE PÓS-GRADUAÇÃO EM PSICOLOGIA
COGNITIVA**

CLASSIFICAÇÃO DE ACESSO A TESES E DISSERTAÇÕES

Considerando a natureza das informações e compromissos assumidos com suas fontes, o acesso a monografias do Programa de Pós-Graduação em Psicologia Cognitiva da Universidade Federal de Pernambuco é definido em três graus:

- “Grau 1”: livre (sem prejuízo das referências ordinárias em citações diretas e indiretas);
- “Grau 2”: com vedação a cópias, no todo ou em parte, sendo, em consequência, restrita a consulta em ambientes de biblioteca com saída controlada;
- “Grau 3”: apenas com autorização expressa do autor, por escrito, devendo, por isso, o texto, ser confiado a bibliotecas que assegurem a restrição, ser mantido em local sob chave ou custódia;

A classificação desta tese se encontra, abaixo, definida por seu autor.

Solicita-se aos depositários e usuários sua fiel observância, a fim de que se preservem as condições éticas e operacionais da pesquisa científica na área da administração.

Título da Tese: **ECCOs 4/10: do papel ao teste adaptativo computadorizado**

Nome do Autor: Renato Guedes dos Santos

Data da aprovação: 23/02/2015

Classificação conforme especificação acima:

Grau 1	<input checked="" type="checkbox"/>
Grau 2	<input type="checkbox"/>
Grau 3	<input type="checkbox"/>

Recife, 23 de fevereiro de 2015

Assinatura do autor

Renato Guedes dos Santos

**ECCOs 4/10:
do papel ao teste adaptativo computadorizado**

Orientador: Prof. Antonio Roazzi, D.Phil.

Tese apresentada ao Programa de Pós-Graduação em Psicologia Cognitiva da Universidade Federal de Pernambuco como requisito para a obtenção do título de Doutor em Psicologia.

Área de concentração: Psicologia Cognitiva

Linha de Pesquisa: Desenvolvimento Cognitivo

**Recife
2015**

Catálogo na fonte
Bibliotecária Maria do Carmo de Paiva, CRB4-1291

S237e Santos, Renato Guedes dos.
ECCOs 4/10 : do papel ao teste adaptativo computadorizado / Renato Guedes dos Santos. – Recife: O autor, 2015.
224 f. : il. ; 30 cm.

Orientador: Prof. Dr. Antonio Roazzi.
Tese (doutorado) - Universidade Federal de Pernambuco. CFCH. Pós-Graduação em Psicologia Cognitiva, 2015.
Inclui referências e anexos.

1. Psicologia Cognitiva. 2. Cognição em crianças. 3. Psicometria. 4. Testes psicológicos. 5. Inovações tecnológicas. I. Roazzi, Antonio (Orientador). II. Título.

153 CDD (22.ed.) UFPE (BCFCH2015-41)

FOLHA DE APROVAÇÃO

Renato Guedes dos Santos

ECCOs 4/10: do papel ao teste adaptativo computadorizado

Tese apresentada ao Programa de Pós-Graduação em Psicologia Cognitiva da Universidade Federal de Pernambuco para obtenção do título de Doutor.

Área de Concentração: Psicologia Cognitiva

Aprovado em: 23 de Fevereiro de 2015

Banca Examinadora

Dr. Antonio Roazzi - Orientador
Universidade Federal de Pernambuco

Dra. Luciana Hodges – Examinador Externo
Faculdade IBGM

Dra. Suely de Melo Santana - Examinador Externo
UNICAP

Dra. Gilda Guimarães - Examinador Interno
Universidade Federal de Pernambuco

Dra. Rafaella Asfora - Examinador Interno
Universidade Federal de Pernambuco

Dedico este trabalho aos meus pais (in memoriam), em especial à minha mãe, Elizabethe Guedes dos Santos, que, apesar do pouco estudo formal, sempre soube se adaptar para superar toda sorte de adversidades que encontrou pelo caminho, conduzindo sua vida com humildade, amor e alegria. Os seus exemplos e conselhos se manterão vivos em mim para sempre.

Agradecimentos

Agradeço à minha família por sempre torcer pelo sucesso dos meus projetos, desempenhando um papel de fundamental importância para a materialização desta e de outras conquistas ao longo da minha vida. Espero retribuir ao carinho e apoio incondicionais que sempre recebi de vocês na mesma intensidade.

Em especial quanto ao esforço de realização deste trabalho, agradeço às minhas filhas Crystal de Menezes Santos, engenheira que me ajudou a desenvolver o programa de computador adaptativo utilizado neste estudo, instrumento essencial para obtenção dos resultados aqui relatados, e Ludmila de Menezes Santos, pedagoga que me ajudou na aplicação de testes junto às crianças, sempre com competência e dedicação.

Meus agradecimentos sinceros ao Prof. Dr. Antonio Roazzi, que assumiu a orientação desse projeto desde o início, do mestrado até o doutorado. Indo sempre além do papel de orientador, brindou-me com sua atenção e disponibilidade, tendo compartilhado comigo seus conhecimentos e a grande rede internacional de relacionamentos da qual faz parte.

Agradeço, também, a todos aqueles que, embora não nomeados, me ajudaram em distintos momentos deste projeto.

“Inteligência é a capacidade de tomar e manter certa direção, adaptar-se a novas situações e de criticar suas próprias ações”

Alfred Binet

Resumo

A utilização de recursos computacionais promete revolucionar a realidade dos testes psicológicos, em particular daqueles voltados para a avaliação cognitiva na infância. Numa lógica de testes que evoluem acompanhando e se adaptando à inovação tecnológica, diríamos que o teste informatizado traduz uma atenção da psicologia à rentabilização das novas tecnologias, podendo até mesmo representar um fator de validade dos testes ao torná-los mais atraentes às pessoas que os realizam. O estado da arte dos testes informatizados reúne a larga capacidade de processamento computacional atualmente disponível com a psicometria moderna, representada pela teoria da resposta ao item (TRI) e a abordagem dos testes adaptativos computadorizados (TAC), possibilitando a construção de testes onde a administração de uma quantidade menor de itens pode produzir estimativas de habilidades mais confiáveis do que nos testes sem o uso desses recursos. Nesse contexto, este estudo utilizou como base o banco de itens da Escala de Avaliação de Competências Cognitivas para Crianças dos 4 aos 10 anos de Idade (ECCOs 4/10), uma bateria de provas voltadas para avaliação cognitiva na infância concebida pelo Instituto de Educação e Psicologia da Universidade do Minho (UMinho), de Portugal, para produzir versões informatizadas e adaptativas de provas da escala. O estudo demonstrou que os testes informatizados podem abreviar e simplificar a aplicação dos testes e a fase de coleta de dados, eliminar transcrições, evitar o desperdício de materiais e ser de fácil transporte quando utilizado em notebooks ou tablets. Além disso, para as crianças os testes computadorizados podem ser mais agradáveis e atraentes do que os testes em lápis e papel, aproximando-os da ideia de jogos eletrônicos, principalmente quando usados com telas sensíveis ao toque. A abordagem adaptativa demonstrou que é possível reduzir drasticamente a duração dos testes e a quantidade de itens que devem ser respondidos pelos participantes e ainda produzir estimativas de habilidades consistentes. Por outro lado, o estudo também concluiu que seria necessário ampliar e recalibrar o banco de itens da ECCOs 4/10 para que os testes adaptativos dessem conta adequadamente das diversas faixas etárias atendidas, em especial para as crianças de tenra idade, sendo recomendada a utilização de banco de itens separados ou recorrer a abordagens multidimensionais.

Palavras-chave: ECCOs 4/10. avaliação cognitiva. testes adaptativos computadorizados. teoria da resposta ao item. psicometria

Abstract

The use of computational resources promises to revolutionize the reality of psychological tests, particularly those focused on the cognitive assessment in childhood. Believing that the psychological tests must also follow and adapt to technological innovation, we can say that the computerized test is an attention of psychology in the face of new technologies and may even represent a factor of validation of tests to make them more attractive to people who perform them. The current state of the art of computerized testing meets the large computational processing capacity currently available with the techniques of modern psychometrics, represented by item response theory (IRT) and the approach of computer-adaptive testing (CAT), allowing the construction of tests where administration to a few items can produce more reliable ability estimates than the estimates produced by the long tests that do not use these features. In this context, this study used the item bank of Escala de Avaliação de Competências Cognitivas para Crianças dos 4 aos 10 anos de Idade (ECCOS 4/10), a battery of tests aimed to cognitive assessment in children conceived by the Instituto de Educação e Psicologia of Universidade do Minho (UMinho), Portugal, to produce computerized and adaptive versions of the scale tests. The study showed that computerized tests can shorten and simplify the application of the tests and data collection, can eliminate transcripts, avoid wasting materials and be easy to transport when used in notebooks or tablets. In addition, for children the computerized tests may be more pleasant and attractive than pencil and paper tests, bringing them closer to the idea of electronic games, especially when used with touch screens. The adaptive approach demonstrated that it is possible to dramatically reduce the duration of the tests and the amount of items that should be answered by the participants and still produce consistent estimates of abilities. On the other hand, the study also concluded that it is necessary to expand and recalibrate the ECCOS 4/10 items bank for adaptive tests could cope adequately with the various age groups, especially for young children, and recommended use the multidimensional approaches or even use separate items bank in the future.

Keywords: ECCOS 4/10. cognitive assessment. computer-adaptive testing. item response theory. psychometrics

Lista de Figuras

Figura 1 - Curva característica do item (CCI)	42
Figura 2 - Esquema de administração do teste adaptativo de Binet em 1905.....	61
Figura 3 - Exemplo de algoritmo para um teste adaptativo computadorizado	68
Figura 4 - Exemplo de administração de itens em teste adaptativo. A cada item administrado o programa vai melhorando a estimativa de habilidade e diminuindo a estimativa do erro da medida.....	78
Figura 5 - Exemplo de item da prova não verbal Comparação de Figuras	89
Figura 6 - Exemplo de item da prova não verbal Elementos em Árvore. Primeiro, mostra-se a árvore com os elementos. Depois de algum tempo a mesma árvore é mostrada sem os elementos e a criança deverá indicar onde eles estavam	91
Figura 7- Exemplo de item da prova não verbal Desenhos Absurdos	92
Figura 8 - Exemplo de item da prova não verbal Imagens Incompletas.....	93
Figura 9 - Exemplo de tarefa da prova não verbal Construção de Padrões.	94
Figura 10 - Exemplo de tarefa da prova Construção de Figuras - com 6 barrinhas de madeira, construir todas as figuras que quiser dentro de um tempo limite.	96
Figura 11 - Adaptação de item na prova Desenhos Absurdos. As crianças não percebiam nenhum absurdo na figura utilizada em Portugal.	104
Figura 12 - Adaptação de item na prova Situações quantitativas. As crianças não compreendiam o significado da palavra meta.....	104
Figura 13 - Correlações entre tarefas, países, idade e dimensão verbal / não verbal.....	107
Figura 14 – Mapa comparativo entre pessoas e itens no continuum da variável latente.	119
Figura 15 - Boxplot idade x pontuação total.....	122
Figura 16 - Código XML que descreve o item 1 da prova Figuras Incompletas. O bloco <pergunta> define os arquivos de imagem que devem ser apresentados, enquanto o bloco <alternativas> designam as imagens das respostas. A resposta correta é a opção A.....	130
Figura 17 - Representação da diferença de traço nas ilustrações do teste com lápis e papel (à esquerda) e no teste informatizado (à direita) na prova Figuras Incompletas	131
Figura 18 - Exemplo real de desempenho no teste adaptativo de raciocínio - prova Figuras Incompletas: o programa vai ajustando a estimativa de habilidade do aluno na medida em que mais itens são administrados.....	159

Figura 19 - Exemplo real de desempenho no teste adaptativo de percepção – prova Comparação de Figuras. Quanto mais itens são administrados, menor o erro estimado para a medida. No caso, foi necessário administrar 21 itens para se obter uma estimativa de erro aceitável para a estimativa de habilidade. 159

Figura 20 - Mapa de itens e alunos referente à prova de percepção e raciocínio. A pouca quantidade de itens prejudica a seleção de itens com dificuldade próximas e expõe os itens para o público. Os itens de raciocínio possuem um equilíbrio melhor na distribuição ao longo da escala. 168

Lista de Quadros

Quadro 1 - Estrutura e provas da ECCOs 4/10	87
Quadro 2 - Distribuição das faixas etárias na amostra portuguesa.	101
Quadro 3 - Distribuição das faixas etárias na amostra brasileira.	102
Quadro 4 - Transformações subjacentes aos itens da prova de Imagens Incompletas.	113

Lista de Tabelas

Tabela 1 - Distribuição amostra brasileira (n=606) por faixa etária, sexo e tipo de escola... 102	102
Tabela 2 - Amostras brasileira e portuguesa por idade, sexo e tipo de escola..... 103	103
Tabela 3 – Classificação dos participantes por idade e sexo. 117	117
Tabela 4 – Sumário da medida de 272 pessoas..... 120	120
Tabela 5 - Sumário da medida de 42 itens. 120	120
Tabela 6 – Saída da Análise dos Principais Componentes (em eigenvalue). 121	121
Tabela 7 - Classificação dos participantes da Prova Elementos em Árvores 133	133
Tabela 8 - Médias e desvios da prova Elementos em Árvores. 136	136
Tabela 9 - Classificação dos participantes da prova Comparação de Figuras. 142	142
Tabela 10 - Distribuição dos participantes no estudo piloto por idade e sexo..... 154	154
Tabela 11 - Distribuição dos participantes do teste formal por idade e sexo 154	154
Tabela 12 - Itens administrados versus estimativas de habilidade 160	160
Tabela 13 - Quantidade de itens administrados por tipo de prova..... 162	162
Tabela 14 – Determinação da correlação de Pearson entre as provas sequenciais e adaptativas de raciocínio e percepção. 164	164
Tabela 15 - Sumário das medidas de alunos e itens na prova de percepção..... 165	165
Tabela 16 - Mapa de dimensionalidade. 166	166

Lista de Siglas e Abreviaturas

APC - análise dos principais componentes

API - application program interface

ASBAV - Armed Services Vocational Aptitude Battery

BPFS - Back Pain Functional Scale

CAT - computer-adaptive testing ou computerized adaptive testing

CA-AT - Computer Adaptive-Attribute Testing

ECCOs 4/7 - Escala de Avaliação de Competências Cognitivas para Crianças dos 4 aos 7 Anos de Idade

ECCOs 4/10 - Escala de Avaliação de Competências Cognitivas para Crianças dos 4 aos 10 Anos de Idade

CCI - curva característica do item

CCT - curva característica do teste

CESPE/UnB - Centro de Seleção e de Promoção de Eventos da Universidade de Brasília

CNS - Conselho Nacional de Saúde

EAP - expected a posteriori estimator

ENCCEJA - Exame para Certificação de Competências de Jovens e Adultos

ENEM - Exame Nacional do Ensino Médio

ETS - Educational Testing Service

FairTest - The National Center for Fair & Open Testing

GRE - Graduate Record Examination

Inep - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

Infit - inlier-sensitive fit

IRT - item response theory

Logit - log odds unit

KR-20 - Kuder-Richardson fórmula 20

MAP - maximum a posteriori estimator

MDT - measurement decision theory

MEC - Ministério da Educação

MEPS - Military Entrance Processing Stations

MIRT - multi-dimensional item response theory

ML - maximum-likelihood estimator

ML1 - modelo logístico de um parâmetro
ML2 - modelo logístico de dois parâmetros
ML3 - modelo logístico de três parâmetros
NAQ-R - Negative Acts Questionnaire-Revised
OCDE - Organização para a Cooperação e Desenvolvimento Econômico
Outfit - outlier-sensitive fit
PIRLS - Progress in International Reading Literacy Study
PISA - Programme for International Student Assessment
RIA - rich internet applications
SAEB - Sistema de Avaliação da Educação Básica
SAT - Scholastic Assessment Test
SEM – standard error of the mean
SQL - Structured Query Language
TRI - teoria da resposta ao item
TAC - teste adaptativo computadorizado
TAI - teste adaptativo informatizado (mesmo significado de TAC)
TCT - teoria clássica dos testes
TIMSS - Trends in International Mathematics and Science Study
TOEFL - Test of English as a Foreign Language
Unesp - Universidade Estadual Paulista
UFPE - Universidade Federal de Pernambuco
UMinho - Universidade do Minho
XML - eXtensible Markup Language
WLE - weighted likelihood estimator

Sumário

INTRODUÇÃO	21
ORGANIZAÇÃO DO TRABALHO	26
1. REFERENCIAL TEÓRICO	30
1.1. A medida em psicologia	30
1.2. Relação entre o computador e a testagem	36
1.3. Teoria da resposta ao item (TRI).....	39
1.3.1. Unidimensionalidade e independência local.....	44
1.3.2. Modelos da TRI	46
1.3.3. Validade, fidedignidade.....	50
1.3.4. Separação.....	54
1.3.5. Estatísticas de aderência ao modelo	55
1.3.6. Unidimensionalidade em Rasch	57
1.4. Teste adaptativo computadorizado (TAC).....	60
1.4.1. Breve história dos testes adaptativos.....	60
1.4.2. Quem está usando os testes adaptativos computadorizados	63
1.4.3. Como o TAC funciona	67
1.4.3.1. Conjunto de itens calibrados.....	70
1.4.3.2. Estimativas de habilidades.....	71
1.4.3.3. Nível de habilidade de entrada (primeiro item).....	73
1.4.3.4. Algoritmo de seleção do próximo item	74
1.4.3.5. Procedimento de pontuação	75
1.4.3.6. Critério (ou regra) de parada	76
1.4.4. Exemplo de TAC	78
1.4.5. Vantagens do TAC	80
1.4.6. Desvantagens do TAC	81
1.4.7. Teste adaptativo multidimensional.....	82
2. ESCALA DE COMPETÊNCIAS COGNITIVAS PARA CRIANÇAS DOS 4 AOS 10 ANOS DE IDADE (ECCOs 4/10)	85
2.1. Introdução.....	85
2.2. A estrutura da ECCOs 4/10	86
2.3. Processos cognitivos da ECCOs 4/10	88
2.3.1. Percepção.....	88
2.3.2. Memória.....	89
2.3.3. Compreensão.....	91
2.3.4. Raciocínio	92
2.3.5. Resolução de problemas.....	93
2.3.6. Pensamento divergente	95
3. ESTUDO 1 – VALIDAÇÃO E ADAPTAÇÃO DA ECCOs 4/10 PARA O BRASIL	98
3.1. Introdução.....	98
3.2. Recapitulando a ECCOs 4/10	98
3.3. Justificativas Validação e adaptação da ECCOs 4/10 para o Brasil.....	99
3.4. Proposta do estudo.....	100
3.5. Procedimento	100
3.6. Descrição da amostra brasileira.....	101

3.7.	Exemplos de adaptações de itens para uso no Brasil.....	103
3.8.	Análise dos dados	106
4.	ESTUDO 2 – PROVA FIGURAS INCOMPLETAS INFORMATIZADA COM TRI	109
4.1.	Introdução.....	109
4.2.	Resgatando um pouco do referencial teórico	111
4.2.1.	A medida em psicologia	111
4.2.2.	Figuras Incompletas - a prova de raciocínio não verbal da ECCOs 4/10	112
4.2.3.	A teoria da resposta ao item.....	114
4.3.	O estudo empírico	115
4.3.1.	Considerações iniciais.....	115
4.3.2.	Objetivos	115
4.3.3.	Método	117
4.3.4.	Resultados.....	118
4.3.5.	Conclusões.....	123
5.	ESTUDO 3: INFORMATIZAÇÃO DAS PROVAS SEQUENCIAIS.....	127
5.1.	Desenvolvimento dos programas de computador	127
5.1.1.	Interface.....	128
5.1.2.	Linguagem de programação.....	129
5.1.3.	Banco de Itens	130
5.1.4.	Banco de dados.....	131
5.2.	Versão informatizada sequencial da prova de memória Elementos em Árvore.....	132
5.2.1.	Objetivo	132
5.2.2.	Método	132
5.2.3.	Resultados.....	135
5.3.	Versão informatizada sequencial da prova de compreensão Desenhos Absurdos	137
5.3.1.	Objetivo	137
5.3.2.	Método	138
5.3.3.	Resultados.....	139
5.4.	Versão informatizada sequencial da prova de percepção Comparação de Figuras.....	141
5.4.1.	Objetivo	141
5.4.2.	Método	142
5.4.3.	Resultados.....	144
6.	ESTUDO 4: TESTE ADAPTATIVO COMPUTADORIZADO (TAC)	146
6.1.	Introdução.....	146
6.2.	A escolha das provas.....	147
6.3.	Construção do programa de computador adaptativo.....	151
6.4.	O estudo empírico	153
6.4.1.	Objetivos	153
6.4.1.1.	<i>Geral</i>	153
6.4.1.2.	<i>Específicos</i>	153
6.4.2.	Método	154
6.4.2.1.	<i>Participantes</i>	154
6.4.2.2.	<i>Instrumentos e procedimentos</i>	155
6.4.2.3.	<i>Procedimentos</i>	157
6.4.3.	Resultados.....	158
7.	CONSIDERAÇÕES GERAIS E RECOMENDAÇÕES	174
	REFERÊNCIAS	186

ANEXO A – Telas da versão informatizada da escala ECCOs 4/10	196
ANEXO B – Estatísticas dos Itens do SPSS da prova de Figuras Incompletas	199
ANEXO C – Estatísticas dos Itens do Winsteps prova Figuras Incompletas	200
ANEXO D – Curvas dos itens, do teste e de informação do teste da prova Figuras Incompletas ...	201
ANEXO E - Termo de Consentimento Livre e Esclarecido	202
ANEXO F - Procedimentos para aplicação do teste Figuras Incompletas	204
ANEXO G - Procedimentos para aplicação do teste Comparação de Figuras	207
ANEXO H - Procedimentos para aplicação do teste Elementos em Árvore	210
ANEXO I - Procedimentos para aplicação do teste Desenhos Absurdos	212
ANEXO J – Ficha de informação do aluno	215
ANEXO K – Estrutura dos bancos de dados.....	216
ANEXO L - Itens da prova Elementos em Árvore - processo de memória	218
ANEXO M - Itens da prova Figuras Incompletas - processo de raciocínio	220
ANEXO N - Itens da prova Comparação de Figuras - processo de percepção	226
ANEXO O - Itens da prova Desenhos Absurdos - processo de compreensão.....	230

INTRODUÇÃO

INTRODUÇÃO

Desde outubro de 1997 que o Instituto de Educação e Psicologia da Universidade do Minho - UMinho, de Portugal, vem desenvolvendo e aperfeiçoando uma escala para avaliação da inteligência em crianças. A versão mais recente dessa escala, fruto do trabalho da Dra. Maria de Lurdes Dias Brito sob a orientação do Prof. Dr. Leandro Almeida, data de 2009 e foi denominada Escala de Avaliação de Competências Cognitivas para Crianças dos 4 aos 10 Anos de Idade – ECCOs 4/10. Considerando que esse grande projeto ainda poderia ser ampliado, Brito (2009) elencou entre as possibilidades de estudos futuros a adaptação da ECCOs 4/10 para populações de outros países de expressão portuguesa e também a elaboração de uma versão informatizada da escala.

Foi olhando para esses objetivos indicados pelos criadores da escala que o Programa de Pós-Graduação em Psicologia Cognitiva do Departamento de Psicologia da Universidade Federal de Pernambuco - UFPE, tradicional parceiro da UMinho em pesquisas no campo da psicologia cognitiva, integrou-se ao projeto e configurou duas frentes de trabalhos a partir do ano de 2009, indo ao encontro do desejo dos autores da escala e dos interesses dos centros acadêmicos envolvidos. Enquanto a primeira frente se debruçou sobre a validação e adaptação da ECCOs 4/10 para uso com a população brasileira, conservando o modelo em lápis e papel da sua versão original, a segunda frente objetivou realizar estudos para o desenvolvimento de testes informatizados tomando a escala portuguesa como base. O trabalho no Brasil foi distribuído em diferentes fases e equipes, mas esteve sempre conduzido pelo Prof. Dr. Antonio Roazzi (UFPE).

Nesse contexto, este estudo está inserido na frente de pesquisa referente à informatização das provas da ECCOs 4/10, continuando, ampliando e enriquecendo o

trabalho que fora apresentado em 2011 na dissertação de mestrado deste mesmo autor. Na época, produziu-se uma versão informatizada da prova não verbal da ECCOs 4/10 destinada à avaliação do raciocínio abstrato denominada Figuras Incompletas. Aquele trabalho fomentou o aprofundamento de estudos sobre métodos, características, vantagens, desvantagens e limites de escalas informatizadas, além de introduzir os procedimentos metodológicos da teoria da resposta ao item (TRI) na prova, servindo como fase preparatória de fundamental importância para este trabalho atual.

Desta vez os estudos resultaram na informatização, aplicação e validação das provas não verbais da ECCOs 4/10 Elementos em Árvores, Comparação de Figuras e Figuras Absurdas, que avaliam os processos cognitivos de memória, percepção e compreensão, respectivamente. Além disso, e, principalmente, sob o ponto de vista tecnológico, para as provas Comparação de Figuras e Figuras Incompletas, também foram desenvolvidas versões adaptativas de testes utilizando os mesmos itens das provas em lápis e papel. Assim, dispõe-se, hoje, de um total de seis provas informatizadas baseadas na ECCOs 4/10, dando conta de quatro processos psicológicos. Quatro dessas provas, referentes aos processos de memória, percepção, compreensão e raciocínio, foram construídas no modelo sequencial, onde os itens são apresentados de acordo com uma sequência previamente estabelecida. As duas provas desenvolvidas no modelo adaptativo dão conta dos processos de percepção e raciocínio.

Importante frisar que o estudo não se limitou a transpor as provas do papel para a tela do computador. Além de aumentar a oferta de provas computadorizadas para avaliação de processos cognitivos, através da convergência de visões das áreas de informática, matemática e psicologia, pretendeu-se reforçar o conhecimento sobre aplicações e limites processuais dos testes informatizados ao tempo em que se adquiria domínio sobre mais uma técnica de uso inédito no Departamento de Psicologia da UFPE: a dos testes adaptativos realizados em computador, sensíveis aos participantes, ajustados automaticamente à capacidade cognitiva

de cada pessoa, normalmente englobados pelo termo teste adaptativo computadorizado (TAC), oriundo do termo em inglês computer-adaptive testing (CAT). Apesar da história do CAT ter seu início remontado aos testes para aferição de inteligência produzidos por Alfred Binet, ainda em 1905, somente nos últimos anos é que eles estão se tornando viáveis e populares em função do aumento de poder de processamento, miniaturização e facilidade de acesso aos computadores, bem como devido à evolução e simplificação das linguagens de programação e suas bibliotecas.

Essa perspectiva de aquisição de novos conhecimentos e de evolução de processos metodológicos, partindo de uma prova tradicional, concebida para uso com lápis e papel, para um teste adaptativo computadorizado, por exemplo, foi uma marca permanente durante toda a realização deste estudo. Afinal, os testes adaptativos computadorizados norteiam a visão atualizada dos testes no âmbito da psicometria, o ramo da psicologia que se utiliza de muita matemática e estatística para "explicar o sentido que têm as respostas dadas pelas pessoas a uma série de itens e propor técnicas de medida dos processos mentais" (Pasquali, 2009).

O estudo demonstrou que os testes informatizados podem se configurar como aliados dos psicólogos por simplificar a aplicação dos testes e toda a fase de coleta de dados, eliminando transcrições e erros decorrentes, reduzir desperdícios de materiais e por ser de fácil transporte se utilizado em notebooks ou tablets. Para as crianças os testes computadorizados podem apresentar vantagens adicionais ao parecerem mais agradáveis e atraentes para quem os realiza, até mesmo afastando um pouco a ideia de que se trata de uma avaliação para dar a impressão de que se trata de um jogo eletrônico, principalmente quando usados com telas sensíveis ao toque. Além disso, apesar de requerer especialização, análise quanto a sua conveniência, planejamento adequado e cuidados específicos para sua elaboração e aplicação, a abordagem adaptativa demonstrou que é possível reduzir drasticamente o tempo de duração dos testes e também a quantidade de itens que devem ser

respondidos pelos participantes e ainda aumentar a confiança nas estimativas de habilidades produzidas. Em geral, o TAC necessita de uma quantidade bem menor de itens respondidos (50% é típico) para produzir escores confiáveis e comparáveis, implicando em menor tempo de aplicação e maior satisfação para quem faz os testes. Em contrapartida, o estudo concluiu que na ECCOs 4/10 apenas as provas não verbais Comparação de Figuras e Figuras Incompletas admitem versões adaptativas e, mesmo assim, para que pudesse ser utilizada na prática seria necessário ajustar e calibrar o banco de itens para dar conta das diversas faixas etárias atendidas pela ECCOs 4/10, em especial para as crianças de tenra idade.

É fácil perceber que ao beneficiar a escala ECCOs 4/10 com provas automatizadas e provas adaptativas que complementam e estendem a versão original, em lápis e papel, estamos lhe conferindo um importante diferencial frente outras abordagens. Assim, podemos entender que são beneficiários diretos deste trabalho aqueles profissionais que se valem da mensuração de habilidades cognitivas no seu dia a dia, bem como todas as crianças que vierem a ser usuárias da escala no formato informatizado algum dia.

Do ponto de vista do Departamento de Psicologia da UFPE, vale ressaltar que a aquisição e o domínio de tecnologias contemporâneas constituem uma demarcação estratégica necessária para qualquer centro de estudos que queira estar na vanguarda das pesquisas. Na área da testagem, os testes adaptativos computadorizados (TAC) e a teoria da resposta ao item (TRI) expressam técnica e modelo que definem o estado da arte dos testes para avaliação psicométrica. Apesar de já utilizados em algumas situações, tanto a TRI quanto a TAC são métodos ainda pouco explorados no campo dos testes psicológicos no Brasil, como veremos com mais detalhes no capítulo I, e, portanto, espera-se que a experiência com este projeto possa ser replicada a outros estudos, fortalecendo a participação do Departamento nessa linha de pesquisa.

Por fim, não podemos deixar de considerar os benefícios indiretos advindos das pesquisas realizadas para este estudo para os próprios pesquisadores. Sem dúvida, esses novos domínios de conhecimentos serão úteis para futuros trabalhos que possam se interessar pela psicométrica moderna, seja no aspecto teórico ou prático.

ORGANIZAÇÃO DO TRABALHO

Esta tese foi estruturada em sete capítulos de forma a conduzir o leitor através do roteiro que foi seguido para a produção dos resultados aqui descritos, navegando pelo entendimento dos princípios da escala ECCOs 4/10 concebida pela Universidade do Minho, passando pelo trabalho referente a sua adaptação para a população brasileira, pelos achados sobre testes informatizados e pela aplicação da teoria da resposta ao item sobre a prova de raciocínio, até a produção e aplicação de testes adaptativos computadorizados utilizando as provas Comparação de Figuras e Figuras Incompletas, culminando com uma comparação dos resultados e reflexões sobre o trabalho desenvolvido.

Ressaltamos que os estudos relatados nos Capítulos III e IV, referentes à validação e adaptação da ECCOs 4/10 para o Brasil e ao desenvolvimento da versão informatizada da prova Figuras Incompletas, não fazem parte diretamente do esforço realizado durante esta tese, pois retratam trabalhos anteriores do autor (versão informatizada da prova Figuras Incompletas) ou estudos paralelos realizados por outros autores (validação e adaptação da ECCOs 4/10 para o Brasil). No entanto, o presente estudo se utiliza dos dados e aquisições desses dois estudos e os estende, em especial no tocante às provas informatizadas, inclusive fazendo a migração da prova Figuras Incompletas para a linguagem Java. Assim, esses estudos anteriores foram descritos neste documento em capítulos específicos visando encapsular a história da escala portuguesa no Brasil e colaborar com a compreensão da trajetória do estudo que agora mais nos interessou, que é a ampliação da oferta de provas informatizadas e a criação de testes adaptativos computadorizados baseados em provas da ECCOs 4/10.

Desta forma, o Capítulo I reflete o referencial teórico que serviu de base para este estudo, visitando conceitos e técnicas dos métodos que serão utilizados ao longo do trabalho.

A relação da testagem com os computadores, a teoria da resposta ao item, o modelo de Rasch e a abordagem dos testes adaptativos computadorizados estão lá abordados. O capítulo também traz algumas percepções quanto à evolução das teorias que embasam a psicometria, trazendo-a para dentro das possibilidades atuais do uso de computadores na geração e aplicação de testes psicológicos que se adaptam à capacidade cognitiva dos participantes.

No Capítulo II apresentamos a Escala de Competências Cognitivas para Crianças dos 4 aos 10 anos de idade (ECCOs 4/10) tal como foi concebida na versão produzida pela UMinho em 2009. Na verdade, trata-se de um prolongamento do referencial teórico, mas que por causa da sua relevância para o trabalho foi destinado um capítulo à parte. A visão composta de inteligência da escala é apresentada através das suas onze provas, entre verbais e não verbais, direcionadas a processos cognitivos diferenciados.

Já o Capítulo III fala sobre o estudo de validação e adaptação da ECCOs 4/10 na sua versão tradicional, em lápis e papel, para a população brasileira. Descreve resumidamente como foi a aplicação das provas no Nordeste do Brasil para mais de seiscentas crianças e jovens, relatando algumas dificuldades e curiosidades percebidas no processo, apresentando uma breve e preliminar análise dos dados obtidos naquele estudo.

No Capítulo IV é apresentado o estudo realizado em 2010 que relata a informatização da prova não verbal de raciocínio abstrato da ECCOs 4/10, denominada Figuras Incompletas, alvo da primeira fase de informatização da escala, que acabou por fornecer subsídios ao trabalho atual. A prova de raciocínio, que fora concebida pelo olhar clássico das teorias dos testes, teve seus itens e resultados analisados pela visão da teoria da resposta ao item, elemento fundamental da moderna psicometria.

O Capítulo V relata o estudo para desenvolvimento das provas Elementos em Árvores, Desenhos Absurdos e Comparação de Figuras no modelo sequencial, onde os itens

são apresentados seguindo uma sequência definida pela ordem de dificuldade. Essas três novas provas se juntam à prova Figuras Incompletas, concebida anteriormente, fazendo com que a ECCOs 4/10 passe a dispor de quatro provas informatizadas no modelo sequencial.

O Capítulo VI descreve o estudo para elaboração e aplicação de testes adaptativos computadorizados baseados nas provas Comparação de Figuras e Figuras Incompletas, referentes aos processos de raciocínio e percepção da ECCOs 4/10, importante eixo de inovação relacionado ao estudo descrito neste documento. Com a inclusão das provas em modelo adaptativo, a ECCOs 4/10 passou a dispor de um total de seis provas, sendo cinco novas, compreendendo quatro provas no modelo sequencial e duas no modelo de teste adaptativo. Ressalta-se que essa categoria de testes representa o que há de mais atual no campo da testagem.

Por fim, no Capítulo VII temos as conclusões e as discussões de ordem geral, os comentários finais e as recomendações para prosseguimento com novos estudos no futuro.

Anexados ao trabalho temos exemplos de provas em papel, de telas de provas informatizadas, a estrutura de dados modelada para o teste, curvas e estatísticas geradas através dos programas SPSS e Winsteps, o termo de consentimento dos responsáveis pelos alunos, os procedimentos para aplicação de cada um dos testes, a ficha de informação acadêmica do aluno, a representação das bases de dados e a representação de todos os itens das quatro provas informatizadas e abordadas ao longo desse estudo.

Capítulo I

Referencial Teórico

1. REFERENCIAL TEÓRICO

1.1. A medida em psicologia

Os testes psicológicos são importantes instrumentos que auxiliam a prática profissional do psicólogo na realização de avaliações psicológicas, no ensino e na pesquisa. De fato, a medição é uma atividade de fundamental importância para a ciência. Segundo Pasquali (2003), uma ciência se constrói com as variáveis que ela estuda e cada área da ciência desenvolve seu próprio conjunto de procedimentos de medição. Nós adquirimos conhecimento sobre pessoas, objetos, eventos e processos através da observação. Para que essas observações ganhem sentido muitas vezes precisamos quantificá-las, ou seja, medir as coisas em que temos interesse científico.

Segundo DeVellis (2003), dentro das ciências sociais e comportamentais, a psicometria, termo que define a medida em psicologia, tem se destacado como uma especialização em benefício da medição de fenômenos psicológicos e sociais. Normalmente, esse procedimento de medição envolve testes (provas, questionários) e as variáveis de interesse são parte de um amplo referencial teórico, procurando-se explicar o sentido que têm as respostas dadas pelos participantes a uma série de tarefas, tipicamente chamadas de itens, normalmente agrupadas em um ou mais testes. Por sua vez, "teste é um processo ou um instrumento padronizado que fornece informações sobre uma amostra de comportamento ou de processos cognitivos de maneira quantificada" (Hogan, 2006, p.30).

Em sentido lato, a definição clássica para o termo "medição" ou "mensuração" foi fixada pelo psicólogo Stanley Smith Stevens nos anos 1940 quando classificou as quatro escalas (nominal, ordinal, intervalar e razão) para medidas em ciências: medir é assinalar números para objetos ou eventos de acordo com regras. O fato de que números podem ser atribuídos segundo regras diferentes leva a diferentes tipos de escalas e diferentes tipos de

medidas (Stevens, 1946). No entanto, essa é uma definição que limita a função da medida, pois medir nas Ciências Sociais representa bem mais do que apenas assinalar números para objetos ou eventos usando regras preestabelecidas. No nosso caso, a medição é um processo pelo qual nós tentamos compreender a própria natureza da variável (Bridgman, 1928) aplicando técnicas matemáticas.

De acordo com Pasquali (2003), "a origem da psicometria deve ser procurada nos trabalhos do estatístico Charles Spearman (1863-1945) que, por sua vez, seguiu os procedimentos fisicalistas da obra de Francis Galton (1822-1911) no que se refere à psicologia" (p.14). Adicionalmente, o mesmo autor identifica duas orientações independentes, uma com preocupação mais prática, voltada ao clínico e ao psicopedagógico, e a outra, mais preocupada com o desenvolvimento da própria teoria psicométrica, perseguida por psicólogos de orientação estatística. As duas correntes se unificariam posteriormente na chamada psicometria clássica.

Pasquali (2003) ensina que a psicometria se insere dentro da teoria da medida em geral e esta, por sua vez, desenvolve uma discussão epistemológica sobre a utilização do número no estudo científico dos fenômenos naturais. Ou seja, trata-se de uma interface entre sistemas teóricos de saber diferentes, tendo a teoria da medida a função de justificar e explicar o sentido que tal interface possui. É importante salientar que a psicometria é um ramo da psicologia e não da estatística. Portanto, suas variáveis de trabalho precisam adquirir conteúdos psicológicos, visto que a psicologia não tem como objeto de estudo parâmetros, mas, sim, processos comportamentais, processos psíquicos.

Já o termo "variável" representa qualquer coisa que possa ser medida e que possa diferir entre entidades ou através do tempo. Em psicologia, normalmente estamos interessados em medir comportamentos e processos psicológicos, variáveis que nem sempre são possíveis de serem observadas. Alguns exemplos de variáveis contínuas em psicologia

são inteligência, extroversão, desajustamento, ciúmes, ansiedade aprendizado e acuidade visual. As pessoas que estudamos variam ao longo dessas variáveis, apresentando escores que podem ser altos ou baixos, existir em maior ou menor quantidade ou se diferenciarem por questões ambientais e sociais ou, ainda, por outros conjuntos similares de quantificadores.

As variáveis que não podem ser medidas diretamente pela observação são chamadas de variáveis latentes. Ayala (2009) diz que, ao contrário da informação colhida diretamente da observação do comportamento, o termo latente se refere à informação obtida a partir de pressupostos ou inferências sobre dados oriundos dos comportamentos, ou seja, as variáveis latentes não são observáveis, mas podem ser acessadas. Ferramentas matemáticas, como a teoria da resposta ao item (TRI) ou a teoria clássica dos testes (TCT), podem ser aplicadas para explicar comportamentos observáveis a partir da perspectiva de variáveis latentes contínuas.

Por sua vez, o conceito de traço latente é permeado por ambiguidades e controvérsias, bastando ver a quantidade de expressões que o representam. Pasquali (2003) dá exemplos de alguns: variável hipotética, variável fonte, fator, construto, conceito, estrutura psíquica, traço cognitivo, processo cognitivo, processo mental, habilidade, aptidão, traço de personalidade, processo elementar de informação, componente cognitivo, tendência e atitude, entre outros. As concepções de traço latente dependem do nível de especificidade que se quer dar a esse construto ou parâmetro. Por exemplo, a capacidade de uma pessoa resolver equações matemáticas pode ser considerada um traço latente, mas esta habilidade, por sua vez, pressupõe a capacidade de somar, subtrair, multiplicar e dividir que também poderiam ser concebidas como traços latentes. Como esclarece Pasquali (2003), a psicometria trabalha com o conceito fatorista de traço latente, ou seja, para a psicometria o traço latente deve ser concebido como um processo psicológico macro, não reducionista, compreendendo uma rotina de execução de tarefas que pode englobar diversas etapas.

Os testes psicológicos pressupõem que qualquer traço latente reconhecível que tenha algum grau de estabilidade e que possa ser quantificado num continuum poderá ser medido. As escalas utilizadas para medição em psicologia correspondem a processos onde se estabelecem as correspondências entre os dados observados e a localização das pessoas no continuum da variável latente, de forma que seja possível compará-las com outras pessoas. Como parte da medição também se faz necessário determinar as definições operacionais das variáveis de interesse. Além disso, o processo de medição deve avaliar a quantidade de informação que as medidas geram sobre os participantes e o quanto essas medidas refletem a variável latente.

Ainda segundo Pasquali (2003), quase todos os atributos em ciências psicossociais são mensuráveis apenas com base em leis e teorias, não se tratando de medidas fundamentais, visto que estas devem permitir operações de associação e concatenação sobre seus atributos mensuráveis, nem de medidas derivadas, que se apoiam em relações sobre medidas fundamentais. Assim, as medidas em psicologia, particularmente na teoria dos testes psicológicos, também são medidas por teoria e trabalham com dois parâmetros: a resposta fornecida pela pessoa (seu comportamento) e o critério. Na TCT o critério representa o desempenho da pessoa, ou seja, também é entendido como comportamento (futuro), enquanto na TRI o critério é o traço latente, causando grande distinção entre essas duas teorias.

Algumas questões são fundamentais para construção de testes com valor científico. O primeiro ponto envolve a fidedignidade do teste, ou seja, a consistência, a estabilidade da medida, pois ela afeta a nossa confiança na medição. Os testes devem ser fidedignos, capazes de medir sem erros. Medir sem erros significa que "o mesmo teste, medindo as mesmas pessoas em ocasiões diferentes, ou testes equivalentes, medindo as mesmas pessoas na mesma ocasião, produzem resultados idênticos" (Pasquali, 2003).

O segundo ponto se refere à validade da medida. Embora existam várias discussões sobre validade de um teste, é comum admitir que um teste é válido se de fato mede o que supostamente deve medir. Lembrando que a existência da medida não significa, necessariamente, que a medida resulta de uma reflexão acurada da variável do interesse teórico, para nossos propósitos vamos entender a validade como sendo o grau em que nossas medidas representam, realmente, manifestações da variável latente de interesse.

O terceiro ponto é que o instrumento usado para a medição não pode ser influenciado por nenhuma das medidas que ele mesmo realizou, ou seja, o instrumento deve possuir a propriedade de invariância, caso contrário as comparações entre os diferentes sujeitos teriam utilidade limitada (Ayala, 2009).

O quarto ponto é que quando interpretamos e usamos nossos dados precisamos ter em conta os diferentes tipos de informação que podem decorrer das observações ou das medidas. Dependendo da abordagem utilizada para medir a variável latente a medida pode carregar certas propriedades que afetem como nós interpretamos sua informação. Ou seja, quando formos interpretar nossos dados devemos fazer uma análise qualitativa antes de quantificá-los. Vieses de itens, questões relativas à unidimensionalidade do teste, itens que pouco ou nada discriminam, pessoas com habilidades extremas e respostas desleixadas são fatores que podem prejudicar a correta interpretação quantitativa dos resultados. Softwares para análise dos dados, tal como o Winsteps, utilizado neste estudo, podem ajudar a identificar e eliminar situações indesejadas de diversas dessas naturezas.

Por fim, vale dizer que os escores dos testes precisam ser interpretados corretamente, utilizando modelos estatísticos e normas baseadas nos escores de grandes grupos. Estas normas nos permitem estimar o quanto um desempenho individual se aproxima ou se afasta da média do grupo ou das pontuações obtidas pela generalidade dos sujeitos da amostra de referência. Quando tais amostras são grandes e devidamente heterogêneas, os resultados

tendem a se distribuir de acordo com a curva gaussiana, estando esta distribuição na origem das normas para a interpretação dos desempenhos individuais.

1.2. Relação entre o computador e a testagem

Os computadores têm influenciado profundamente o desenvolvimento da testagem. Se nas primeiras fases dessa relação entre testagem e computador este foi mantido nos bastidores auxiliando o pesquisador na formatação, processamento e entendimento dos dados coletados nas pesquisas, nas fases seguintes tivemos o computador elevado ao plano principal, primeiro como aplicador de testes previamente concebidos, substituindo os cadernos de testes impressos, depois como elaborador de testes sorteando os itens que serão apresentados.

Mas não ficou por aí. Nos últimos anos assistimos a uma grande popularização dos computadores devido a sua miniaturização e baixo preço associado ao aumento extraordinário do seu poder de processamento e plasticidade. Com computadores robustos e onipresentes, a portabilidade das provas e de suas bases de dados experimentou um crescimento sem precedentes. O processo de transcrição dos dados coletados em papel para os programas de análise estatística podem, agora, ser eliminados, relegando os erros de digitação para o passado. Novas estratégias e possibilidades de trabalho estão sendo criadas visando modernizar, refinar, simplificar e dar suporte ao processo de testagem, como os testes dinâmicos e os itens multimídia.

Como veremos adiante, na testagem adaptativa (também chamada de dinâmica ou sob medida) o computador não só apresenta os itens aos participantes, como também seleciona o próximo item a ser apresentado com base nas respostas fornecidas nos itens e testes anteriores (Hogan, 2006), produzindo testes únicos, otimizados para níveis diferenciados de habilidades e, portanto, mais adequados para expressar nossa diversidade de processos cognitivos. Os testes adaptativos estão revolucionando o campo da testagem, representando uma das áreas que mais cresce atualmente dentro da psicometria, ainda segundo Hogan (2006).

A utilização do computador para aplicação de testes também vem propiciando a utilização de uma nova categoria de itens: a dos itens multimídia e interativos. Esses tipos de itens podem englobar estímulos visuais, como figuras, vídeos e animações, estímulos sonoros e até mesmo táteis empregando joysticks com vibração. Santos (2007) relata que a utilização de computadores com crianças normalmente resulta em aumento da motivação, diminuição do comportamento inadequado à atividade ou situação em questão, maior atenção e aumento do rendimento quando comparado com os métodos tradicionais.

Para os psicólogos a informatização das provas pode representar redução de despesas com materiais, aumento sem precedentes da portabilidade tanto das provas quanto das bases de dados, eliminação do processo de transcrever os dados coletados em papel para os programas de análise estatística (deixando os erros de digitação no passado) e agilidade na coleta de dados e na geração de relatórios para interpretação das informações.

Numa lógica de testes psicológicos que evoluem acompanhando e se adaptando à inovação tecnológica, diríamos que o teste informatizado traduz uma atenção da psicologia à rentabilização das novas tecnologias, podendo isso representar um fator de validade dos testes, pois, em teoria, os tornam mais atraentes às pessoas que os realizam. De fato, nos dias de hoje, as crianças muitas vezes parecem mais interessadas em manipular equipamentos do que escrever com lápis e papel, incentivadas que são desde o nascimento pelo contato precoce com as telas sensíveis ao toque dos tablets e celulares onde podem jogar, colorir, se comunicar com outras pessoas, ver fotos e assistir desenhos ou filmes.

Millsap (2000) faz um extenso levantamento de trabalhos que buscaram identificar diferenças nos desempenhos dos participantes em testes informatizados versus testes com lápis e papel, mas deduziu que a literatura existente é inconclusiva, ambígua e, por vezes, contraditória, haja vista a grande diversidade de situações, métodos e objetivos das pesquisas relatadas. Além disso, transpor para o computador escalas concebidas para serem utilizadas

com cadernos de provas em papel, de maneira tradicional, como é o caso da ECCOs 4/10, requer grandes cuidados e pode não ser uma tarefa tão simples ou direta, sendo muito importante visualizar os limites impostos pela própria estrutura em que a escala foi concebida e avaliar os ganhos que a informatização pode, de fato, permitir, pois não haverá garantias de que a informatização das provas proporcionará todos os benefícios desejados. Além disso, em escalas que utilizam várias provas para avaliar processos complementares, como também é o caso da ECCOs, objeto central deste trabalho, nem sempre se conseguirá atingir resultados similares de satisfação para todas as provas. Nesses casos, à primeira vista, o mais sensato e seguro é acreditar que teremos uma diversidade de situações, tanto no tocante da migração do papel para o meio eletrônico em si, como das funcionalidades e mesmo das vantagens obtidas por cada uma das provas na versão eletrônica.

1.3. Teoria da resposta ao item (TRI)

A teoria da resposta ao item (TRI), também denominada teoria do traço latente, teoria da curva característica do item ou, ainda, psicometria moderna, é uma abordagem para relacionar os dados observados com a localização das pessoas no continuum da variável latente ou, em outras palavras, um sistema de modelos que define uma forma de estabelecer a correspondência entre variáveis latentes e suas manifestações (Ayala, 2009). Basicamente, trata-se de um paradigma baseado em modelos matemáticos destinado à concepção, análise e pontuação de testes, questionários e instrumentos similares que se destinem à mensuração de habilidades, atitudes e outras variáveis.

As bases da TRI não são novas, foram erguidas nos Estados Unidos na década de 1950, notadamente a partir dos trabalhos de Paul Lazarsfeld (1959), Frederic Lord (1952) e do dinamarquês George Rasch (1960), mas seu uso somente se popularizou a partir dos anos 1980 em função do avanço dos microcomputadores e da disponibilidade de softwares apropriados, ambos necessários para dar conta dos seus complexos algoritmos matemáticos. Também não é recente o conceito utilizado pela TRI de que os itens de um teste podem ser associados a uma localização no continuum do traço latente, podendo ser visto em Thurstone (1925, 1928), Andrich (1978), Lumsden (1978) e Yen (1986).

O nome *teoria da resposta ao item* se deve ao foco da teoria sobre o item em oposição ao foco no nível do teste existente na teoria clássica dos testes. Assim, a TRI modela as respostas de cada participante em uma dada habilidade para cada item do teste. O termo item, por sua vez, é genérico, cobrindo todos os tipos de itens informativos. Os itens podem se apresentar como tarefas (questões) de múltipla escolha, mas também podem ser sentenças em questionários que admitem respostas onde é indicado o nível de concordância do participante

quanto a essas sentenças, como na escala Likert. Em outros formatos de itens se é possível indicar a existência ou ausência de algum comportamento, sinal ou sintoma.

Nestas primeiras décadas do século XXI, a maioria dos grandes exames voltados à avaliação educacional (testes psicométricos de realização acadêmica) usam a TRI. Segundo Nota Técnica do Ministério da Educação (2012), o uso da TRI em avaliações educacionais teve início no Brasil com o Sistema de Avaliação da Educação Básica (SAEB) em 1995 e, posteriormente, foi implementado também no Exame para Certificação de Competências de Jovens e Adultos (ENCCEJA), na Prova Brasil e no Exame Nacional do Ensino Médio (ENEM). De acordo com a nota do Ministério da Educação, no âmbito internacional, a TRI vem sendo utilizada largamente por diversos países, tais como Estados Unidos, França, Holanda, Coreia do Sul e China, sem falar nos mais de sessenta países participantes do Programa Internacional de Avaliação de Estudantes (PISA).

Ainda conforme a Nota Técnica do Ministério da Educação, um dos grandes e mais consolidado exemplo de avaliação utilizando a TRI é o Test of English as a Foreign Language (TOEFL), que avalia a proficiência na língua inglesa. Surgido em 1964, este exame já avaliou mais de 25 milhões de alunos e tem sido aplicado por mais de 4.500 centros em 165 países. Importante citar que no TOEFL os alunos marcam o horário em um dos centros credenciados e podem realizá-lo várias vezes ao ano. A prova tem seções adaptativas, realizadas em computador, e, apesar de que a cada aplicação o candidato recebe uma prova distinta, todos os resultados são comparáveis e considerados isonômicos.

Outro exame educacional bastante importante que também se utiliza da TRI é o Scholastic Assessment Test (SAT) que serve de critério para admissão nas universidades norte-americanas, semelhante ao nosso ENEM. Tal como o TOEFL, apesar de as avaliações serem realizadas sete vezes ao ano, portanto, em momentos distintos e com provas diferentes,

a existência de uma escala padrão possibilita a comparabilidade entre o desempenho dos participantes. Também são importantes e bastante conhecidos os testes Progress in International Reading Literacy Study (PIRLS), que avalia os estudantes do quarto ano das escolas norte-americanas em leitura, e o Trends in International Mathematics and Science Study (TIMSS) que avalia estudantes do quarto, oitavo e décimo-segundo anos de mais de 60 países em ciências e matemática.

A hoje chamada teoria clássica dos testes (TCT), vinculada à psicometria tradicional, predominava na construção de testes até a popularização da TRI nos anos 1980. A partir daí, a TRI vem se consolidando cada vez mais e substituindo grande parte da teoria clássica da psicometria. Segundo Pasquali e colaboradores (2010), "a TCT surgiu dentro da concepção monista materialista que imperava nas ciências em geral desde o empirismo inglês do século XVII, enquanto a TRI faz suposição de uma concepção dualista interacionista do ser humano" (p. 105).

Para Hambleton, Swaminathan e Rogers (1991), o enorme impacto da TRI se deveu ao fato dela superar certas limitações teóricas graves que a psicometria tradicional contém, em especial a dependência dos participantes da amostra (group-dependent ou sample-dependent), a incapacidade de produzir os mesmos escores com os mesmos participantes em testes diferentes medindo a mesma aptidão (test-dependent), as dificuldades para garantir a fidedignidade do testes, além de sua própria orientação para o teste total, inviabilizando a análise por item. Em contrapartida, a TRI avalia as características do item, não depende da amostra de participantes, fornece escores semelhantes independentemente do teste utilizado, orienta a análise de cada item individualmente, permite acrescentar ou retirar itens de um teste facilmente, não exige rigor no paralelismo da fidedignidade e oferece uma medida de precisão para cada nível de aptidão.

Basicamente, segundo da Silva et al. (2012), a TRI adota dois postulados básicos:

- (i) O desempenho do participante em um item do teste pode ser explicado a partir de um conjunto de traços latentes, onde o desempenho é o efeito e os traços latentes são a causa;
- (ii) A relação entre o desempenho na tarefa e o conjunto dos traços latentes pode ser descrita por uma equação monotônica crescente, chamada de curva característica do item (CCI), conforme exemplificado na Figura 1, onde podemos observar que pessoas com aptidão maior terão maior probabilidade de responder corretamente ao item e vice-versa (Θ - teta - é a aptidão e $P_i(\Theta)$ a probabilidade de uma pessoa responder corretamente ao item).

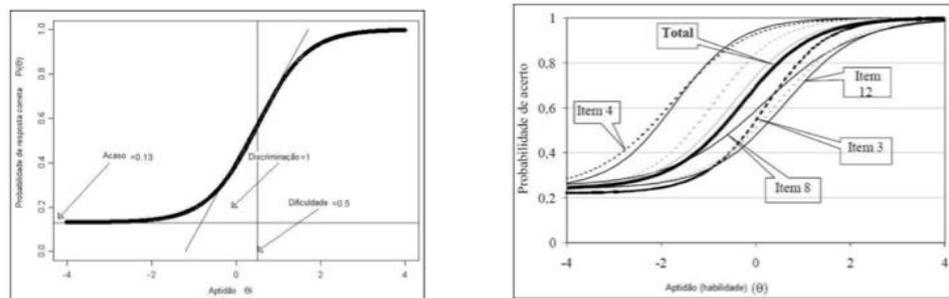


Figura 1 - Curva característica do item (CCI)

Dessa forma, ao analisarmos as respostas dadas por uma pessoa a uma série de itens podemos inferir sobre seu traço latente, hipotetizando relações dessa pessoa com seu nível de traço latente através de uma equação matemática que descreve a forma de função que estas relações assumem.

Nos casos em que algum dos modelos da TRI adira aos dados observados, ela oferece vantagens importantes sobre a TCT. Por exemplo, com a TRI as estimativas de localização das pessoas no *continuum* da variável latente são invariáveis em relação ao instrumento, sendo a precisão dessas estimativas conhecida no nível individual e não apenas no nível de grupo. Além disso, ao contrário da TCT, na TRI somos capazes de fazer declarações

preditivas sobre o desempenho dos respondentes bem como examinar a sustentabilidade ou não do modelo em relação aos dados, uma vez que ela reúne modelos estatísticos usados para fazer previsões, estimativas ou inferências sobre as habilidades (ou competências) medidas em um teste.

A TRI vem se tornando o modelo psicométrico básico no campo dos testes, sendo também a técnica predominante na construção dos testes adaptativos (Pasquali, 2003). Enquanto a TCT se preocupa em explicar o resultado final total de um teste, isto é, a soma das respostas dadas a uma série de itens expressa no chamado score total, a TRI se interessa especificamente por cada um dos itens do teste e quer saber qual é a probabilidade e quais são os fatores que afetam esta probabilidade de cada item individualmente ser acertado ou errado. Assim, a TCT tem interesse em produzir testes de qualidade, enquanto a TRI procura produzir itens de qualidade.

Segundo Torrecilla (2006), os fatores propiciados pela utilização da TRI que realmente garantem o sucesso do teste são: a) a elaboração de provas constituídas por itens independentes dos grupos de aplicação; b) a garantia de que a habilidade do aluno seja traduzida por um score independentemente da dificuldade do teste; c) a construção de itens relacionados perfeitamente à habilidade que se pretende medir; e d) a garantia de fidedignidade e comparabilidade dos resultados numa perspectiva espacial e temporal, além de não se fazer necessário o uso de provas estritamente paralelas.

Ayala (2009) ressalta que o termo “teoria” empregado na TRI deve ser entendido no sentido de que a TRI estabelece um paradigma que tenta explicar todos os fatos com que ela pode ser confrontada. De fato, a TRI não representa uma teoria no sentido tradicional, porque não explica a razão que levou a pessoa a fornecer uma resposta particular para um item ou como essa pessoa decidiu o que responder. Ao invés disso, a TRI é uma teoria de

estimação estatística que usa as caracterizações latentes de indivíduos e itens para prever as respostas observadas. Os processos cognitivos usados por um indivíduo para responder a um item não são modelados na maioria dos modelos de TRI.

Há muitos possíveis domínios de construtos para os quais a TRI pode ser aplicada. Estes envolvem construtos psicológicos, tais como motivação, ansiedade social, o desenvolvimento cognitivo, as preferências dos consumidores, a proficiência, e assim por diante. Seja qual for o construto de interesse, supomos que ele se manifesta através da resposta de um indivíduo a uma série de itens.

Temos que a riqueza na avaliação psicológica dentro do enfoque da TRI consiste em construir bancos de itens válidos para avaliar os traços latentes, itens com os quais podemos construir tantos testes válidos quanto quisermos ou o número de itens permitir. Assim, como escreveu Pasquali (1997), a tarefa do psicometrista já não será mais de validar e normatizar testes e sim de parametrizar tarefas ou itens. Com isso se quer dizer que a tarefa consiste, agora, em redigir a carteira de identidade de cada item contendo os seus parâmetros distintivos, tais como o seu coeficiente de validade (a carga no traço latente), seu índice de discriminação, nível de dificuldade, seu índice de disfunção cultural (DIF), e outros.

1.3.1. Unidimensionalidade e independência local

Duas suposições essenciais para a aplicação de modelos baseados na TRI são a unidimensionalidade e a independência local.

Sobre a independência local, o postulado afirma que se forem mantidas constantes as aptidões que afetam o teste, menos o teta (habilidade) dominante, as respostas das pessoas a quaisquer dois itens são estatisticamente independentes. Ou seja, o desempenho da pessoa em um item não deve afetar seu desempenho em outro item, sendo cada item respondido

exclusivamente em função do tamanho do seu teta dominante. Esta suposição é importante e útil porque significa que para examinandos com uma aptidão dada a probabilidade de resposta a um conjunto de itens é igual aos produtos das probabilidades das respostas do examinando a cada item individualmente.

Quanto à unidimensionalidade, não obstante a existência do desenvolvimento de modelos para testes com respostas multideterminadas, a grande maioria dos modelos da TRI postula que deve existir apenas uma aptidão (representada pela letra grega teta) responsável pela realização de um conjunto de tarefas (os itens de um teste) para que a medição possa ser realizada – a bem da verdade esse é um postulado intrínseco a qualquer tipo de medida. Por outro lado, como afirmam Pasquali e Primi (2003), qualquer desempenho humano é sempre multideterminado ou multimotivado, dado que na execução de qualquer tarefa participam mais de um traço latente, de acordo com o modelo geral do traço latente (que nos remete aos anos 1930). Dessa forma, para satisfazer o postulado da unidimensionalidade é suficiente admitir a existência de uma aptidão dominante responsável pelo desempenho num conjunto de itens de um teste.

De fato, a questão da unidimensionalidade pode ser muito complicada, porque ela depende do propósito do instrumento. Por exemplo, um teste de aritmética (adição, subtração, multiplicação, divisão) é unidimensional na perspectiva do administrador escolar decidir se a criança deve avançar para a próxima série escolar, mas o mesmo teste é multidimensional na perspectiva do psicopedagogo diagnosticando dificuldades na aprendizagem. Na prática, a unidimensionalidade é mais conceitual do que factual, mais qualitativa do que quantitativa, uma ideia ou intuição mais que a experiência. Voltaremos a abordar a unidimensionalidade no tópico do modelo de Rasch.

1.3.2. Modelos da TRI

Para que possamos extrair as características de itens e pessoas e prever as chances que as pessoas terão de responder acertadamente determinado item de um teste, como proposto pela TRI, precisamos vincular um modelo estatístico que possa suportar essas capacidades e representar nossos dados. Como não temos acesso direto a sentimentos ou processos psicológicos, coletamos dados que representam o processo de interesse e, então, usamos esses dados para construir modelos estatísticos, reduzindo o processo psicológico a um modelo estatístico (Field, 2009). Claro que os modelos podem diferir da realidade de diversas formas, pois sempre carregam algum erro consigo, mas o que devemos tentar construir é um modelo que melhor se ajuste à situação de interesse, com o menor erro possível, baseado nos dados disponíveis.

Uma importante forma em que os modelos da TRI diferem é quanto ao número de parâmetros em que eles estão interessados. Na prática, existem três modelos matemáticos predominantes para expressar a relação de probabilidade de sucesso em um item e a aptidão medida por um teste: são os modelos logísticos de um (ML1), dois (ML2) ou três (ML3) parâmetros. Nos modelos ML1, apenas a dificuldade do item é considerada; nos modelos ML2, considera-se a dificuldade e a capacidade de discriminação dos itens; já os modelos ML3 consideram a dificuldade, a discriminação e a sensibilidade a falsos positivos (normalmente respostas dadas ao acaso). Quando obedecidos os pressupostos exigidos pelo modelo, os parâmetros dos itens (dificuldade, discriminação e acerto ao acaso) são invariantes, ou seja, são consideravelmente constantes independentemente da amostra na qual são aplicadas. Logicamente, o modelo mais simples da TRI é o que caracteriza cada item em termos de um único parâmetro. Este parâmetro é a localização do item no *continuum* latente que representa o construto.

Um modelo de um parâmetro muito popular atualmente é o modelo de Rasch (1960). Esse foi o modelo utilizado nas pesquisas aqui relatadas e será detalhado a seguir.

O modelo de Rasch

Apesar do modelo de Rasch ser considerado um modelo de um parâmetro (1PL), muitos preferem vê-lo como uma abordagem completamente diferente de conceber a relação entre dados e teoria. Assim como outras abordagens de modelagem estatística, a TRI enfatiza a primazia do ajuste de um modelo aos dados observados, enquanto o modelo de Rasch enfatiza a primazia dos requisitos para a medição fundamental, sendo a adequação ao modelo um requisito importante, mas secundário, para a produção da medida. Em termos operacionais, isso significa que as abordagens da TRI incluem parâmetros adicionais para refletir os padrões observados nos dados, enquanto que na abordagem Rasch as alegações relativas à presença de um traço latente só podem ser consideradas válidas quando ambos (a) os dados de ajustarem ao modelo de Rasch, e (b) itens de teste e participantes obedecerem ao modelo. Assim, a abordagem Rasch pode ser vista como uma abordagem de confirmação, por oposição às abordagens experimentais que tentam modelar os dados observados.

Neste ponto é importante que nos detenhamos atentamente aos conceitos e estatísticas propiciadas pelo modelo de Rasch, em especial na forma como foi implementado no Winsteps, o software para produção de análises estatísticas com Rasch que também utilizamos nesse trabalho, visto que a maior parte da análise desta pesquisa depende dessa compreensão.

Segundo Bond e Fox (2007), assim como em outras teorias, para um modelo ser útil à investigação de aspectos da condição humana representada nas teorias do desenvolvimento, é preciso incorporar as seguintes propriedades:

- a. Ser sensível à aquisição ordenada de conhecimentos ou habilidades sob investigação (isto é, ele deve revelar a ordem de desenvolvimento ou aquisição).
- b. Conseguir estimar as distâncias entre as habilidades das pessoas (ou seja, deve nos dizer quanto uma pessoa é mais desenvolvida, mais capaz, ou mais reabilitada do que a outra pessoa).
- c. Permitir determinar se o padrão geral de desenvolvimento entre os itens e as pessoas é suficiente para explicar o padrão de desenvolvimento apresentado por cada item e cada pessoa.

As análises baseadas no modelo de Rasch para medições são particularmente orientadas para investigações em uma grande faixa das ciências humanas exatamente por suportar essas características.

De forma resumida, a análise de Rasch é um método para a obtenção de medidas objetivas, fundamentais e aditivas (qualificada por erros padrão e estatísticas para ajuste do controle de qualidade) a partir de observações estocásticas de respostas ordenadas. O matemático dinamarquês Georg Rasch formulou sua abordagem em 1953 para analisar as respostas a uma série de testes de leitura.

Rasch desenvolveu um modelo matemático para construção de medidas baseado na relação probabilística entre a dificuldade de qualquer item e a habilidade de qualquer pessoa, sendo que a diferença entre essas duas medidas governa a probabilidade de qualquer pessoa ter sucesso num item em particular. O membro mais simples da família de modelos Rasch é um modelo de probabilidade binomial para uma resposta dicotômica (Rasch, 1960; 1980), ou seja, respostas tipo certo ou errado. Rasch formulou seus modelos probabilísticos para alguns testes de inteligência e de realização usando um princípio deliciosamente simples: uma pessoa com habilidade maior do que outra deverá ter uma probabilidade maior de resolver qualquer item do mesmo tipo. Do mesmo modo, se temos um item mais difícil do que outro

significa que para qualquer pessoa a probabilidade de resolver o item mais fácil será maior do que a probabilidade de resolver o item mais difícil (Rasch, 1960).

Apesar de simples, até mesmo óbvio, esse entendimento tem um resultado surpreendente para a análise de itens. Com medidas produzidas por esse modelo é possível identificar as habilidades das pessoas na amostra de dados para calibragem do teste e a partir das particularidades dessas habilidades estimar a facilidade do item. Os escores das pessoas obtidos no teste podem ser usados para remover a influência de suas habilidades sobre a análise dos itens. Como consequência, a calibragem do teste é feita independentemente das pessoas.

Para Bond e Fox (2007), a característica central do modelo de Rasch é uma tabela de probabilidades esperadas desenhada para responder à seguinte questão central: qual é a probabilidade de uma pessoa com habilidade X responder corretamente a um item com dificuldade Y ? A resposta é que a probabilidade de sucesso depende da diferença entre a habilidade da pessoa e a dificuldade do item.

Considerando-se uma mesma população sob as condições do modelo de Rasch, as medidas obtidas a partir dos escores brutos total das pessoas e dos itens são estatisticamente equivalentes para pessoas e itens, independentemente de quais pessoas ou itens sejam analisados.

O modelo de Rasch incorpora um método para ordenar pessoas de acordo com suas habilidades e ordenar itens de acordo com seus níveis de dificuldade. Os programas baseados na análise de Rasch, tal como o Winsteps, executam uma transformação logarítmica sobre os dados dos itens e pessoas para converter os dados ordinais em dados intervalares. Essas transformações representam a estimativa de habilidade das pessoas e a dificuldade dos itens

que foi detectada no conjunto de dados, ou seja, itens e pessoas posicionadas e comparáveis ao longo do *continuum*.

Segundo Bond e Fox (2007), as estimativas das habilidades das pessoas e das dificuldades dos itens são apresentadas ao longo de uma escala logit (log odds unit), uma escala intervalar na qual os intervalos entre o posicionamento de pessoas e itens têm um valor ou significado consistente. Normalmente, o modelo de Rasch fixa em 50% a probabilidade de sucesso para qualquer pessoa em relação a um item localizado no mesmo ponto da escala logit de itens e pessoas, ou seja, quando a habilidade da pessoa (em logit) for igual à dificuldade do item (também em logit) sua chance de sucesso no item será de 50%.

1.3.3. Validade, fidedignidade

A análise dos itens é um método de inspecionar itens de um teste tanto qualitativamente quanto estatisticamente, visando garantir que todos os itens tenham um critério mínimo de qualidade. O objetivo de ambas as análises é o mesmo: identificar eventuais itens problemáticos no teste, pois eles afetarão a fidedignidade e a validade do teste como um todo.

Segundo Varma (2010), itens podem ser problemáticos por diversas razões: redação ruim, inadequação ou desatualização de gráficos, desenhos, diagramas, falta de clareza nas opções de resposta, por conter distratores que levem o aluno a tentar adivinhar a resposta correta, por apresentar conteúdos (construtos) diferentes dos demais itens (multidimensionalidade) ou por conter vieses a favor ou contra gêneros, grupos étnicos e outros.

Fidedignidade (reliability) e validade são assuntos fundamentais para as medidas em psicologia. Enquanto a fidedignidade se refere a quanto uma variável influencia um conjunto

de itens, a validade se preocupa em verificar se a variável é a causa subjacente da covariação do item. Na medida em que uma escala é fidedigna, a variação nos escores na escala pode ser atribuída ao escore verdadeiro de algum fenômeno que exerce uma influência causal sobre todos os itens. Entretanto, determinar que uma escala é fidedigna não garante que a variável latente compartilhada por todos os itens seja, de fato, a variável de interesse do desenvolvedor da escala. A adequação da escala como a medida de uma variável específica (por exemplo, estresse psicológico) é um assunto da validade.

Quando falamos de fidedignidade de um teste queremos saber se ele é capaz de produzir os mesmos resultados se administrado mais de uma vez para um mesmo grupo, ou seja, refere-se à estabilidade da medida. Outro indicador de fidedignidade é que o teste deve se comportar da mesma maneira para diferentes grupos, significando que os itens devem apresentar o mesmo índice quando ordenado pelo seu “p-value” - um indicador da dificuldade do item que detalharemos adiante. Já quando falamos de validade de um teste, queremos saber se o teste está medindo, realmente, o que ele se propõe a medir.

Segundo DeVellis (2003), as escalas serão fidedignas na medida em que elas consistirem de itens confiáveis que compartilhem uma variável latente comum. Existem vários métodos (coeficiente alfa, matriz de covariância, teste-reteste, etc.) para computar fidedignidade e eles podem ter diferentes utilidades em situações particulares.

A consistência interna de um teste normalmente é fornecida pelo coeficiente alfa de Cronbach (Cronbach, 1951), correspondendo, fortemente, à clássica definição de fidedignidade como sendo a proporção da variância em uma escala que pode ser atribuída ao escore verdadeiro da variável latente. O software Winsteps implementa uma versão análoga ao alfa de Cronbach para itens dicotômicos chamada Kuder-Richardson fórmula 20 (KR-20). Um coeficiente acima de 0.9 para o KR-20, normalmente, indica um teste homogêneo.

Através do software SPSS também é possível extrair um outro interessante indicador da fidedignidade do teste baseado no alfa de Cronbach, que é o “alfa de Cronbach se o item for deletado”. A sutileza desse indicador é que podemos avaliar o que aconteceria com a fidedignidade do teste se um determinado item fosse excluído. Ao comparar o alfa de Cronbach do teste completo com o alfa de Cronbach sem um determinado item, podemos constatar se a eliminação desse item aumentará ou diminuirá a fidedignidade do teste completo, ou seja, se a informação oriunda do item contribui ou não com o teste.

Segundo Varma (2010), duas outras estatísticas podem ser usadas para determinar se um item de teste é provável de ser válido e fidedigno: correlação point-biserial e p-value.

A correlação point-biserial é a correlação entre os escores certos e errados dos participantes em um dado item e o escore total desses participantes na soma de seus escores em todos os demais itens. É um tipo especial de correlação entre uma variável dicotômica (uma resposta certa ou errada: 1 ou 0) e uma variável contínua (o escore total que vai de zero até o número de itens). Como em todas as correlações, a point-biserial varia entre -1.0 até +1.0. Valores grandes e positivos indicam participantes com altos escores no teste total que também acertaram o item (como esperado) e participantes com baixos escores no teste total que erraram o item (como esperado). Valores baixos para point-biserial implicam que os participantes que acertaram o item tendem a ter um escore baixo no teste total (indica anormalidade) e que os participantes que erraram o item tendem a ter um bom escore no teste total (também anormal).

Itens com valores muito baixos para a correlação point-biserial devem ser investigados. Alguma coisa na redação, apresentação ou conteúdo desses itens pode explicar a baixa correlação. Entretanto, mesmo se nada parecer errado com esses itens, recomenda-se que eles sejam retirados do teste. Varma (2010) diz que bons itens apresentarão correlação

point-biserial acima de 0,25 e recomenda excluir os itens com índice abaixo de 0,15. Linacre (1998) sugere investigar e excluir, quando for o caso, apenas os casos de correlações negativas.

Já o p-value de um item nos dá a proporção de participantes que acertaram o item. Pode ser convertido para percentagem quando multiplicado por cem, representando a percentagem de participantes que acertaram o item. Varia de 0 até 1 e indica a dificuldade (ou facilidade, mais precisamente) do item. O valor de p-value será alto para itens fáceis e baixo para itens difíceis. De uma forma geral, os testes serão mais confiáveis quando os p-values derem cobertura a toda a faixa de possibilidades de 0.0 até 1.0 e tiver uma concentração maior de itens com dificuldade mediana, ou seja, com p-value em torno de 0,5.

Resumindo, a correlação point-biserial deve ser usada para avaliar a qualidade do item e p-value deve ser usado para avaliar a dificuldade do item.

Segundo o Winsteps Help Manual for Rasch, no Winsteps o índice de fidedignidade do participante é equivalente ao índice de fidedignidade do teste nos modelos tradicionais. Valores baixos indicam medidas de pessoas com baixa habilidade ou um número muito pequeno de itens. Esse índice pode ser aumentado incluindo no teste pessoas com mais habilidades extremas (altas e baixas) ou prolongando o teste. Melhorar a segmentação do teste também pode ajudar a elevar a fidedignidade.

Fidedignidade também significa reprodutibilidade da localização relativa da medida, mas não informa sobre a qualidade dos dados. Assim, alta fidedignidade significa que existe uma grande probabilidade de que as pessoas ou itens estimados com habilidade elevada realmente tenham medidas maiores do que as pessoas ou itens estimados com valores baixos. Índices altos de fidedignidade serão obtidos através de uma amostra ampla e/ou de um erro baixo de medição. Então, alta fidedignidade para a medida das pessoas requer uma amostra

de pessoas com uma ampla gama de habilidades e/ou de um instrumento com muitos itens. Alta fidedignidade para o item requer um teste com itens que deem cobertura a uma grande faixa de dificuldades e/ou uma grande amostra de pessoas. Normalmente, baixa fidedignidade do item indica que a amostra é muito pequena para permitir estabelecer uma hierarquia reprodutível de dificuldade do item.

1.3.4. Separação

Ainda no conceito da fidedignidade, o modelo de Rasch também provê índices que ajudam o investigador a determinar se existem itens e pessoas suficientes ao longo do *continuum* mensurado para garantir a adequada localização desses elementos nesse *continuum*.

A fidedignidade para a medição da habilidade das pessoas exige não só capacidade de produzir estimativas bem orientadas por um conjunto adequado de itens, mas também deve ser capaz de diferenciar capacidades em toda a amostra de modo que das medidas possa emergir a hierarquia de habilidades (ou desenvolvimento) das pessoas sobre esse construto, ou seja, permita identificar a separação dos sujeitos (Fox & Jones, 1998). Portanto, alta fidedignidade das pessoas significa que temos uma escala em que algumas pessoas terão maior pontuação e outras pessoas pontuações menores e que nós podemos esperar coerência dessas inferências. Analogamente, a fidedignidade para os itens indica se os itens manterão as mesmas características quando submetidos a outras pessoas.

A separação do item é a distância em logits entre itens com dificuldades variadas. É representado por dois cálculos, o índice de confiabilidade (reliability index) e o índice de separação do item (item separation index). Como vimos, o índice de confiabilidade é equivalente ao alfa de Cronbach ou o KR-20, representando a estimativa de replicabilidade de posicionamento dos itens dentro da hierarquia de dificuldade de pessoas com habilidades

diferentes. O índice de separação do item é expresso em unidades de erro padrão e calculado pela divisão do desvio padrão ajustado pela média do erro de medição. A confiabilidade da separação do item é determinada pelo grau em que as calibrações dos itens estão suficientemente separadas para definir diferentes níveis de capacidade. A separação das habilidades das pessoas segue a mesma lógica.

A separação implementada em Rasch é representada pelo índice de separação (G), que é a razão entre a estimativa do desvio padrão verdadeiro da amostra (pessoa ou item) e a raiz quadrada da média do erro de medição ($rmsa$) da amostra. Ele quantifica "confiança" de uma forma simples e direta e tem uma interpretação clara: a proporção entre a variância "verdadeira" e o erro da variância.

Segundo Bond e Fox (2007), no modelo de Rasch a separação é estimada para as medidas de pessoas e itens. O índice de separação das pessoas (person reliability index) e o índice de separação dos itens (item reliability index) são estimativas de como diferenciar pessoas e itens no *continuum*. Ou seja, pode-se estimar a replicabilidade do posicionamento da pessoa sobre outros itens medindo o mesmo construto. Essa estimativa é baseada no mesmo conceito do alfa de Cronbach, isto é, a fração da variância da resposta observada que é reproduzível.

1.3.5. Estatísticas de aderência ao modelo

As estatísticas de aderência (fit) servem para estimar a magnitude do desajuste dos dados das pessoas e dos itens ao modelo. Segundo Linacre (2002), no contexto de Rasch as estatísticas que indicam precisão e permitem previsibilidade de aderência dos dados ao modelo são outfit, infit, média-quadrada e as estatísticas padronizadas, conforme descritas a seguir.

Outfit (outlier-sensitive fit) é a soma dos quadrados dos residuais padronizados baseada na estatística convencional do qui-quadrado. É uma estatística para identificação de outliers, mais sensível a respostas inesperadas nos extremos, ou seja, quando a pessoa erra um item considerado muito fácil para sua habilidade ou acerta um item considerado muito difícil para sua habilidade.

Infit (inlier-sensitive fit), também baseada no qui-quadrado, mas tendo cada observação ponderada pela sua informação estatística (variância do modelo). É utilizada para ponderação da informação sobre os itens que não estão sendo acertados por pessoas com habilidade próxima à habilidade requerida pelo item.

Média-quadrada (mean-square - MNSQ) é uma estatística de aderência que mostra o tamanho da aleatoriedade dos dados, ou seja, a quantidade de distorção do sistema de medida. Corresponde ao qui-quadrado dividido pelo seu grau de liberdade, sendo esperado um valor próximo de 1.0 (sempre será positivo). Valores menores que 1.0 indicam excesso de previsibilidade (redundância, dados superdimensionados para o modelo). Valores maiores que 1.0 indicam imprevisibilidade (interferências não modeladas, dados subdimensionados para o modelo). A expectativa do MNSQ tanto para outfit quanto para infit em sistemas modelados em Rasch é que seus valores estejam orientados para 1.0. Normalmente, valores entre -2 e +2 são admitidos (Bond & Fox, 2007, p.57 e 62), esperando-se que se aproximem de 1.0 quanto maior for a amostra.

Estatísticas padronizadas (standardized - ZSTD) são estatísticas tipo t-test da hipótese “Os meus dados se ajustam perfeitamente ao modelo?”. São reportadas como z-escores, unidades de desvio da curva normal. Mostra a improbabilidade dos dados, isto é, sua significância. Terá valor zero quando os dados aderirem ao modelo. Valores menores que

zero também indicam previsibilidade e maiores que zero lacunas na previsibilidade. Segundo o Winsteps Help for Rasch Analysis, se o MNSQ for aceitável, ZSTD pode ser ignorado.

A análise de aderência avalia o quanto nossos dados cooperam com a construção da medição. Ela nos dá uma ferramenta para monitorar respostas de pessoas e itens. Nós podemos avaliar qualquer conjunto de itens ou pessoas para determinar onde o desajuste ocorre. A análise de aderência provê uma técnica de controle de qualidade requerida para supervisionar e validar os itens de um teste e as respostas das pessoas. Quando os dados estão dentro das nossas expectativas de aderência, nós temos o controle requerido para confiança de calibragem dos itens e medição das pessoas. Quando desajustes são identificados nós podemos localizar essas ocasiões e iniciar um estudo mais frequente dos itens e pessoas envolvidas.

1.3.6. Unidimensionalidade em Rasch

A unidimensionalidade, conceito de fundamental importância para a TRI, particularmente para o modelo de Rasch, é frágil e pode ser violada de muitas maneiras, de acordo com o Winsteps Help for Rasch Analysis. Se considerarmos todos os testes estatísticos conhecidos para verificar violações dessa natureza, qualquer base de dados irá falhar com algum deles, indubitavelmente. Conseqüentemente, a questão não é se os dados são perfeitamente unidimensionais, porque eles não são. A pergunta que interessa e deve ser feita é se a multidimensionalidade presente nos dados é grande o suficiente para ameaçar a validade dos resultados da análise desses dados usando o modelo de Rasch.

Uma das ferramentas que vem sendo utilizada na análise de modelos Rasch em detrimento da tradicional análise de fatores é a análise dos principais componentes (APC). Trata-se de uma análise de residuais (respostas observadas menos suas expectativas). Wright (1996) usou a APC para investigar se não havia mais que um componente da variância

explicando a estrutura dos dados das respostas aos itens. O autor postula que se os dados forem unidimensionais, então os componentes nos residuais serão apenas ruídos, interferências. Essa mesma visão é defendida por Linacre (2008).

A ideia por trás do modelo de Rasch é que as informações dos dados, a coerência entre eles, possam ser explicadas pelas variáveis latentes. Então, as partes dos dados que permanecerem inexplicadas, os residuais, serão consideradas interferências aleatórias.

Segundo o Winsteps Help for Rasch Analysis, a principal diferença entre a análise fatorial tradicional (common-factor analysis) e a APC é que enquanto na primeira nós tentamos otimizar as *commonalities*, fazemos maximizações e rotações para dar uma estrutura mais forte possível ao fator e onde as cargas dos fatores serão interpretadas como correlações com os desejados fatores latentes, na APC de resíduos nós tentamos falsear a hipótese de que os resíduos são interferências aleatórias procurando encontrar o componente que explica a maior quantidade possível de variância nos resíduos. Esse é o primeiro contraste, ou seja, o primeiro componente da APC na matriz de correlação dos residuais. Se o eigenvalue do primeiro contraste é pequeno (usualmente até 2.0), então o primeiro contraste está no nível apenas de ruído e a hipótese de interferência aleatória não é falsa.

A ideia de conservar apenas os componentes que estão acima do chamado nível de ruído é uma prática comum em psicometria. O teste *scree* de Cattell (1966) e a regra de Kaiser (1960) são os procedimentos mais utilizados para determinar o número de componentes. Ambos são baseados na inspeção dos valores eigenvalues da matriz de correlação. A recomendação de Cattell é manter apenas os componentes acima do ponto de inflexão em uma relação de eigenvalues ordenados em ordem decrescente. Kaiser (1960) recomenda que apenas os eigenvalues iguais ou maiores do que um sejam mantidos.

Basicamente, o procedimento de análise da aderência dos dados ao modelo em Rasch envolve três etapas: o que foi observado, o que era esperado observar e as diferenças (os residuais) entre o esperado e o observado. Quanto menor esta diferença, maior o ajuste dos dados ao modelo. Linacre (1998) sugere três passos para investigar a dimensionalidade dos dados em Rasch e, conseqüentemente, definir a aderência dos dados ao modelo:

- (1) Identificar e examinar correlações point-biserials negativas
- (2) Corrigir os desajustes (misfit) de pessoas e itens usando os indicadores infit e outfit fornecidos pelo modelo de Rasch
- (3) Verificar a unidimensionalidade usando a análise de residuais de Rasch (ao invés da tradicional análise fatorial) para verificar a importância da parcela dos dados que não se ajustou ao modelo.

1.4. Teste adaptativo computadorizado (TAC)

Teste adaptativo é um termo amplo utilizado para descrever métodos que alteram dinamicamente as condições, o fluxo, o conteúdo ou os limites de um teste com base na fabricação ou seleção de dados do teste com ajuda da análise estatística. O termo é utilizado na psicologia, mas também poderá ser encontrado em áreas bem distantes, como na construção e testes de circuitos integrados.

Focando no âmbito da psicometria, o teste adaptativo computadorizado (TAC), muitas vezes denominado como teste adaptativo informatizado (TAI) ou ainda por CAT (do termo original em inglês computer-adaptive testing), é uma forma de teste baseado em computador que tem a capacidade de se adaptar dinamicamente ao nível de proficiência demonstrado pelo participante durante a realização do próprio teste. Conforme o website da International Association For Computerized Adaptive Testing (<http://www.iacat.org/what-is-cat>), o TAC é o redesenho de instrumentos de medições psicológicas e educacionais usando computadores, sendo indicado para testes de mensuração de habilidade ou realização, bem como para medidas de personalidade e variáveis atitudinais. No fundo, seu objetivo é selecionar, para cada examinando, um conjunto de itens para o teste a partir de um banco de itens pré-calibrados que simultaneamente seja mais efetivo e eficiente para posicionar a pessoa no continuum do traço latente.

1.4.1. Breve história dos testes adaptativos

O TAC representa o mais poderoso sucessor de uma série de aplicações bem sucedidas de avaliações individualizadas que teve início nos primórdios da mensuração psicológica durante o desenvolvimento do teste de QI de Alfred Binet (Binet & Simon, 1905) (mais tarde publicado como Stanford-Binet QI Test). Binet foi um expoente nessa área devido a seus testes de inteligência (Linacre, 2000). Como sua preocupação era com o

diagnóstico de cada indivíduo, ao invés do grupo, ele percebeu que não era justo aplicar o mesmo teste para todas as pessoas e adaptou o teste para as situações individuais classificando grupos de itens por níveis, de acordo com a idade cronológica. Assim, Binet começava a testar o participante com o que ele considerava ser o subconjunto de itens adequado para o nível de capacidade do examinando. Se o examinando acertava, Binet o apresentava a subconjuntos de itens sucessivamente mais difíceis até que ele falhasse algumas vezes. Por outro lado, se o candidato falhasse no subconjunto inicial de itens, Binet administrava subconjuntos de itens sucessivamente mais fáceis até que o candidato acertasse com frequência. Com essas informações, Binet gerava estimativas sobre o nível de habilidade dos participantes. A Figura 2 ilustra o procedimento utilizado no teste adaptativo de Binet.

Mental Age	Items	Adaptive Branching	Number Administered	Proportion Correct
10.5			—	—
Ceiling Level → 10	51- 52- 53- 54- 55- 56- 57- 58- 59- 60-		10	0.00
9.5	41+ 42+ 43+ 44- 45- 46+ 47- 48- 49- 50-		10	.40
Starting Level → 9	1+ 2+ 3- 4+ 5+ 6+ 7- 8- 9- 10+		10	.60
8.5	11+ 12- 13+ 14+ 15+ 16- 17+ 18+ 19+ 20+		10	.80
8	21+ 22+ 23+ 24+ 25+ 26+ 27+ 28- 29+ 30+		10	.90
Basal Level → 7.5	31+ 32+ 33+ 34+ 35+ 36+ 37+ 38+ 39+ 40+	10	1.00	
7			—	—
6.5			—	—
Total			60	.617

Figura 2 - Esquema de administração do teste adaptativo de Binet em 1905

É importante notar que o procedimento de aplicação do teste de Binet idealizado no início do século passado é um procedimento totalmente adaptativo, já trazendo consigo

muitos dos conceitos que são reproduzidos nos mais modernos testes adaptativos computadorizados produzidos atualmente. Vejamos:

1. Ele usa um banco pré-calibrado de itens por nível de idade cronológica, onde cerca de 50% das crianças da faixa etária respondeu ao item corretamente.
2. O teste é administrado individualmente por um psicólogo treinado e cada bloco de itens foi projetado para representar o nível de dificuldade que fosse mais adequado para a idade cronológica de cada examinando.
3. Tem uma opção de partida (item inicial) variável, ou seja, o administrador inicia o teste com o bloco de itens que melhor represente seu palpite sobre o provável nível de habilidade do examinando - tipicamente a idade cronológica do examinando era utilizada como parâmetro, mas poderia ser administrado um bloco de maior ou menor dificuldade, caso existissem informações adicionais que justificassem a mudança.
4. O teste usa um método bem definido de pontuação, calculado pelo administrador do teste imediatamente após a aplicação de cada bloco de itens.
5. Existe uma regra para determinar qual deve ser o próximo conjunto de itens que será administrado para o examinando baseada no seu desempenho nos conjuntos de itens anteriores, aumentando ou diminuindo a dificuldade do teste.
6. Uma regra pré-estabelecida terminava o teste quando os níveis superiores e inferiores da habilidade do examinando fossem determinados.
7. Por último, calculava-se a pontuação final do examinando no teste ponderando-se a sua idade cronológica sobre os resultados em cada conjunto de itens.

Segundo Linacre (2000), o procedimento de testes Flexilevel de Lord (1980) e suas variantes, tais como o procedimento Step, de Henning (1987), e o Testlets, de Lewis e Sheehan (1990), são refinamentos do método de Binet. Nesses testes, os itens são estratificados por nível de dificuldade e vários subconjuntos de itens são formados em cada nível. A execução do teste acontece administrando esses subconjuntos de itens e se movendo para cima ou para baixo de acordo com a taxa de sucesso em cada subconjunto. Após a administração de vários subconjuntos, a estimativa final da capacidade do candidato é obtida. Apesar da abordagem minimalista, esses métodos podem produzir os mesmos resultados que as técnicas mais sofisticadas de TAC usadas atualmente (Yao, 1991).

O uso de computadores veio facilitar o desenvolvimento dos testes adaptativos, sua administração e a seleção de itens. Reckase (1974) trouxe um dos primeiros exemplos da metodologia de testes adaptativos computadorizados. No início, a escassez e o alto custo dos programas e equipamentos limitava a utilização do TAC. Mas, devido à evolução e popularização dos computadores, o uso de TAC vem se tornando lugar comum, obrigatório em muitos casos.

1.4.2. Quem está usando os testes adaptativos computadorizados

Assim como vimos no tópico 1.3 que a TRI vem crescendo rapidamente como um novo paradigma para elaboração, análise e pontuação de testes, o movimento em direção aos testes adaptativos computadorizados também tem sido muito grande. Ainda nos primórdios dos TAC, em 1989, a Educational Testing Service (ETS) publicou um primeiro estudo com os resultados da aplicação de uma versão computadorizada e adaptativa do famoso Test of English as a Foreign Language (TOEFL), destinado à avaliação da proficiência na língua inglesa. O estudo concluiu que o teste funcionou satisfatoriamente, obtendo-se maior precisão nos extremos da distribuição de habilidade e sendo realizado em um tempo menor do que a

aplicação original, em lápis e papel. Além disso, vejam só, mesmo considerando que 72% dos participantes nunca tinham sido expostos a um computador antes da experiência do teste, apenas 41% disseram preferir fazer a prova no estilo lápis e papel (Hicks, 1989). Atualmente, algumas seções do TOEFL são administradas no modelo adaptativo computadorizado.

Logo em seguida, nos anos 1990, foi a vez do Armed Services Vocational Aptitude Battery (ASVAB) se render à abordagem dos testes adaptativos computadorizados. Conforme o site <http://official-asvab.com>, o ASVAB é uma bateria de testes para avaliação de múltiplas aptidões que tem a função de prever o sucesso acadêmico e profissional no serviço militar, sendo administrado anualmente para mais de um milhão de pessoas nos Estados Unidos. Após vinte anos de pesquisas, a versão adaptativa do ASVAB, denominada CAT-ASVAB, foi implementada operacionalmente durante os anos 1996 e 1997 pelo ministério da defesa daquele país em todos os centros de admissão de militares (Military Entrance Processing Stations - MEPS), sendo considerada a primeira bateria de testes adaptativos administrada em larga escala para um ambiente de missão crítica, como a dos militares.

Pegando exemplos mais recentes, em 2011, um estudo da Educational Testing Service (ETS) realizado nos EUA relatou que alguns Estados norte-americanos, tais como Idaho, Óregon, Dakota do Sul, Delaware e Carolina do Norte, já estavam utilizando testes adaptativos computadorizados baseados em TRI para avaliação educacional nos ensinos fundamental ou médio. A Carolina do Norte, inclusive, publicara um conjunto de recomendações para os Estados que pretendessem migrar para a "próxima geração das avaliações" (Stone & Davey, 2011). Um ano depois, em 2012, uma matéria da revista Education Week intitulada "Shifting to Adaptive Testing" já afirmava que pelo menos vinte Estados norte-americanos pretendiam passar a utilizar testes adaptativos até 2016 e que uma coalisão chamada Smarter Balanced Assessment Consortium recebera fundos federais para desenvolver testes adaptativos de alcance nacional com alta tecnologia, incorporando vídeos

e gráficos em itens interativos, projetados tanto para identificar a proficiência dos alunos quanto para serem mais envolventes (Davis, 2012).

Assim como as entidades voltadas para o ensino formal, os cursos para ensino de idiomas também vêm se beneficiando dos testes adaptativos, principalmente aqueles cujos modelos se baseiam em computadores, uma vez que permitem automaticamente avaliar e acompanhar a proficiência dos seus alunos, indicando os conteúdos que devem ser estudados de forma ajustada ao desenvolvimento individual dos alunos, como no caso da DynEd International (<http://web2.dyned.com/us/products/placementtest/>), uma gigante mundial no ramo do ensino de inglês que possui dezenas de milhões de estudantes nos cinco continentes e cujo método de ensino prevê que cerca de 80% do curso seja realizado a distância.

Os testes adaptativos computadorizados que nem sempre são baseados na TRI também vêm favorecendo a clínica médica e psicológica. Uma investigação feita por Gibbons et al. (2008) sobre uso de TAC em avaliações psiquiátricas revelou que o tamanho reduzido dos testes, o direcionamento individualizado dos itens com administração apenas de itens relevantes e a autonomia do teste na seleção e apresentação dos itens pode diminuir drasticamente o tempo de administração dos testes, significando menos cansaço para médicos e pacientes que podem usar o tempo que sobra para atividades mais produtivas. Além disso, o mesmo estudo também ressaltou que, virtualmente, qualquer questionário pode ser personalizado com o uso de TAC, podendo facilitar a criação de medidas de saúde mental para crianças que fossem mais ajustadas para o estágio de desenvolvimento da criança do que os questionários tradicionais, bem como ser usado para avaliar toda a matriz de psicopatologia e fazer um diagnóstico confiável na maioria das situações clínicas.

A prática tem demonstrado que Gibbons et al. tem razão. A cada dia surgem diversos novas investigações e aplicações onde os testes adaptativos computadorizados são utilizados

para apoiar a clínica médica, obtendo resultados precisos, através de testes mais objetivos e administrados em um tempo menor, trazendo, dessa forma, mais conforto para médicos e pacientes. Podemos citar, como exemplos, estudos para avaliação do nível de bulling (Ma, Chien, Wang, Li, & Yui, 2014), de ansiedade (Gibbons et al., 2014), de depressão (Fliege et al., 2009) e (Smits, Cuijpers & van Straten, 2011), do equilíbrio em pacientes com acidente vascular cerebral (Hsueh et al., 2010), da dor lombar em contraste com a Back Pain Functional Scale (BPFS) (Hart, Mioduski, Werneke & Stratford, 2006), do impacto da dor de cabeça (Ware et al., 2003) e para avaliação do sofrimento em pacientes com câncer (Walker, Böhnke, Cerny & Strasser, 2010), entre muitas outras possibilidades.

Algumas iniciativas desprendidas da visão do lucro imediato também ajudam a difundir os TAC. A University of Cambridge, através do seu Psychometrics Centre, oferece gratuitamente a plataforma Concerto (<http://www.psychometrics.cam.ac.uk/newconcerto>), um ambiente baseado na internet de código aberto para construção e administração de testes adaptativos online de qualquer porte e natureza, desde simples questionários até complexos testes baseados na TRI. A plataforma Concerto pode ser acessada de computadores e dispositivos móveis, como tablets e smartphones, admitindo, inclusive, o uso de itens multimídia.

No Brasil, ainda não é possível enxergar nada semelhante à explosão de estudos e aplicações que pode ser observada em países como Estados Unidos e Inglaterra, mas o assunto também tem estado no foco de entidades públicas e privadas, principalmente aquelas ligadas à avaliação da aprendizagem escolar. Segundo matéria publicada em junho de 2012 no blog Porvir (<http://porvir.org/porcriar/testes-sob-medida-sao-tendencia-em-grandes-exames/20120603>), o Centro de Seleção e de Promoção de Eventos da Universidade de Brasília (CESPE/UnB), principal organizador de concursos públicos do país, teria anunciado que começaria a utilizar os testes adaptativos em concursos públicos e em exames de

certificação em escala nacional. Projetos no mesmo sentido também estariam em andamento no Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), órgão do Ministério da Educação (MEC) responsável por avaliações nacionais como o Enem e a Prova Brasil, e na Fundação Vunesp, que faz o vestibular da Universidade Estadual Paulista (Unesp) e outros concursos. Ainda no Brasil, empresas privadas, tais como a Água-Marinha Mídia Educacional (www.aguamarinha.com.br) e o Geek Labs (www.geekie.com.br) também já oferecem serviços para a avaliação do aprendizado e orientação educacional de estudantes baseados em testes adaptativos.

1.4.3. Como o TAC funciona

Na forma mais comum dos testes adaptativos computadorizados baseados na TRI, os itens vão sendo selecionados pelo computador a partir de um banco de itens com características devidamente conhecidas de forma a combinar a dificuldade do item com o nível estimado da habilidade (ou proficiência, capacidade, nível de aptidão, etc.) de cada pessoa que realiza o teste. Se o participante for bem sucedido em um item, outro item um pouco mais desafiador será apresentado em seguida. Caso contrário, um item com menor dificuldade será apresentado. Geralmente, essa técnica permite convergir rapidamente para a sequência de itens que delimitam o nível de habilidade efetiva do participante. Normalmente, o teste é interrompido quando a capacidade do participante é determinada com a precisão necessária, mas outros critérios de parada podem ser definidos, tal como a quantidade total de itens administrados. Ao final, o resultado do teste pode ser fornecido imediatamente. A avaliação de novos itens para inclusão no banco de itens, bem como a validação da qualidade dos itens existentes pode se dar simultaneamente com a administração dos testes.

Thissen e Mislevy (2000) resumiram o método básico para TAC como um algoritmo iterativo consistindo dos seguintes passos:

1. Um item condizente com a estimativa de habilidade corrente do participante é selecionado dentro do conjunto de itens disponíveis;
2. O item escolhido é apresentado para o participante, que, por sua vez, pode produzir uma resposta correta ou incorreta;
3. A estimativa de habilidade do respondente é atualizada com base em todas as respostas obtidas até então;
4. Os passos de 1 a 3 são repetidos até que um critério de parada pré-definido seja satisfeito.

Uma representação gráfica de um teste adaptativo computadorizado está ilustrada na Figura 3, a seguir.

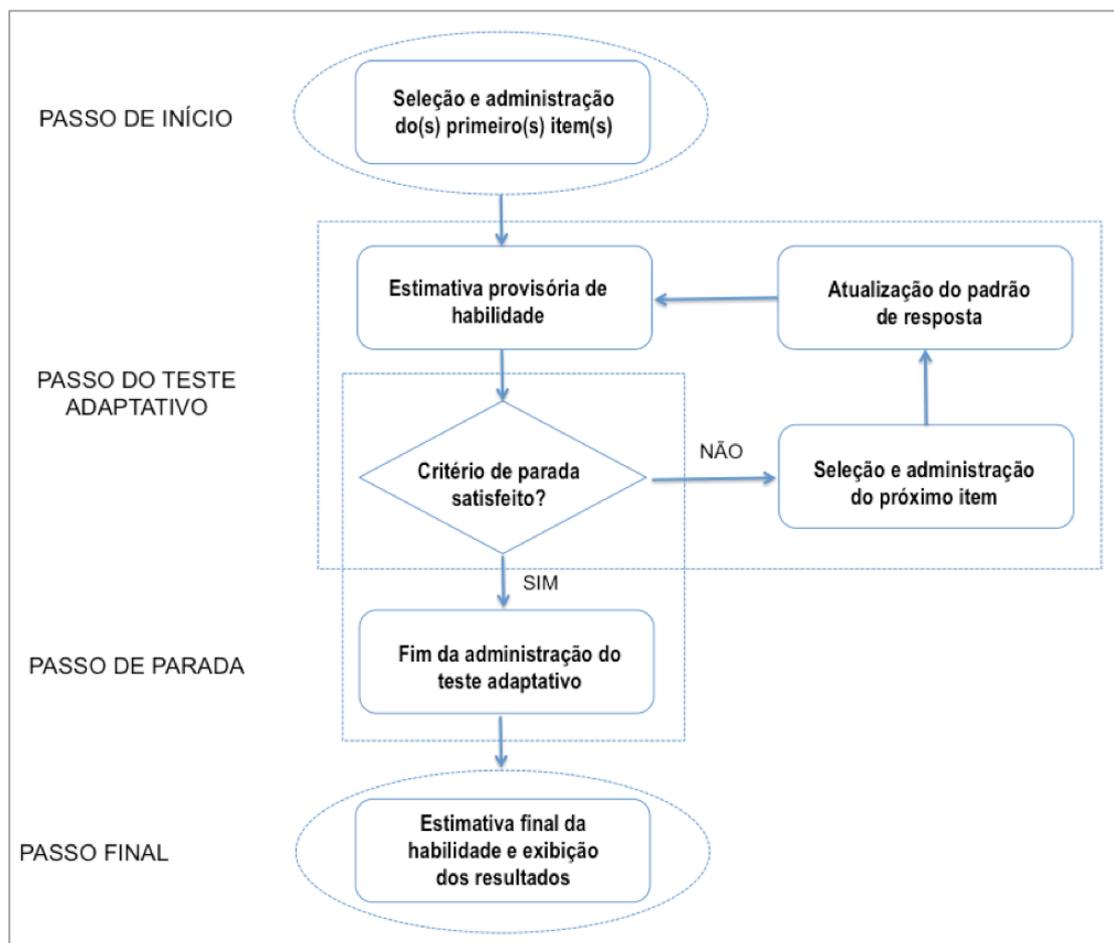


Figura 3 - Exemplo de algoritmo para um teste adaptativo computadorizado

Como resultado da administração adaptativa de itens, os diferentes participantes receberão testes também diferentes (Green, 2000). Normalmente, a TRI é a tecnologia utilizada para equiparar os escores obtidos através desses diferentes conjuntos de itens, bem como também a metodologia preferida para a seleção dos itens durante a aplicação dos testes (Thissen and Mislevy, 2000). Porém, é saudável registrar que nem todas as abordagens de testes adaptativos precisam ser baseadas na teoria da resposta ao item. Por exemplo, quando um teste está sendo usado para fins de classificação o problema pode ser abordado a partir da perspectiva da TRI ou a partir da perspectiva da teoria da decisão. Lawrence Rudner propôs a abordagem de uma teoria da decisão de medição que denominou measurement decision theory (MDT) para determinar maestria (testes de aprovação ou reprovação) ou outras decisões de classificação dicotômica. Não avançaremos nesse campo por estar fora do escopo deste estudo, mas seu site (<http://echo.edres.org:8080/mdt/>) descreve como essa abordagem funciona e traz um tutorial interativo sobre MDT, bem como apresenta outros recursos e referências sobre MDT e questões conexas.

De acordo com Weiss e Kingsbury (1984), deixando-se de lado questões práticas, tais como os processos de elaboração e pré-testagem dos itens, pode-se elencar cinco componentes técnicos no processo de construção de um teste adaptativo computadorizado:

1. Conjunto de itens calibrados
2. Nível de habilidade de entrada (primeiro item)
3. Algoritmo de seleção do próximo item
4. Procedimento de pontuação
5. Critério de parada

Importante perceber que os componentes 2, 3, 4 e 5 envolvem, direta ou indiretamente, estimativas dinâmicas de habilidades, significando que um TAC será tão bem sucedido quanto for sua capacidade para calcular acertadamente as estimativas de habilidades dos participantes.

1.4.3.1. Conjunto de itens calibrados

Antes de mais nada, um banco de itens deve estar disponível para o algoritmo do programa de computador que está por trás do TAC fazer suas escolhas. O banco deve ser calibrado de acordo com o modelo psicométrico que será utilizado como base para o programa - normalmente, os modelos da TRI são empregados nesses testes (Weiss & Kingsbury, 1984). A TRI se tornou popular porque coloca pessoas e itens na mesma métrica, aspecto bastante útil para a seleção de itens.

De acordo com Thompson e Weiss (2011), muitas vezes os testes podem requerer situações forçadas por determinadas características não psicométricas. Um exemplo típico dessa natureza é quando há imposições de conteúdos. Por exemplo, um teste de matemática pode exigir certa distribuição de itens de forma a abranger álgebra, geometria e probabilidade. Nesses casos, estratégias como a do balanceamento de conteúdo (Olea et.al. 1999, Kingsbury e Zara, 1989, Wainer e Kiely, 1987) ou de testlets (Wainer & Kiely, 1987) podem ser empregadas na construção do banco de itens. O balanceamento de conteúdos divide o banco de itens em várias seções, cada uma delas representando um conteúdo específico que se deseja avaliar, eliminando a necessidade de segmentar o banco de itens em vários pequenos bancos. Já os testlets representam grupos de itens relacionados a um determinado conteúdo desenvolvidos como se fossem “unidades de teste” com uma quantidade predeterminada de itens. A construção híbrida de um banco de itens que implementa o balanceamento de conteúdo e os testlets pode ser muito proveitosa, já que é

possível garantir que um determinado exame avalie várias habilidades em um único exame. No entanto, restrições desse tipo podem reduzir a eficiência do algoritmo adaptativo porque eles impedem o processo de seleção natural de escolha dos itens. Como essas restrições podem ser muito importantes a partir de uma perspectiva mais ampla, elas devem ser consideradas quando do projeto e da construção do banco de itens e do algoritmo do teste.

1.4.3.2. *Estimativas de habilidades*

As estimativas de habilidades representam um ponto de central importância nos testes adaptativos. Ao contrário do que ocorre nos testes sequenciais, nos testes adaptativos podemos distinguir três estágios onde se é necessário produzir estimativas de habilidades: (1) a estimação de habilidade para iniciar o procedimento de seleção de itens no teste; (2) a estimação de habilidade durante o teste para adaptar a seleção de cada item do teste à habilidade demonstrada pelos participantes; e (3) a estimação de habilidade ao final do teste para reportar o escore do participante. Cada um desses estágios envolve seus próprios requisitos e problemas (van der Linden & Pashley, 2010).

Diversos métodos para produzir as estimativas de habilidades têm sido usados no aprimoramento dos testes adaptativos computadorizados. No passado, o método da máxima verossimilhança, maximum-likelihood (ML) estimator, no termo em inglês, era a escolha mais popular (van der Linden & Pashley 2010). Formulada por Ronald A. Fisher (1922), a estimativa por máxima verossimilhança foi um dos mais importantes desenvolvimentos em estatística do século XX e representa os valores dos parâmetros que maximizam a probabilidade de que os dados observados teriam sido gerados (Aldrich, 1997).

Atualmente, técnicas de estimação alternativas ao ML têm sido empregadas com frequência, tal como o weighted likelihood estimator (WLE), de Warm (1989), um modelo ponderado de estimação, e os modelos com abordagem das probabilidades Bayesianas, como

o Bayes modal (BM) ou maximum a posteriori (MAP) estimator e o expected a posteriori (EAP) estimator. Probabilidade Bayesiana é o nome dado a várias interpretações relacionadas de probabilidade que têm em comum a noção de probabilidade como uma espécie de crença parcial, ao invés de uma frequência, e são baseadas no teorema de Bayes, que relaciona a probabilidade atual com a probabilidade prévia ("Bayes' theorem," n.d.). Alguns termos estatísticos são de especial interesse quando falamos em testes adaptativos:

Erro padrão (SE - standard error) - é o desvio-padrão da distribuição amostral de uma estatística. O erro padrão é um termo estatístico que mede a precisão com que uma amostra representa uma população ou um conjunto de dados. No caso da medida de habilidade, o erro padrão é uma estimativa do desvio padrão da estimativa de capacidade que pode ser esperada para um determinado participante.

Escore verdadeiro - é o escore que um teste absolutamente confiável, sem nenhuma margem de erro, forneceria a quem o realizasse. Porém, desde que todos os testes contêm erros, o escore verdadeiro é apenas um conceito teórico. No atual estado da arte dos testes psicométricos nós não temos condições de conhecer o verdadeiro escore de um indivíduo, mas podemos, entretanto, determinar uma estimativa do escore verdadeiro dessa pessoa e estimar a quantidade de erro existente nessa estimativa.

Estimativa de habilidade - é o escore, parcial ou final, que a pessoa recebe ao fazer um teste real, podendo ser entendida como o escore verdadeiro adicionado do erro estimado na medida.

Confiabilidade e informação: uma das principais contribuições da teoria de resposta ao item à psicometria é a extensão do conceito de confiabilidade. Confiabilidade refere-se à precisão da medição, o grau no qual a medição está livre de erros. Na

teoria clássica dos testes, o índice tradicionalmente utilizado para nos dizer sobre a utilidade de um teste é o erro padrão da medição (SEM – standard error of the mean), assumindo que os escores brutos (raw scores) sobre os testes e os itens do teste sejam uma combinação do escore verdadeiro com erros aleatórios. Assim, o SEM refere-se à "distribuição de erros aleatórios em torno do escore verdadeiro" (Kline, 2005, p. 92) e um único valor para o SEM é dado para o teste como um todo. Quanto menor o valor do SEM, mais confiável será a pontuação do teste. Por outro lado, "a TRI utiliza o conceito de informação do teste e informação do item. A informação é [inversamente relacionada com] o SEM e é calculada separadamente para diferentes níveis de habilidade. A função de informação do teste indica o grau de certeza com que cada nível de habilidade está sendo estimado pelo teste" (Thorpe et al., 2007, p. 179). A informação também passa a ser uma função dependente dos parâmetros do modelo. Por exemplo, de acordo com a teoria da informação de Fisher, a informação fornecida pelo item no caso do modelo 1PL para respostas dicotômicas é, simplesmente, a multiplicação da probabilidade de uma resposta correta pela probabilidade de uma resposta incorreta. A TRI deixa claro que a precisão da medida não é uniforme em toda a gama de resultados do teste - as pontuações nas extremidades da escala de habilidade de um teste, têm, geralmente, mais erros do que aquelas localizadas na faixa central da escala.

1.4.3.3. Nível de habilidade de entrada (primeiro item)

No TAC os itens são selecionados com base no desempenho do participante a cada ponto no teste. A questão é como determinar o primeiro item quando ainda não se conhece nada sobre a habilidade do participante, ou seja, como estimar a capacidade inicial de um

participante desconhecido? Apresentar um item muito acima da sua capacidade pode gerar estresse e, ao contrário, um item muito fácil pode gerar desinteresse.

Como soluções possíveis para esse problema, muitas vezes são administrados alguns itens no início do teste apenas para criar a primeira estimativa de habilidade; noutras vezes, admite-se que a capacidade do participante seja mediana e um item com dificuldade média ou de média para fácil será apresentado; quando disponíveis informações anteriores sobre a habilidade do participante, ela também pode ser usada para determinar o primeiro item a ser exibido no teste. Métodos Bayesianos, tais como o *expected a posteriori* (EAP), têm sido utilizados para estimar a habilidade do participante com poucos itens administrados. Havemos de nos preocupar, também, com a exposição do item e evitar que a estratégia adotada leve a apresentar o mesmo item inicial para todos os participantes (van der Linden & Pashley, 2010), deixando-o muito conhecido.

Na verdade, nenhuma dessas soluções dá conta totalmente do problema da estimativa inicial de habilidade, mas se o teste não for muito curto (apenas 10 itens, por exemplo) haverá tempo para o TAC se ajustar e se recuperar de uma estimativa inicial ruim (van der Linden & Pashley, 2010).

1.4.3.4. Algoritmo de seleção do próximo item

Como a TRI permite colocar a habilidade dos participantes e a dificuldade dos itens na mesma métrica, se o programa do computador tiver uma estimativa da capacidade do examinando ele será capaz de selecionar um item que seja mais apropriado para essa estimativa (Wainer & Mislevy, 2000). Tecnicamente, a seleção do item é feita pelo nível de dificuldade compatível com a proficiência que vai sendo demonstrada pelo participante ou escolhendo o item que apresente a maior discriminação (que representa a quantidade de informação) dentro do nível de habilidade estimado.

O método de máxima informação de Fisher (Maximum Fisher Information) é largamente utilizado para seleção dos itens que devem ser apresentados. O método seleciona o próximo item como sendo aquele que maximiza a informação dos itens existentes e disponíveis no banco de itens (Baker, 1992). Por levar em consideração a estimativa de habilidade calculada para a pessoa a cada momento do teste, maximizar a informação de Fisher significa, intuitivamente, selecionar um item com dificuldade que corresponda exatamente ao nível de habilidade estimada para o examinando, considerando os itens até então administrados. No caso dos modelos logísticos de três parâmetros, itens com maior discriminação, ou equivalentemente, alto valor do parâmetro alfa, serão preferencialmente selecionados pelo algoritmo.

Novamente, a exposição exagerada do item também pode ser um fator de preocupação em muitos casos. Quando não se deseja permitir que os itens fiquem conhecidos é comum utilizar métodos específicos para realizar o controle da exposição do item.

1.4.3.5. Procedimento de pontuação

Depois que um item é administrado, o TAC atualiza a estimativa do nível de habilidade do participante. Se o participante foi bem sucedido no item, provavelmente, o TAC irá elevar um pouco a estimativa de habilidade do participante e vice-versa. Para isso, utiliza-se a função de resposta ao item oriunda da TRI para se obter uma função de probabilidade da capacidade do examinando. Os métodos máxima verossimilhança (maximum-likelihood estimation- ML) e estimativa Bayesiana, ou uma combinação deles, são muito utilizados para calcular o nível de habilidade do participante.

1.4.3.6. *Critério (ou regra) de parada*

A decisão quanto ao momento de finalizar um teste adaptativo é outro fator crucial para o sucesso da aplicação. Testes muito curtos podem levar a estimativas imprecisas de habilidade. Por outro lado, se o teste for muito longo, então teremos tempo e recursos desperdiçados, itens expostos desnecessariamente e o participante pode cansar e comprometer o seu nível de desempenho levando a resultados inválidos. Assim, um ou mais critérios devem ser fixados para determinar quando o teste deve ser encerrado. Segundo Linacre (2000), um teste adaptativo computadorizado irá parar quando encontrar uma das seguintes situações:

- a) Quando o banco de itens foi exaurido (todos os itens foram administrados);
- b) Quando o número máximo de itens previsto para o teste foi alcançado;
- c) Quando a medida de habilidade for estimada com precisão suficiente;
- d) Em testes onde o resultado é apenas aprovar ou reprovar (pass-fail) o participante, quando a medida de habilidade corrente já permitir uma decisão estatística que não possa mais ser alterada, seja pelo nível de habilidade demonstrado, seja pela pequena quantidade de itens restantes;
- e) Se o participante estiver exibindo comportamento inapropriado ou desinteressado.

Por outro lado, um TAC não deve parar antes que:

- a) Um número mínimo previsto de itens seja administrado;
- b) Todos os tópicos previstos no teste terem sido cobertos;

- c) Tenha-se administrado um número suficiente de itens para evitar questionamentos. Testes muito longos para uns e muito curtos para outros podem levar os participantes a anteciparem o resultado do teste e gerar situações desconfortáveis, principalmente em testes onde a aprovação ou reprovação está em jogo.

Para se obter o máximo de benefícios do TAC, nem o tempo nem a quantidade de itens do teste deveriam ser impostos como critérios de parada. Porém, recomenda-se garantir que um número mínimo de itens seja exibido para garantir maior confiabilidade da medida, assim como se determinar um número máximo de itens para evitar que o teste seja muito longo. Normalmente, o melhor indicador de que as estimativas calculadas pelo teste são aceitáveis e que o teste pode ser finalizado será quando o erro padrão associado à estimativa de habilidade for pequeno e se mantiver estável ou decrescente ao longo da administração de alguns itens.

1.4.4. Exemplo de TAC

Apenas para efeito de exemplificar a técnica, baseado em Linacre (2000) vamos supor a administração de um teste qualquer com respostas dicotômicas (múltipla escolha, por exemplo) para uma criança com proficiência de 50 unidades em uma escala que vai de 0 a 100 unidades. Ressalta-se que o computador não sabe o nível de proficiência da criança a priori. A aplicação do teste simulado também pode ser acompanhada na Figura 4.

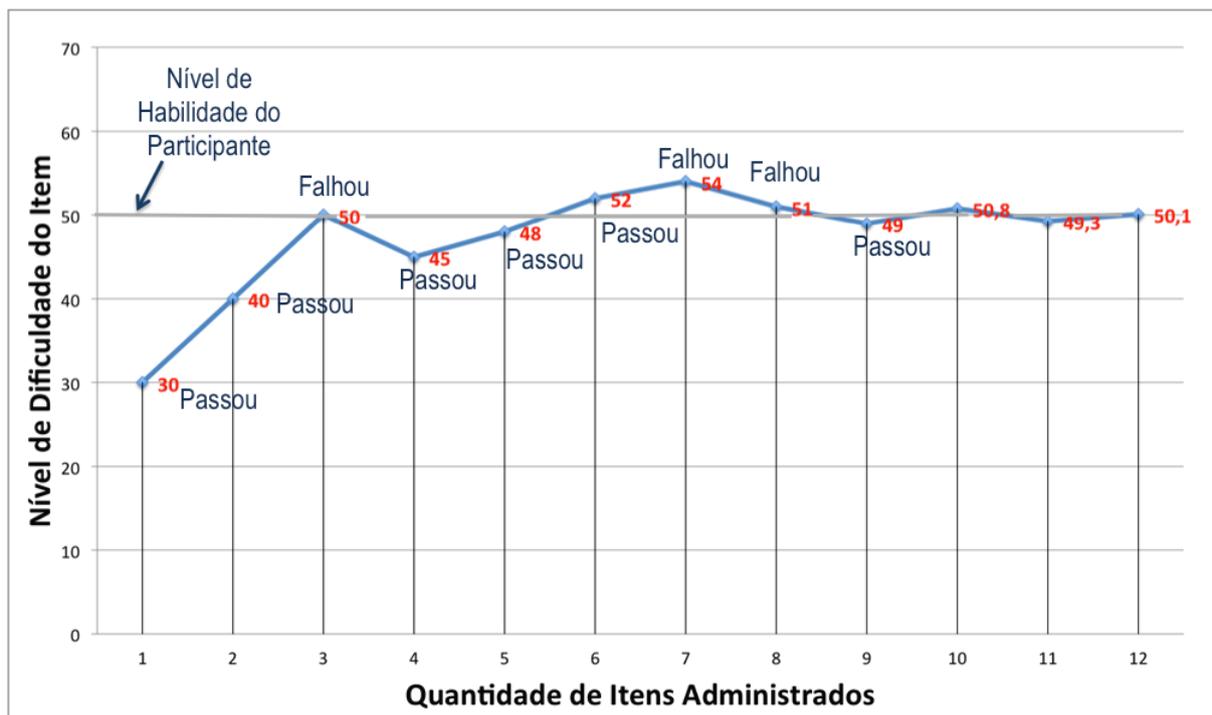


Figura 4 - Exemplo de administração de itens em teste adaptativo. A cada item administrado o programa vai melhorando a estimativa de habilidade e diminuindo a estimativa do erro da medida.

1. Suponha que foi escolhido para primeiro item a ser administrado um item com 30 unidades de dificuldade. Como o nível de habilidade da criança é de 50 unidades, muito provavelmente ela deverá ser bem sucedida nesse primeiro item.
2. Em seguida, o computador seleciona um item mais difícil, de 40 unidades, e a criança é bem sucedida novamente.

3. O computador seleciona um item mais difícil ainda, de 50 unidades de dificuldade. Neste caso, a habilidade da criança e a dificuldade do item têm o mesmo nível, significando que a criança terá 50% de chance de sucesso. Suponhamos que a criança não acerte.
4. Agora, o computador administrará um item um pouco mais fácil do que o de 50 unidades, mas que deve ser mais difícil do que 40 unidades, que foi o nível de dificuldade do último item onde a criança foi bem sucedida. Então, um item com 45 unidades é administrado. A criança passa.
5. O computador administra um item mais difícil, de 48 unidades e a criança passa novamente.
6. Tendo em vista o sucesso da criança nos itens entre 40 e 48 unidades, há evidências para suspeitar que o fracasso da criança no item de 50 unidades possa ter sido por desatenção ou qualquer especificidade do item. Então, o computador administra um item de dificuldade 52. Como a dificuldade desse item é apenas um pouco superior ao nível de habilidade da criança, sua chance de sucesso estará só um pouco abaixo de 50%. A criança acerta o item.
7. O computador sobe o nível e administra um item com dificuldade de 54 unidades. A criança não acerta.
8. O computador recua o nível de dificuldade e administra um item de 51 unidades. A criança falha, novamente.
9. Um item de 49 unidades é administrado. A criança acerta.
10. O processo continua. O programa de computador vai ficando cada vez mais certo de que o nível de habilidade da criança está perto de 50 unidades. Quanto mais itens são administrados, mais precisa essa estimativa de capacidade se torna.

Alguns critérios inseridos na programação definem quando o computador deve encerrar a aplicação do teste. Observando-se o gráfico da Figura 4 fica fácil perceber a tendência do teste de estimar a habilidade final da criança para algo em torno de 50 unidades.

1.4.5. Vantagens do TAC

Diversas são as vantagens da aplicação de testes adaptativos computadorizados em comparação com os testes sequenciais tradicionais. Podemos citar:

- Os testes podem ser mais curtos e rápidos, devido à administração de uma quantidade menor de itens do que no teste com papel e lápis (Linacre, 2000) (Rudner, 1998).
- Itens irrelevantes, com pouca ou nenhuma informação, são identificados e eliminados (Linacre, 2000).
- Sem muitos itens incompatíveis com a habilidade do participante, o teste pode representar uma experiência melhor, mais desafiadora, atraente e justa (Linacre, 2000).
- Os testes podem ser realizados por demanda, com horários flexíveis e os resultados obtidos imediatamente (Rudner, 1998).
- Apesar de continuar a exigência de supervisão da aplicação do teste, não há necessidade de fichas nem de treinamentos extensos para os administradores dos testes. As diferenças quanto aos aplicadores dos testes são eliminadas (Rudner, 1998).
- Há melhor controle de exposição dos itens e segurança do teste.

- Pode prover escores acurados para uma grande faixa de nível de habilidade, enquanto os testes tradicionais são mais acurados para habilidades medianas (Rudner, 1998).
- Permite atualizar o banco de itens de forma simples e rápida.
- Pode ser uma excelente escolha em testes de auto avaliação, muito úteis na educação a distância, por exemplo (Piton-Gonçalves et al., 2009).

1.4.6. Desvantagens do TAC

As desvantagens do teste adaptativo computadorizado incluem a necessidade e o custo de equipamentos e instalações, a pouca oferta de programas de computador para administração de testes dessa natureza (embora crescente), bem como a própria dificuldade de alguns aplicadores de testes lidarem com softwares e equipamentos de informática,

Além disso, em 1992, o The National Center for Fair & Open Testing (FairTest), uma organização educacional norte-americana que aborda questões relacionadas com a justiça e a precisão da aplicação e pontuação de testes em estudantes, já listava alguns problemas ainda não resolvidos em TAC e pontos onde devemos ter cuidado:

- Há relatos, corroborados por pesquisas, de problemas na equivalência da estimativa de habilidade entre os testes tradicionais e os TAC que podem mascarar variações individuais. Além disso, a performance em alguns tipos de perguntas pode ser diferente nos dois tipos de testes (Bugbee and Bernt, 1990).
- Testes computadorizados constroem mais os examinandos do que os testes de papel e lápis. Em versões adaptativas computadorizadas os examinandos não podem, por exemplo, sublinhar partes do texto, riscar opções eliminadas

ou pular questões para voltar depois - todas são estratégias bastante utilizadas, principalmente em testes de realização. Estudos também sugerem que demoramos mais para ler ou detectar erros na tela do computador do que em materiais impressos (Bugbee e Bernt, 1990).

- Os usuários mais habilidosos com computadores podem ser favorecidos em itens ou programas que requeiram maior interatividade, tal como rolagem de tela.
- Com cada participante recebendo um conjunto diferente de perguntas, eventuais injustiças podem não ser percebidas. É essencial que os testes administrem itens suficientes para garantir que a sua estimativa de habilidade final seja verdadeira. Um intervalo de confiança aceitável deve garantir a confiabilidade do teste Rudner (1998).
- O teste adaptativo não é aplicável para todas as disciplinas e competências. A maioria dos TAC são baseados em modelos da TRI, mas a TRI não é aplicável a todas as habilidades e tipos de item. O mesmo também acontece nos testes tradicionais.

1.4.7. Teste adaptativo multidimensional

Embora não seja foco desse trabalho, aproveitamos para registrar que abordagens multidimensionais da teoria da resposta ao item, designadas pelo termo inglês multidimensional item response theory (MIRT), vêm ganhando atenção justamente por permitir combinar e avaliar mais de uma habilidade em um único teste (Reckase, 1985, 1997).

Um teste adaptativo construído sobre MIRT, chamado teste adaptativo multidimensional, oferece ao menos duas vantagens sobre os testes adaptativos

unidimensionais: (1) a abordagem multidimensional inclui mais informações do que a unidimensional, uma vez que múltiplas subescalas de habilidade estão sendo avaliadas e correlacionadas; e (2) a abordagem multidimensional pode balancear os conteúdos automaticamente, sem a necessidade de recorrer totalmente a técnicas de balanceamento de conteúdo (Wang & Chang, 2011).

Capítulo II

Escala de Competências Cognitivas para Crianças dos 4 aos 10 anos de Idade (ECCOs 4/10)

2. ESCALA DE COMPETÊNCIAS COGNITIVAS PARA CRIANÇAS DOS 4 AOS 10 ANOS DE IDADE (ECCOs 4/10)

2.1. Introdução

A inteligência tem sido um dos construtos mais estudado pela psicologia. Segundo Roazzi (2008), "o início do estudo científico da inteligência se confunde com as próprias origens da psicologia como ciência, ambos ocorrendo entre o final do século XIX até o início do século XX". Mas, apesar de tamanha relevância, Brito (2009) nos diz que a necessidade de novos instrumentos para avaliação cognitiva para as faixas etárias mais baixas continua a ser uma realidade, pois são poucos os testes para avaliação da inteligência e da cognição destinados à infância disponíveis e atualizados para a língua portuguesa.

Indo ao encontro dessa necessidade, o Instituto de Educação e Psicologia da Universidade do Minho (UMinho), de Portugal, vem desenvolvendo pesquisas no sentido de aumentar a oferta de ferramentas para avaliação da inteligência em crianças. Nesse sentido, em janeiro de 2009 foi publicada a Escala de Avaliação de Competências Cognitivas para Crianças dos 4 aos 10 anos de Idade (ECCOs 4/10), uma bateria extensa de testes voltada para avaliação cognitiva na infância, particularmente nos anos que respeitam a fase pré-escolar e o primeiro ciclo do ensino básico, constituindo uma versão ampliada e adaptada de sua congênere, a ECCOs 4/7. Aferida à zona Norte de Portugal continental, a ECCOs 4/10 é o resultado de um longo trabalho iniciado no mestrado e continuado no doutorado da Dra. Maria de Lurdes Dias Brito (UMinho) sob a orientação do Prof. Dr. Leandro Almeida (UMinho).

2.2. A estrutura da ECCOs 4/10

Confluindo diversas posturas teóricas sobre a questão da cognição, a ECCOs 4/10 foi elaborada assumindo a visão de Almeida (2008) de que não se pode aceitar uma única definição de inteligência porque ela não existe, importando antes atender à diversidade de formas que a cognição humana pode assumir face às realidades socioculturais dos indivíduos. Segundo Brito (2009), subjacente ao modelo de avaliação apresentado na ECCOs 4/10 está a ideia de uma inteligência composta, abordando desde processos cognitivos mais simples até processos cognitivos mais exigentes, para os quais são apresentadas tarefas que mantêm a sua tipologia ao longo das idades abarcadas, adequando o nível de exigência dos seus itens ao longo dessas várias faixas etárias.

Brito e Almeida (2009) relataram que na ECCOs 4/10 se procurou que cada processo cognitivo pudesse ser avaliado através de provas apoiadas sobre aspectos verbais e linguísticos, onde se requer maiores valências de comunicação por parte da criança, e, também, por provas que possam libertá-la tanto quanto possível desse domínio e lhe permitam um desempenho que apele a conteúdos figurativos, perceptivos, manipulativos, espaciais e de coordenação psicomotora, ainda que esta diferenciação não seja estanque. Dessa forma, as tarefas da ECCOs 4/10 podem ser essencialmente verbais e não verbais, assim como as Escalas de Wechsler ou de Stanford-Binet (Vernon, 1950; Kaufman, 1979; Almeida, 1994; Seabra-Santos, 1998). Esta diferenciação apoia-se, ainda, nos contributos da psicologia cognitiva e da neuropsicologia quando se sugere que o desempenho cognitivo se diferencia nos indivíduos de acordo com os processos cognitivos envolvidos nas tarefas e o conteúdo das mesmas (Grieve, 1995; Springer, & Deutsch, 1994; Habib, 2003).

Ao final, a ECCOs 4/10 combina seis processos cognitivos com tarefas de conteúdos essencialmente verbal (exceto para o processo de percepção) e tarefas de conteúdos

essencialmente não verbal, resultando num total de onze provas, definindo e fundamentando cada um dos mesmos em abordagens teóricas atuais. Essa estrutura está condensada no quadro a seguir.

Quadro 1 - Estrutura e provas da ECCOs 4/10

ECCOs 4/10		
Processos Cognitivos	Conteúdos	
	Verbal	Não verbal
Percepção	-----	Comparação de Figuras
Memória	Elementos em Frases	Elementos em Árvores
Compreensão	Frases Absurdas	Desenhos Absurdos
Raciocínio	Frases Incompletas	Figuras Incompletas
Resolução de problemas	Situações Quantitativas	Composição de Padrões
Pensamento divergente	Construção de Histórias	Construção de Figuras

Na ECCOs 4/10, procurou-se afixar as provas em uma sequência temporal de execução com uma exigência cognitiva cada vez mais complexa. Partiu-se de processos cognitivos mais básicos e que trabalham a recepção e codificação da informação vinda dos sentidos para a sua articulação com informação existente na memória, sendo este processo essencial para a atribuição de significado ao que é recebido. Estas fases são igualmente requeridas quando as situações com que a criança depara lhe exigem um raciocínio mais elaborado ou a resolução de um problema, processos já de si mais exigentes em termos cognitivos (Brito, 2009).

2.3. Processos cognitivos da ECCOs 4/10

As descrições que se seguem sobre as definições dos processos cognitivos da ECCOs 4/10 foram amplamente baseadas no seu Manual de Aplicação.

2.3.1. Percepção

No seio das teorias da inteligência e da cognição, bem como no modelo Cattell-Horn-Carroll (CHC) (McGrew & Flanagan, 1998), destacam-se três aptidões relacionadas com o processo perceptivo: (1) o processamento visual, entendido como a capacidade para gerar, perceber, reter, analisar, manipular e transformar imagens visuais (Primi & Almeida, 2002); (2) o processamento auditivo, relacionado com a percepção, análise e síntese de padrões sonoros envolvendo particularmente distorções ou estruturas musicais complexas (Primi & Almeida, 2002); e (3) a velocidade de processamento ou rapidez de decisão, ligada à velocidade em reagir ou tomar decisões (Primi & Almeida, 2002). Contudo, outras provas de avaliação cognitiva para crianças, tais como a SON-R 5 1/2-17 (*Snijders-Oomen Non-Verbal Intelligence Test*, 1989) centram a avaliação do processo cognitivo em provas que requerem a diferenciação de figuras em fundos complexos e, assim, mais ligadas a competências *gestálticas*.

Na ECCOs 4/10 a prova perceptiva está centrada na avaliação da percepção visual, articulando a capacidade de discriminação e de acuidade visuais com a velocidade desse processamento. Por outro lado, e entendendo que o processo perceptivo é indissociável do processo atencional, as tarefas presentes nesta prova, exigem concentração e atenção por parte da criança para análise e captação de detalhes.

Prova não verbal - Comparação de Figuras: apresenta-se à criança cinco imagens que diferem apenas em pequenos pormenores, devendo ela indicar as duas imagens que são

exatamente iguais. A prova tem um tempo máximo de execução, ajustável em função da idade do participante.

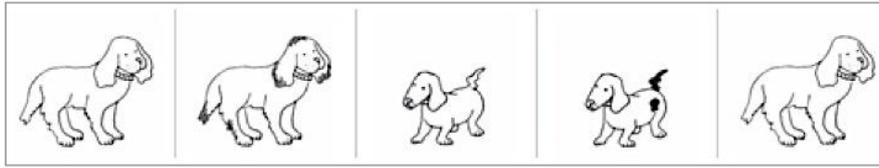


Figura 5 - Exemplo de item da prova não verbal Comparação de Figuras

2.3.2. Memória

De acordo com Garcia-Martinez e colaboradores (1997), a memória é um processo cognitivo complexo responsável pelo armazenamento e pela recuperação da informação (Tarrida, 2002). Em 1968, Atkinson e Shiffrin apresentaram o modelo de múltiplos armazenamentos, estruturados em três níveis: a memória sensorial, a memória de curto prazo e a memória de longo prazo. A memória sensorial corresponde às diferentes modalidades sensoriais e possui grande capacidade de registo, mas uma duração de apenas alguns milésimos de segundo. Muito desta informação se desvanece, acedendo o restante à memória de curto prazo. Neste nível, é possível a manipulação consciente da informação, muito embora o seu arquivo seja frágil pelas limitações de durabilidade que apresenta ou pelo fato da chegada de nova informação fazer decair ou substituir a existente. O armazenamento seguinte é a memória a longo prazo. A passagem da informação de curto para longo prazo obedece a sistemas de filtragem de tal modo que somente parte da informação permanece mais tempo guardada (sendo a restante esquecida), organizando-se, para isso, em categorias, associações ou hierarquias (Tiberghien, 1999). A capacidade de retenção e a duração da informação nesta memória são ilimitadas. As baterias de avaliação cognitiva privilegiam provas baseadas na memória de curto prazo, embora a avaliação da memória de longo prazo

apareça avaliada nas baterias de Woodcock-Johnson-Revised (WJ-R) ou no Diferencial Abilities Scale (DAS).

Em 1974, o modelo apresentado por Baddeley e Hitch fez nascer um novo conceito de memória de trabalho (*working memory*). Tratava-se de um sistema de capacidade limitada que interfere no armazenamento e na manipulação da informação durante a realização da tarefa cognitiva (Bayliss, Jarrold, Gunn, & Baddeley, 2003), não só mantendo ativa a informação que surge do meio, como manipulando-a e transformando-a para que responda às exigências de planificação e de orientação da conduta (Tirapu-Ustárrroz, & Muñoz-Céspedes, 2005). Este modelo de memória de trabalho está presente em algumas baterias de avaliação cognitiva destinadas à infância (como é o caso da WISC-IV ou da Stanford-Binet V). Na ECCOs 4/10, as provas de memória baseiam-se no conceito de memória a curto prazo, mas exigindo compreensão e retenção da informação e aproximando-as da memória de trabalho (Brito, Almeida, Viola, & Chaves, 2006).

Prova verbal - Elementos em Frases: trata-se de uma prova de memória auditiva na qual, a partir de pequenos fatos relacionados com as vivências da criança e através de termos que se encontram contextualizados, se procura avaliar as capacidades de atenção, retenção e evocação de elementos simples, no quadro de um contexto significativo.

Prova não verbal - Elementos em Árvores: o objetivo desta prova é avaliar a capacidade para reter e evocar elementos simples apresentados anteriormente. Procurou-se que o material utilizado tivesse maior significado para a criança, tornando-o agradável e lúdico, apresentando uma certa uniformidade de item para item.

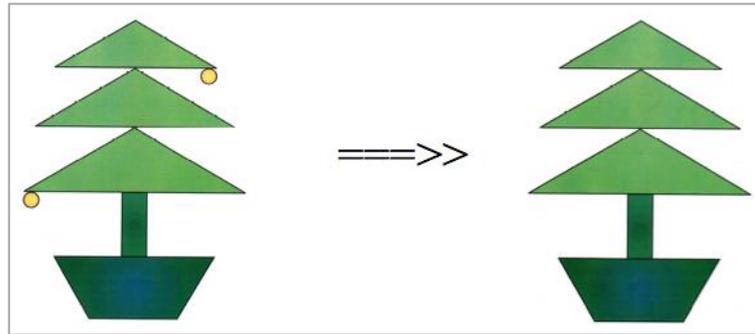


Figura 6 - Exemplo de item da prova não verbal Elementos em Árvore. Primeiro, mostra-se a árvore com os elementos. Depois de algum tempo a mesma árvore é mostrada sem os elementos e a criança deverá indicar onde eles estavam

2.3.3. Compreensão

Compreender implica estabelecer conexões entre conhecimentos, previamente arquivados na memória e dados que nos chegam da situação em análise, integrando elementos perceptivos e conceptuais (Kintsch, 1998). Assim, compreender é construir significações (Richard, 1995) a partir de representações, ou seja, de informações tidas em conta pelo processo cognitivo quando se realiza uma dada tarefa.

Na ECCOs 4/10, o processo de compreensão é abordado na realidade social, na forma como a criança conceitualiza as outras pessoas e como aborda os pensamentos, emoções, intenções e pontos de vista dos outros (Shantz, 1975).

Prova verbal - Frases Absurdas: é composta por pequenas frases nas quais é pedido à criança que identifique o que está errado, o que é absurdo em cada uma delas, procurando se avaliar a capacidade de compreensão e requerendo a organização dos elementos expressos verbalmente.

Prova não verbal - Desenhos Absurdos: é constituída por imagens relativas a acontecimentos familiares à criança, mas apresentando incongruências de elementos e contextos, através dos quais se pretende avaliar competências relativas ao conhecimento e

compreensão do meio e à organização semântica ou contextual dos diversos elementos constantes das figuras.



Figura 7- Exemplo de item da prova não verbal Desenhos Absurdos

2.3.4. Raciocínio

O raciocínio foi, desde cedo, um dos fatores presentes nas teorias fatoriais da inteligência. Em 1938, Thurstone considerava o raciocínio aritmético, a dedução e a indução, nas suas nove aptidões mentais primárias, entendendo-as como fatores autônomos. Os dois primeiros, sendo menos evidentes (Brody & Brody, 1976, citados por Almeida, 1988), foram reunidos por alguns autores num só fator que designaram por raciocínio, abarcando a indução e a dedução. Segundo Almeida (1988, 1994), o fator de raciocínio era definido como a capacidade para resolver problemas lógicos, através de tarefas que implicassem a descoberta e a aplicação de uma lei geral de sucessão de dígitos ou de transformação de figuras. Já em Guilford (1967), embora não se fale em raciocínio, encontra-se a operação de produção que, sendo convergente, diz respeito à possibilidade de recorrer a deduções lógicas ou a inferências. Nos diferentes estudos, o raciocínio apresenta-se como independente entre as várias aptidões mentais e, na sua essência, abarca a capacidade de recurso à lógica e à inferência para atingir uma solução única que responda à questão em causa. Já nos modelos ditos hierárquicos, a presença de uma inteligência fluída, encarada como uma capacidade

mental mais básica e liberta de aprendizagens (Cattell, 1971), faz aproximar este fator do que designamos por raciocínio.

Na ECCOs 4/10, o raciocínio é avaliado através de tarefas analógicas, à semelhança do que acontece noutras provas de avaliação cognitiva, tais como a WISC-IV, o DAS ou a SB-V (Brito, 2009).

Prova verbal - Frases Incompletas: está organizada num formato próximo ao de analogias, procurando-se que a criança forme uma pequena frase com as três palavras que lhe são lidas e que descubra a palavra que completa logicamente essa frase.

Prova não verbal - Imagens (ou figuras) Incompletas: tem também subjacente o modelo de analogia, requerendo a seleção de uma dentre cinco alternativas de respostas para que se conclua logicamente a analogia.

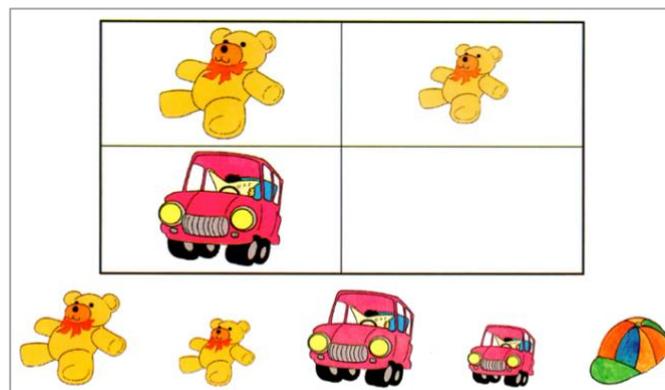


Figura 8 - Exemplo de item da prova não verbal Imagens Incompletas.

2.3.5. Resolução de problemas

A resolução de problemas é uma competência cognitiva complexa, talvez até a atividade humana mais inteligente (Chi & Glaser, 1992). Resolver problemas implica três elementos essenciais (Chi & Glaser, 1992): o estado inicial, o estado meta e o conjunto de recursos ou operadores que permitem a passagem do primeiro para o último estado (French & Colman, 1995). A Psicologia da *Gestalt* enfatiza na resolução de problemas a forma como os

elementos se organizam entre si, passando a busca da solução pela reorganização dos elementos num todo, dando-lhes uma nova (e boa) forma ou *gestalt*. A teoria do processamento da informação (Newell & Simon, 1972) enfatiza os processos cognitivos que se vão sucedendo ao longo dessa realização (Chi & Glaser, 1992). A atenção está no desenrolar das fases ao longo das quais a informação vai sendo processada, interessando conhecer os processos subjacentes desde a entrada da informação à sua saída. O primeiro passo será o da representação mental do problema que implica que este seja interpretado e compreendido por quem o recebe face aos conhecimentos e experiências prévios. O passo seguinte é o da aplicação de operadores com vista ao alcance de um estado ou solução desejados.

Prova verbal - Situações Quantitativas: é composta por um conjunto de problemas com que a criança pode se deparar no quotidiano, pressupondo competências que vão desde a pré-contagem e contagem à lógica simples.

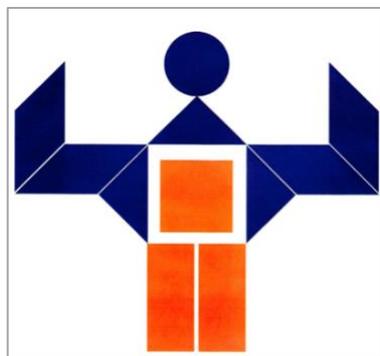


Figura 9 - Exemplo de tarefa da prova não verbal Construção de Padrões.

Prova não verbal - Construção de Padrões: requer competências de integração visório motora, de síntese de informação parte-todo, de orientação espacial, de lateralização e sequencialização de elementos e, ainda, de coordenação motora fina em face da natureza manipulativa da tarefa. O grau de dificuldade aparece relacionado com o número e o tipo de peças constituintes de cada desenho, bem como com a posição que cada uma delas ocupa no todo.

2.3.6. Pensamento divergente

Associando pensamento divergente à criatividade podemos considerar que, na psicologia, é longa a teorização em torno deste construto necessário à adaptação dos indivíduos às situações, problemas e mudanças. As situações e os objetos podem ser vistos como possuidores de funções e utilidades diferentes daquelas a que habitualmente estão associados, estabelecendo-se propriedades e relações novas que permitem ultrapassar as relações lógicas e dedutivas (Tarrida, 2002). Assim, a possibilidade de responder às exigências do meio passaria pelo recurso à lógica e ao estabelecimento de deduções entre os elementos disponíveis e pelo recurso a soluções várias, através de critérios pouco restritivos, valorizando-se a variedade e quantidade de soluções encontradas.

A produção divergente foi frequentemente operacionalizada através dos fatores cognitivos de fluência, flexibilidade, originalidade e elaboração. O primeiro se relaciona com a produção quantitativa, sendo certo que, quanto maior for o número de ideias produzidas, maior será a possibilidade de se encontrar a solução adequada (Wechsler, 2002). O fator de flexibilidade exige a capacidade para abordar e interpretar coisas e situações de uma forma nova, requerendo-se variabilidade de significado, interpretação e uso (Sandoval, 1993). O terceiro fator da produção divergente é a originalidade, ou seja, a capacidade de produzir respostas raras.

O pensamento divergente é pela primeira vez incluído na ECCOs, assumindo-se, à luz do modelo de Guilford, a necessidade de abordar uma produção divergente como complemento ou dimensão da cognição.

Prova verbal - Construção de Histórias: lê-se uma pequena história à criança para a qual lhe será pedido um ou vários fins.

Prova não verbal - Construção de Figuras: procura-se que a criança crie um número ilimitado de imagens dentro de um tempo limite, recorrendo a seis pequenas barras de madeira.



Figura 10 - Exemplo de tarefa da prova Construção de Figuras - com 6 barrinhas de madeira, construir todas as figuras que quiser dentro de um tempo limite.

Capítulo III

Estudo 1 - Validação e adaptação da ECCOs 4/10 para o Brasil

3. ESTUDO 1 – VALIDAÇÃO E ADAPTAÇÃO DA ECCOs 4/10 PARA O BRASIL

3.1. Introdução

Em 2009, o *Instituto de Educação e Psicologia da Universidade do Minho - UMinho*, de Portugal, liberou uma escala para avaliação da inteligência em crianças denominada *Escala de Avaliação de Competências Cognitivas para Crianças dos 4 aos 10 Anos de Idade – ECCOs 4/10*, produto do trabalho da Dra. Maria de Lurdes Dias Brito sob a orientação do Prof. Dr. Leandro Almeida.

Em função do intercâmbio daquele instituto de pesquisa com o Programa de Pós-Graduação em Psicologia Cognitiva da UFPE, em seguida à liberação da escala se deu o início de um trabalho para validação e adaptação da ECCOs 4/10 para a população brasileira. Esse trabalho foi conduzido pelo Prof. Dr. Antonio Roazzi (UFPE) e contou com a participação, principalmente, da Dra. Luciana Hodges (UFPE), Dra. Rafaella Asfora (UFPE) e Dra. Maira Roazzi (EAR).

Importante salientar que este estudo referente à validação e adaptação da ECCOs 4/10 no modelo lápis e papel para o Brasil não foi foco do trabalho realizado durante esta tese, estando descrito neste documento apenas no intuito de registrar a história da escala portuguesa no Brasil e ajudar a compreender a dimensão do projeto como um todo. Por outro lado, também ressaltamos que os dados coletados durante o estudo com a prova lápis e papel serviram para promover conceitos e análises em confronto com os dados dos demais estudos.

3.2. Recapitulando a ECCOs 4/10

Como vimos no capítulo anterior, a ECCOs 4/10 tem como referência os modelos compósitos de inteligência, avaliando o desempenho intelectual através da articulação de seis processos cognitivos organizados em um nível crescente de complexidade utilizando tarefas

verbais e tarefas não verbais. Os recursos das provas compõem-se de materiais lúdicos, coloridos, e procurou-se utilizar itens que remontam a figuras ou situações que fazem parte do cotidiano da criança. A tipologia das provas se mantém nas várias idades. Ao longo de cada prova os itens vão ficando mais complexos em dificuldade ou exigência de cotação.

Cada prova dispõe de critérios para determinar o item inicial a ser aplicado, tipicamente definido em função da idade da criança, e critérios de parada da aplicação do teste, normalmente em função de sucessivas falhas ou quando a criança demonstra desatenção ou cansaço. Não há ordem rigorosa para a apresentação das tarefas, podendo ser adaptada para cada criança, mas sugere-se que as provas verbais e não verbais sejam aplicadas alternadamente e que os mesmos processos cognitivos não sejam invocados seguidamente. No início de cada prova existem itens de exemplo com o propósito de assegurar o entendimento do que está sendo pedido por parte da criança.

3.3. Justificativas Validação e adaptação da ECCOs 4/10 para o Brasil

A relevância do trabalho pode ser claramente justificada devido à falta de instrumentos com suporte teórico claro para a avaliação das capacidades cognitivas na infância. Segundo Brito (2009), a necessidade de novos instrumentos de avaliação da inteligência para as faixas etárias mais baixas continua a ser uma realidade, pois são poucos os testes para avaliação da inteligência e da cognição destinados à infância disponíveis e atualizados para a língua portuguesa.

Embora existam outros instrumentos de avaliação cognitiva no Brasil, não se verifica instrumentos de avaliação que agreguem vários processos cognitivos e que avaliem um leque ampliado de faixa etária, sobretudo em crianças que estão na educação infantil, tal como apresentados na ECCOs 4/10. Estudos como este podem contribuir para suprir uma lacuna na área de avaliação psicológica e educacional.

3.4. Proposta do estudo

Tratou-se de uma proposta de estudo realizado na região Nordeste do Brasil (cidade do Recife, no estado de Pernambuco) sobre as características psicométricas da ECCOs 4/10, a propósito da adaptação e da validação dessa escala, onde se procurou:

1. Avaliar as relações entre o gênero, idade, e tipo de escola frente ao desempenho na escala de avaliação cognitiva.
2. Estabelecer relações de similaridade/dissimilaridade nas habilidades cognitivas avaliadas entre os dados coletados no Brasil e aqueles coletados em Portugal e Moçambique.
3. Chegar a um instrumento que possibilitasse comparações e usos em diferentes países de língua portuguesa.

3.5. Procedimento

O procedimento consistiu em fazer um estudo piloto e a adaptação do material para sua adequação à população brasileira.

A proposta consistia na aplicação individual de todas as onze provas da ECCOs 4/10 em cerca de 560 crianças, equitativamente distribuídas por sexo, tipo de escola (pública e particular) e 14 faixas etárias compreendidas entre os 4 e os 10 anos de acordo com a quantidade de meses de vida da criança. Ao final, foram entrevistadas 606 crianças.

A população-alvo para a pesquisa de adaptação consistiu de escolares entre 4 anos e 0 meses até 10 anos e 11 meses de idade residentes na cidade do Recife.

3.6. Descrição da amostra brasileira

Para determinar o nível socioeconômico dos alunos, usou-se como critério o tipo de escola (pública ou particular). Esse foi o mesmo critério usado na padronização brasileira dos testes Matrizes Progressivas Coloridas de Raven e WISC – III. Justifica-se esse critério no fato de que, em geral, crianças brasileiras de nível socioeconômico médio a alto frequentam escolas particulares, enquanto as de nível socioeconômico baixo frequentam escolas públicas.

Outros eventuais critérios para determinação do nível socioeconômico foram desprezados para evitar distorções nas informações e consequente falhas na análise dos dados. No caso da renda familiar, por exemplo, sabe-se que há uma tendência a distorcer esse dado ou não revelá-lo. Já quanto a utilizar o nível educacional e ocupacional dos pais para procurar estabelecer o nível socioeconômico, além de não existirem estatísticas oficiais brasileiras sobre isso, muitos profissionais aceitam empregos com nível inferior à sua formação.

Os quadros seguintes resumem a relação entre as idades e as 14 faixas etárias que foram utilizadas no estudo em Portugal e no Brasil.

Quadro 2 - Distribuição das faixas etárias na amostra portuguesa.

Grupo etário	Grupo normativo	Faixa etária
4	4 anos	4 anos aos 4 anos 1 mês
	4 ½ anos	4 nos 5 meses aos 4 anos 7 mês
5	5 anos	4 anos 11 meses aos 5 anos 1 mês
	5 ½ anos	5 anos 5 meses aos 5 anos 1 mês
6	6 anos	5 anos 11 meses aos 6 anos 1 mês
	6 ½ anos	6 anos 5 meses aos 6 anos 7 mês
7	7 anos	6 anos 11 meses aos 7 anos 1 mês
	7 ½ anos	7 anos 5 meses aos 7 anos 7 meses
8	8 anos	7 anos 11 meses aos 8 anos 1 mês
	8 ½ anos	8 anos 5 meses aos 8 anos 7 meses
9	9 anos	8 anos 11 meses aos 9 anos 1 mês
	9 ½ anos	9 anos 5 meses aos 9 anos 7 meses
10	10 anos	9 anos 11 meses aos 10 anos 1 mês
	10 ½ anos	10 anos 5 meses aos 10 anos 7 meses

Quadro 3 - Distribuição das faixas etárias na amostra brasileira.

Grupo etário	Grupo normativo	Faixa etária
4	4 anos	4 anos aos 4 anos 5 meses 29 dias
	4 ½ anos	4 anos 6 meses aos 4 anos 11 meses 29 dias
5	5 anos	5 anos aos 5 anos 5 meses 29 dias
	5 ½ anos	5 anos 6 meses aos 5 anos 11 meses 29 dias
6	6 anos	6 anos aos 6 anos 5 meses 29 dias
	6 ½ anos	6 anos 6 meses aos 6 anos 11 meses 29 dias
7	7 anos	7 anos aos 7 anos 5 meses 29 dias
	7 ½ anos	7 anos 6 meses aos 7 anos 11 meses 29 dias
8	8 anos	8 anos aos 8 anos 5 meses 29 dias
	8 ½ anos	8 anos 6 meses aos 8 anos 11 meses 29 dias
9	9 anos	9 anos aos 9 anos 5 meses 29 dias
	9 ½ anos	9 anos 6 meses aos 9 anos 11 meses 29 dias
10	10 anos	10 anos aos 10 anos 5 meses 29 dias
	10 ½ anos	10 anos 6 meses aos 10 anos 11 meses 29 dias

Como se pode notar, em Portugal foram entrevistadas crianças com até 10 anos e 7 meses, enquanto no Brasil foram entrevistadas crianças até 10 anos 11 meses. Essa ampliação da faixa etária no caso do estudo brasileiro foi necessária devido à dificuldade das escolas e dos pais aceitarem a participação nas pesquisas.

A distribuição da amostra brasileira (n=606) e portuguesa (n=539) foram resumidas nas tabelas a seguir:

Tabela 1 - Distribuição amostra brasileira (n=606) por faixa etária, sexo e tipo de escola.

Faixa etária	Sexo masculino, escola pública (n=153)	Sexo feminino, escola pública (n=152)	Sexo masculino, escola particular (n=152)	Sexo feminino, escola particular (n=148)
4 anos	10	9	10	10
4a 6m	10	10	10	10
5 anos	11	11	11	10
5a 6m	12	13	11	10
6 anos	10	10	11	11
6a 6m	11	11	11	12
7 anos	13	11	10	12
7a 6m	12	12	10	8
8 anos	10	12	12	10
8a 6m	12	12	11	11
9 anos	11	11	12	10
9a 6m	10	10	13	10
10 anos	11	11	10	13
10a 6m	10	10	10	11

Tabela 2 - Amostras brasileira e portuguesa por idade, sexo e tipo de escola.

Amostra Pesquisada (Brasil N=606 / Portugal N=539)								
Grupo	4 anos	5 anos	6 anos	7 anos	8 anos	9 anos	10 anos	Total
Pública Brasil	39	47	42	48	47	42	42	307
Particular Brasil	40	42	45	40	43	45	44	299
Portugal	77	77	77	77	77	77	77	539
Total	156	166	164	165	167	164	163	1145
Masculino								
Pública Brasil	20	24	21	25	23	21	21	155
Particular Brasil	20	22	22	20	22	24	20	150
Portugal	39	39	38	39	38	38	38	269
Feminino								
Pública Brasil	19	23	21	23	24	21	21	152
Particular Brasil	20	20	23	20	21	21	24	149
Portugal	38	38	39	38	39	39	39	270

3.7. Exemplos de adaptações de itens para uso no Brasil

Durante a aplicação dos testes foram sendo observadas a inadequação de alguns itens para uso com a população brasileira. Tais diferenças se reportam a questões culturais ou às representações cotidianas portuguesas que não se repetem no Nordeste brasileiro. As principais ocorrências e as soluções que foram dadas estão relatadas em seguida.

No caso da prova Desenhos Absurdos, destinada à avaliação da compreensão através da percepção por parte da criança de situações improváveis ou impossíveis, o fato de uma árvore frondosa, como uma mangueira, apresentar cachos de uva não causou estranheza às crianças, haja vista a maioria das crianças não conhecerem a parreira, trepadeira que produz a uva, comum em Portugal. A solução que foi dada consistiu em substituir as árvores e frutos por espécies disponíveis na região (palmeiras e maçãs) como ilustrado na Figura 11.

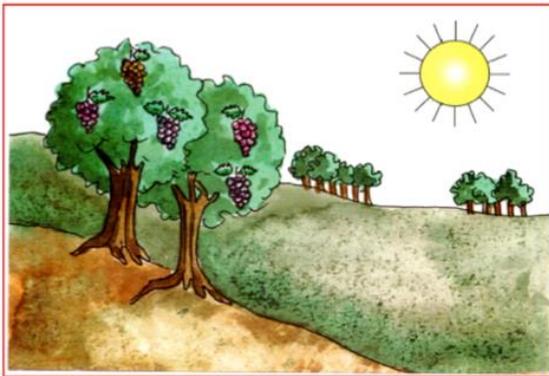
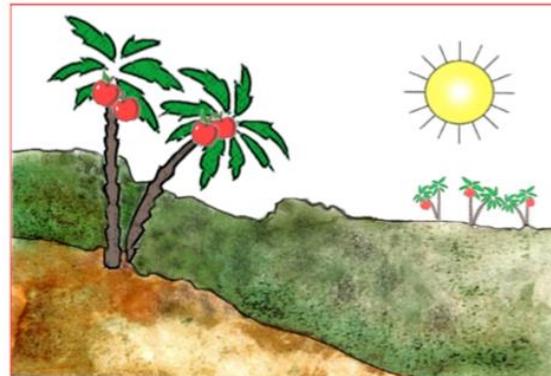
Portugal**Brasil**

Figura 11 - Adaptação de item na prova Desenhos Absurdos. As crianças não percebiam nenhum absurdo na figura utilizada em Portugal.

Na prova Situações Quantitativas, que faz parte da avaliação do processo cognitivo de resolução de problemas, a situação que teve que ser adaptada se deveu a diferenças entre os vocabulários do Brasil e de Portugal. A palavra "meta", utilizada lá, não era compreendida pelas crianças brasileiras e foi substituída por "chegada".

Portugal**Brasil**

Figura 12 - Adaptação de item na prova Situações quantitativas. As crianças não compreendiam o significado da palavra meta.

A prova Elementos em Frases, de memória, também requereu um forte trabalho para ser adaptada para uso no Brasil. Trata-se de uma prova que requer a memória auditiva, na qual se procura avaliar as capacidades de atenção, retenção e evocação de elementos simples

a partir de pequenos fatos relacionados com as vivências da criança e através de termos que se encontram no quadro de um contexto significativo. A questão é que as palavras precisavam ser dissílabas e fazer completo sentido para as crianças. Algumas tentativas foram analisadas até se chegar na situação final:

- 1a. tentativa: nomes de cidades pernambucanas: "Nas férias Antônio visitou: Pombos, Gravatá, Bezerros, Caruaru, Pesqueira, Arcoverde e Petrolina."

O problema aqui foi que, além das palavras não serem dissílabas, crianças de outros estados desconheceriam essas cidades.

- 2a. Tentativa: nomes de cidades do mundo. "Nas férias Antônio visitou: Paris, Roma, Lisboa, Belém, Londres, Tóquio e Madri."

O problema aqui foi que, muitas crianças, principalmente as de baixa renda, não reconheciam os nomes dessas cidades e memorizavam errado. Por exemplo: "Visitou Maris (referindo-se a Paris), Real (referindo-se ao time Real Madri), Boa (Lisboa), tem um que começa com T".

- 3a. Tentativa: cidades brasileiras.

Problema: não foi possível encontrar nomes de sete cidades brasileiras que fossem, ao mesmo tempo, conhecidas e dissílabas.

A solução escolhida acabou por inserir nome de lugares ao invés de cidades: "Nas férias Antônio visitou: praias, circos, parques, clubes, jardins, lojas, festas".

3.8. Análise dos dados

A partir de uma visão metodológica sistêmica optou-se por uma análise quantitativa de tipo estrutural que faz uso de análises multidimensionais (MDS) que respeitam a integridade dos dados coletados, especialmente sua dimensão qualitativa e relacional em todo o processo investigativo, do delineamento de pesquisa à coleta e interpretação dos dados (ver Guttman, 1968, 1991; Roazzi, 1995; Roazzi & Dias, 2001).

Os dados foram submetidos a uma análise multidimensional não-métrica do tipo SSA (Análise da Estrutura de Similaridade, Guttman, 1968; 1991) complementada com o método das “variáveis externas enquanto pontos” (Cohen & Amar, 1999; Roazzi & Dias, 2001), para descrição rigorosa das relações empíricas entre diferentes dimensões da escala ECCOs. A análise SSA faz parte da família de técnicas de escalagem multidimensionais (MDS – ‘Multidimensional Scaling’, Guttman, 1968), as quais permitem a partir de julgamentos de similaridade converter distâncias e similaridades de natureza psicológica em distâncias euclidianas, o que permite um julgamento analítico entre estruturas mentais complexas através de representações geométricas. No nosso caso as provas do ECCOs que permitem uma análise do desempenho cognitivo da criança tomando-se processos cognitivos e conteúdos. O método das “variáveis externas enquanto pontos” permite a localização nessas representações geométricas de variáveis ou subpopulações de interesse da pesquisa - idade, sexo e país, verificando sua dinâmica de estruturação no fenômeno sob investigação.

Na projeção demonstrada na Figura 13 observa-se claramente a diferenciação da dimensão verbal da dimensão não-verbal em uma estrutura modular – provas verbais no centro e não-verbais na periferia. Isto significa que as tarefas verbais apresentam correlações bem mais altas entre si e estruturalmente têm um maior peso por se localizar na área central, enquanto as não verbais estão distribuídas em regiões periféricas.

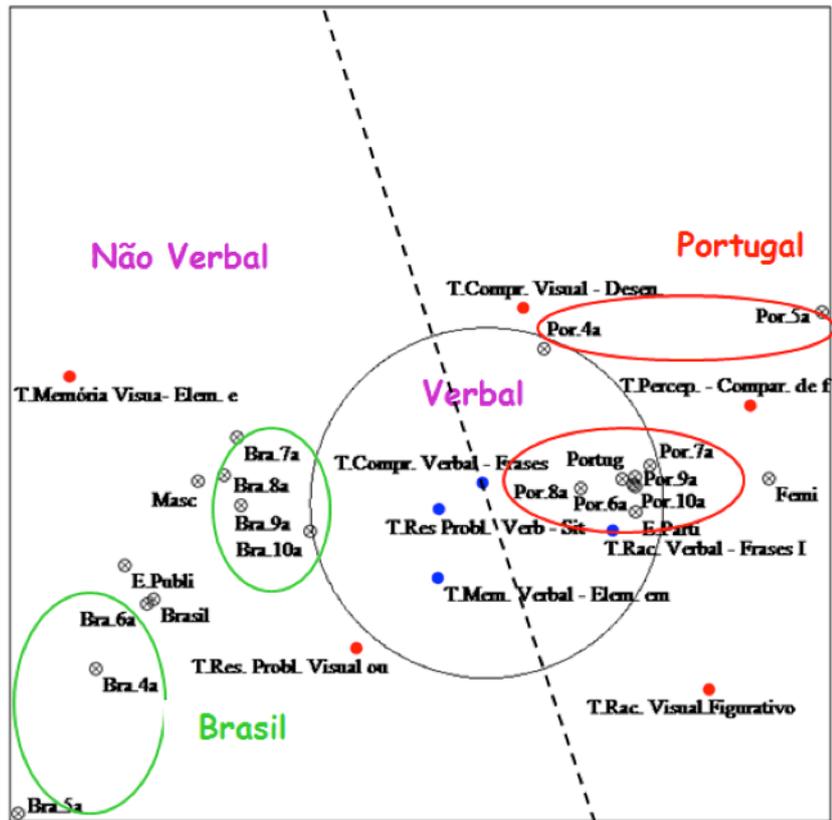


Figura 13 - Correlações entre tarefas, países, idade e dimensão verbal / não verbal.

Capítulo IV

Estudo 2 - Prova Figuras Incompletas informatizada com TRI

4. ESTUDO 2 – PROVA FIGURAS INCOMPLETAS INFORMATIZADA COM TRI

4.1. Introdução

Este estudo, ora apresentado, para a construção e validação da versão informatizada da prova Figuras Incompletas da ECCOs 4/10 foi realizado entre 2010 e 2011, estando apresentado na dissertação de mestrado deste autor. Durante este novo trabalho de tese o estudo foi ampliado, utilizando-se o conhecimento incorporado, uma parte do programa de computador, os dados e os resultados obtidos, assim como a análise realizada pela ótica da TRI. Todos esses pontos serviram de base para a ampliação dos testes informatizados da escala e, principalmente, para o estudo sobre construção de testes adaptativos computadorizados descrito no capítulo VI que deu continuidade e estendeu esse trabalho de informatização das provas da ECCOs 4/10.

Assim, o objetivo deste estudo foi construir à luz da teoria da resposta ao item (TRI) uma versão informatizada da prova Figuras Incompletas, uma prova para avaliação do raciocínio abstrato integrante da *Escala de Avaliação de Competências Cognitivas para Crianças dos 4 aos 10 Anos de Idade* (ECCOs 4/10). Além de fomentar o uso e domínio da TRI como método no desenvolvimento de testes e escalas, importou validar os resultados da versão informatizada em comparação com a prova original, procurando identificar se as características psicométricas foram mantidas e se a utilização da versão para computador se traduziu em benefícios para as crianças e psicólogos, detectando eventuais problemas e limites processuais ou tecnológicos. O estudo constatou vantagens na aplicação de testes informatizados e um maior engajamento das crianças. A utilização da TRI também contribuiu para ampliar o entendimento sobre a estrutura da prova e sobre os alunos, inclusive demonstrando onde se pode intervir na prova para aprimorar o equilíbrio de dificuldade da escala.

Neste capítulo descrevemos o processo de construção e validação da prova informatizada, incluindo a comparação entre os resultados das provas tradicional e informatizada, as vantagens e desvantagens de utilizar uma prova eletrônica, os problemas e facilidades encontrados, bem como os desafios e riscos que puderam ser percebidos na construção e aplicação desse instrumento. Visando conduzir o leitor através do roteiro adotado para a produção dos resultados descritos, este capítulo foi estruturado em duas partes. Na primeira parte, resgatamos um pouco do referencial teórico que orientou o estudo, incluindo a prova de raciocínio abstrato que foi alvo desse trabalho de informatização e validação. A segunda parte relata o estudo empírico, desde a preparação para a pesquisa até a aplicação da prova, seguindo-se das conclusões e das considerações finais e recomendações. Ao final deste documento estão anexadas algumas telas da prova Figuras Incompletas informatizada, a estrutura de dados modelada para o teste, as curvas e estatísticas emitidas pelos programas SPSS e Winsteps, o termo de consentimento dos responsáveis pelos participantes, os procedimentos para aplicação do teste e a ficha de informação acadêmica dos alunos.

4.2. Resgatando um pouco do referencial teórico

4.2.1. A medida em psicologia

Como vimos no capítulo I, no âmbito das ciências sociais e comportamentais, a psicometria tem se destacado como uma especialização em benefício da medição de fenômenos psicológicos e sociais (DeVellis, 2003). Mais do que apenas assinalar números para objetos ou eventos de acordo com regras preestabelecidas, a medição é um processo pelo qual nós tentamos compreender a natureza da variável aplicando técnicas matemáticas (Bridgman, 1928). Uma variável representa qualquer coisa que possa ser medida e que possa diferir entre entidades ou através do tempo. Em psicologia, normalmente, queremos medir comportamentos e processos psicológicos, variáveis que nem sempre permitem ser observadas, tais como inteligência, desajustamento ou ansiedade. As pessoas que estudamos variam ao longo dessas variáveis, apresentando escores que podem ser altos ou baixos, existir em maior ou menor quantidade ou diferenciarem-se segundo um outro conjunto similar de quantificadores.

As variáveis que não podem ser medidas diretamente pela observação são chamadas de variáveis (ou traços) latentes. Estas somente podem ser acessadas a partir de pressupostos ou inferências sobre dados oriundos dos comportamentos (Ayala, 2009). Ferramentas matemáticas, como a teoria da resposta ao item (TRI) ou a teoria clássica dos testes (TCT), podem ser aplicadas para explicar comportamentos observáveis a partir da perspectiva de variáveis latentes contínuas. Para a psicometria o traço latente representa um processo psicológico macro, não reducionista, compreendendo uma rotina de execução de tarefas que pode englobar diversas etapas (Pasquali, 2003). A princípio, qualquer traço latente reconhecível que tenha algum grau de estabilidade e que possa ser quantificado num *continuum* poderá ser medido.

Normalmente, o procedimento de medição em psicologia envolve testes onde se busca explicar o sentido que têm as respostas das pessoas a uma série de tarefas (os itens do teste). Na definição de Hogan (2006), teste representa um processo ou um instrumento padronizado que fornece informações sobre uma amostra de comportamento ou de processos cognitivos de maneira quantificada. As escalas utilizadas para medição em psicologia correspondem a processos onde são estabelecidas as correspondências entre os dados observados e a localização das pessoas na variável latente, permitindo compará-las com outras.

4.2.2. Figuras Incompletas - a prova de raciocínio não verbal da ECCOs 4/10

No capítulo II, vimos que a Escala de Avaliação de Competências Cognitivas para Crianças dos 4 aos 10 anos de Idade (ECCOs 4/10) é uma bateria extensa de testes voltada para avaliação cognitiva na infância, particularmente nos anos que respeitam a fase pré-escolar e o primeiro ciclo do ensino básico. Subjacente ao modelo de avaliação apresentado na ECCOs 4/10 está a ideia de uma inteligência composta, abordando desde os processos cognitivos mais simples até os mais exigentes e para os quais são apresentadas tarefas que mantêm a sua tipologia ao longo das idades abarcadas, adequando o nível de exigência das suas tarefas ao longo dessas várias faixas etárias (Brito, 2009).

Detendo-nos no raciocínio, processo cognitivo que mais nos interessa no estudo apresentado neste capítulo, na ECCOs 4/10, buscou-se avaliar o raciocínio através de tarefas analógicas, um processo pelo qual são avaliados e gerados argumentos lógicos, podendo assumir as formas do raciocínio dedutivo, condicional, silogístico ou indutivo. Brito (2009) destaca que os testes de inteligência, dentre os quais o de raciocínio se insere, são uma fonte válida de coleta de informação e que pela comparação com grupos de referência se produzem análises entre o esperado e o encontrado, servindo para orientar intervenções. Os resultados da avaliação cognitiva também constituem um bom preditor do desempenho nas situações

escolares de aprendizagem e rendimento (Seabra-Santos, 2000) ou até da capacidade intelectual adulta (Bracken & Walker, 1997).

A ECCOs 4/10 abrange o raciocínio em termos dedutivo e indutivo, recorrendo a tarefas de analogias, que representam o tipo de tarefa mais utilizado na avaliação do raciocínio. Alguns itens são compostos por imagens de objetos pertencentes ao quotidiano da criança enquanto em outros as imagens são figuras geométricas sem significado figurativo. O nível de dificuldade dos itens vai crescendo e passando a exigir da criança o estabelecimento de relações progressivamente mais complexas. Brito (2009) condensou no quadro a seguir as transformações subjacentes aos itens da prova de Imagens Incompletas.

Quadro 4 - Transformações subjacentes aos itens da prova de Imagens Incompletas.

Uma transformação	Itens	Duas transformações	Itens	Três transformações	Itens
Tamanho	1; 2; 13	Inserção/Forma	18	Tamanho/Forma/Inserção	34
Quantidade	3; 4	Inserção/ Número	19	Tamanho/Forma/Número	35
Cor	5; 6; 17	Cor/Forma	20	Tamanho/Número/Cor	36
Forma	7; 8; 16	Cor/Rotação	22; 30	Tamanho/Número/Forma	28
Rotação	9; 10; 15	Cor/Número	23; 26	Tamanho/Cor/Rotação	41
Inserção	11; 12; 14	Cor/Inserção	39	Forma/Cor/Número	25
		Tamanho/Rotação	21; 38	Rotação/Forma/Inserção	37
		Tamanho/Inserção	33	Rotação/Cor/Inserção	42
		Rotação/Número	32	Rotação/Cor/Número	29
		Rotação/Inserção	31	Inserção/Cor/Número	24;40
		Rotação/Forma	27	Inserção/Número/Tamanho	43

Para compreender a lógica implícita na prova de raciocínio vamos utilizar a descrição de Brito (2009) sobre a prova Figuras Incompletas: "numa tarefa analógica, pretende-se a análise e comparação de três termos apresentados (A, B e C) e a procura de um quarto termo (D) com o qual se completará corretamente a analogia". Segundo Sternberg (1992), esse tipo de tarefa implica nos seguintes processos:

- (i) codificação ou a descoberta de atributos, para os quais é necessário evocar a memória de longo prazo, atributos relevantes dos termos em causa que permitam, no caso dos itens verbais, a recolha de aspectos semânticos ou, no caso de itens figurativos, de formas, propriedades específicas ou posições;

- (ii) comparação de atributos (inferência, mapeamento e aplicação). A inferência permitirá relacionar A com B e descobrir aspectos que possam interligá-los. Em seguida, um processo semelhante operar-se-á entre os termos A e C (mapeamento ou correspondência), possibilitando-se, então, que a regra inferida de A-B seja aplicada nas características de C de modo a encontrar o termo D para completar a analogia;
- (iii) avaliação, que pode passar por processos de justificação em situações de maior complexidade do item quando surgir mais do que uma alternativa aceitável para o termo D ou não houver nenhuma que satisfaça plenamente a relação, recorrendo-se então ao processo de justificação e avaliação das alternativas disponíveis;

4.2.3. A teoria da resposta ao item

Como abordado no capítulo I, a teoria da resposta ao item (TRI) constitui o mais moderno e atual paradigma para a montagem de testes e interpretação dos dados coletados e desde meados dos anos 1980 vem se tornando a técnica predominante no campo da testagem (Pasquali, 2003). A TRI supera algumas limitações da teoria clássica dos testes (TCT), padrão vigente para construção de testes até então, reunindo modelos estatísticos para fazer predições, estimativas ou inferências sobre as habilidades (competências, aptidões) medidas em um teste que nos permite estabelecer correspondências entre variáveis latentes e suas manifestações (Ayala, 2009).

Para este estudo foi utilizado o modelo de Rasch de 1 parâmetro. Rasch desenvolveu um modelo matemático para construção de medidas baseado na relação probabilística entre a dificuldade de qualquer item e a habilidade de qualquer pessoa, sendo que a diferença entre essas duas medidas governa a probabilidade de qualquer pessoa ter sucesso num item em particular. Seu modelo incorpora um método para ordenar pessoas de acordo com suas habilidades e ordenar itens de acordo com seus níveis de dificuldade. Segundo Bond e Fox (2007), as análises baseadas no modelo de Rasch são sensíveis à aquisição ordenada das

habilidades investigadas, consegue estimar as distâncias entre os níveis de habilidade das pessoas e permite determinar se o padrão geral de desenvolvimento entre os itens e as pessoas é suficiente para explicar o padrão de desenvolvimento apresentado por cada item e cada pessoa, sendo particularmente orientadas para investigações nas ciências humanas.

4.3. O estudo empírico

4.3.1. Considerações iniciais

A prova Figuras Incompletas da ECCOs 4/10 é composta por 43 itens, conforme demonstrado no Anexo M, que são apresentados em ordem crescente de dificuldade. A versão informatizada da prova não modificou sua proposta original nem descartou nenhum item a priori. Manteve-se exatamente a mesma lógica, os mesmos itens originais e a mesma sequência de apresentação. No entanto, como o objetivo maior neste trabalho não era avaliar a capacidade cognitiva das crianças, mas verificar a viabilidade e os impactos de uma prova informatizada, não seguimos à risca os critérios estabelecidos no manual da ECCOs 4/10 para finalização da prova ou determinação do item inicial em função da idade do aluno, optando por submeter as crianças de todas as idades a todos os itens da prova enquanto estas demonstrassem interesse em continuar. Visto que as crianças poderiam se comportar de maneira diferente diante de um teste no computador, concluímos que todos os itens presentes na prova com papel e lápis deveriam ser revalidados pela TRI para a versão informatizada.

4.3.2. Objetivos

Este estudo teve por objetivo construir à luz da TRI uma versão informatizada da prova de avaliação do raciocínio não verbal da ECCOs 4/10, denominada Figuras Incompletas, importando validar seus resultados em comparação com a prova original. Adicionalmente, também desejou-se:

- Detectar eventuais problemas e limites processuais ou tecnológicos.
- Investigar se a utilização do computador favoreceria a aplicação do teste, seja pelo fato da substituição do papel, seja pelo aumento do caráter lúdico do teste permitido pela ferramenta computacional. Seria o teste aplicado através do computador mais prazeroso ou mais estimulante para o aluno?
- Observar se a aplicação da prova no computador aumentaria a atenção e concentração na tarefa por parte do aluno, haja vista que a criança passaria a ser um elemento ativo no teste, inclusive com controle da passagem dos itens.
- Verificar se a prova informatizada poderia ser aplicada em grupos, visto não carecer de tanta assistência do psicólogo quanto a prova tradicional.
- Em relação à teoria da resposta ao item (TRI) como conceito e ferramenta de suporte e análise para a definição do teste, nosso interesse foi compreender como montar provas com melhor distribuição dos itens, como predizer a chance da criança ser bem sucedida diante de um item e identificar se a prova concebida no conceito da TCT estaria bem estruturada na ótica da TRI.

4.3.3. Método

Participantes

Participaram da pesquisa 304 estudantes de duas escolas particulares para classe média dos municípios de Recife e Jaboatão dos Guararapes, em Pernambuco. A distribuição por idade e sexo do grupo está representada na tabela a seguir.

Tabela 3 – Classificação dos participantes por idade e sexo.

Classificação por sexo	Total de Alunos	Classificação por Idade						
		4 anos	5 anos	6 anos	7 anos	8 anos	9 anos	10 anos
Masculino	151	11	19	22	18	25	26	30
Feminino	153	11	15	13	24	29	34	27
Total:	304	22	34	35	42	54	60	57

Os alunos foram convidados a participar da pesquisa através do termo reproduzido no Anexo E, em conformidade com a resolução CNS 196/96 e suas complementares do Conselho Nacional de Saúde.

Instrumento

Para apresentação dos itens, coleta de dados e pontuação foi desenvolvido um software construído em ActionScript, uma linguagem de programação orientada a objetos integrada ao Adobe Flash, baseada em scripts e padronizada pela *European Computer Manufacturers Association*.

Procedimentos

Foram utilizados computadores com telas sensíveis ao toque para aplicação dos testes, que ocorreu dentro do ambiente escolar.

Os dados das respostas aos testes foram complementados por informações fornecidas pelas escolas sobre aspectos socioeconômicos, acadêmicos e de eventuais comprometimentos físicos, sensoriais, psicológicos ou mentais da criança. Esses dados complementares também foram inseridos diretamente no programa de teste. Para a análise dos dados foram utilizados os programas SPSS versão 18 e Winsteps Rasch Measurement versão 3.70.1.1, este destinado exclusivamente a análises de aplicações baseadas na TRI e utilizando o modelo de Rasch.

Um estudo piloto com 63 estudantes das mesmas escolas foi realizado com objetivo de identificar a unidimensionalidade do teste, a distribuição das dificuldades dos itens na escala e eventuais necessidades de ajustes no programa ou na forma de sua aplicação.

4.3.4. Resultados

Em uma análise preliminar decorrente do estudo piloto, percebeu-se que alguns itens tinham ordem de dificuldade diferente daquelas indicadas no teste original. O item 10 mostrou-se inadequado e foi retirado do teste. Além disso, algumas figuras foram retrabalhadas para melhor identificação por parte das crianças. As crianças demonstraram boa aceitação do teste informatizado, normalmente se referindo a ele como jogo.

Após os ajustes, o teste foi aplicado para os 304 alunos participantes e as habilidades das pessoas e as dificuldades dos itens foram convertidas para uma mesma escala (logit). Por definição do método de análise pela TRI, para evitar interferências nas medidas, foi necessário excluir da amostra 32 participantes que poderiam perturbar a análise do teste como um todo caso fossem mantidos, haja vista apresentarem valores altos para outfit, um sinal de inconsistência no padrão das respostas. Dessa forma, prosseguimos com as análises passando a considerar apenas 272 alunos e 42 itens. O mapa de distribuição de pessoas e itens está representado na Figura 14, exibida adiante.

algumas trocas de posicionamento da dificuldade dos itens no *continuum*, tal como o i12 que apresentou índice de dificuldade inferior à i03.

As principais estatísticas globais sobre itens e pessoas do teste, apresentadas nas tabelas 4 e 5 demonstram um teste equilibrado e consistente, com índices de fidedignidade elevados, outfits e infits médios próximos ao ideal de 1.0, medidas em torno de zero e erro padrão baixo. O desvio padrão elevado já era esperado, tanto pelo objetivo da prova de avaliar um construto complexo quanto pelo fato desta estar orientada para crianças de 4 a 10 anos, portanto, em fases diferentes de maturidade. Os índices de separação e confiabilidade bem acima de 2.0 indicam que os dados são bem definidos e que a confiabilidade para posicioná-los ao longo da escala é boa.

Tabela 4 – Sumário da medida de 272 pessoas.

SUMÁRIO DA MEDIDA DE 272 PESSOAS								
	ESCORE TOTAL	CONTAGEM TOTAL	MEDIDA	ERRO MODELO	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MÉDIA	20.5	42.0	.01	.41	1.00	.0	.95	.0
DES. PD.	7.6	.0	1.28	.06	.23	1.1	.38	.8
MÁXIMO	39.0	42.0	3.39	.82	1.74	3.3	1.99	2.1
MÍNIMO	2.0	42.0	-4.34	.38	.47	-2.7	.26	-2.0
REAL RMSE	.43	DES. VERD	1.21	SEPARAÇÃO	2.77	CONFIABILIDADE	.88	
MODELO RMSE	.42	DES. VERD	1.21	SEPARAÇÃO	2.92	CONFIABILIDADE	.89	
ERRO PADRÃO DA MÉDIA DA PESSOA = .08								
CORRELAÇÃO ENTRE ESCORE_TOTAL E MEDIDA DA PESSOA = .99								
ALFA DE CRONBACH (KR-20) ESCORE BRUTO FIDEDIGNIDADE DO "TESTE" = .89								

Tabela 5 - Sumário da medida de 42 itens.

SUMÁRIO DA MEDIDA DE 42 ITENS								
	ESCORE TOTAL	CONTAGEM TOTAL	MEDIDA	ERRO MODELO	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MÉDIA	132.8	272.0	.00	.17	1.00	-.1	.95	-.2
DES. PD.	69.2	.0	1.68	.05	.15	1.7	.26	1.3
MÁXIMO	265.0	272.0	2.41	.40	1.42	4.5	1.61	3.3
MÍNIMO	35.0	272.0	-4.48	.14	.77	-3.5	.35	-2.7
REAL RMSE	.18	DES. VERD	1.67	SEPARAÇÃO	9.30	CONFIABILIDADE	.99	
MODELO RMSE	.18	DES. VERD	1.67	SEPARAÇÃO	9.54	CONFIABILIDADE	.99	
ERRO PADRÃO DA MÉDIA DO ITEM = .26								
CORRELAÇÃO ENTRE ESCORE TOTAL E MEDIDA DO ITEM = -.99								
11424 PT DE DADOS. LOG-LIKELIHOOD QUI-QUADRADO: 10260.04 c/ 11111 d.f. p=1.0000								
RMSR (residual) Global (excluindo escores extremos): .3828								

A unidimensionalidade dos dados foi verificada com o método da análise dos principais componentes – APC. Sabemos que o propósito da APC de resíduos não é construir variáveis (como na análise fatorial), mas explicar a variância. Assim, primeiro de tudo devemos verificar se o contraste nos residuais é grande o suficiente para explicar muito da variância. Como o primeiro contraste ficou no nível de ruído (até 2 eigenvalues), podemos concluir que não há uma segunda dimensão e que podemos prosseguir com a análise.

Tabela 6 – Saída da Análise dos Principais Componentes (em eigenvalue).

Tabela de of variância residual padronizada (em eigenvalue)				
		-- Empírico --		Modelado
Total da variância bruta nas observações	=	71.6	100.0%	100.0%
Variância bruta explicada pelas medidas	=	29.6	41.4%	41.3%
Variância bruta explicada pelas pessoas	=	13.0	18.1%	18.1%
Variância bruta explicada pelos itens	=	16.7	23.3%	23.2%
Variância bruta inexplicada (total)	=	42.0	58.6%	58.7%
Variância inexplicada no primeiro contraste	=	2.0	2.8%	4.7%
Variância inexplicada no segundo contraste	=	1.9	2.7%	4.5%
Variância inexplicada no terceiro contraste	=	1.8	2.5%	4.3%
Variância inexplicada no quarto contraste	=	1.6	2.3%	3.9%
Variância inexplicada no quinto contraste	=	1.5	2.2%	3.7%

Corroborando com a visão da unidimensionalidade dos dados o fato de que simulações sobre bases de dados maiores, geradas com o Winsteps, indicam redução do eigenvalue no primeiro contraste e aumento da variância explicada pelas medidas. Por exemplo, para uma base simulada de 1.000 respostas, a variância inexplicada no primeiro fator caiu para 1.4 eigenvalues.

Quanto à fidedignidade e validade do teste, nenhum dos itens apresentou correlação point-biserial negativa, ressaltando a qualidade dos itens, conforme defendido por Linacre (2008). Apenas os itens i19 (point-biserial=0,12) e i41 (point-biserial=0,11) apresentaram correlação inferior aos 0,15 ao recomendado por Varma (2010).

A alta consistência do teste também foi confirmada pelos indicadores alfa de Cronbach com valor de 0,888 e os coeficientes Spearman-Brown com 0,816 e Guttman pelo método Split-half de 0,815 atribuídos pelo SPSS. Tão importante quanto os indicadores

globais de consistência é o “alfa de Cronbach se o item for excluído”, representando como ficaria a confiabilidade do teste na ausência de cada um dos itens. Mais uma vez, constatamos que a retirada de qualquer item não teria influência significativa na confiança do teste. A função de informação do teste junto com as curvas características dos itens e do teste, apresentadas em anexo, permitem uma visão gráfica complementar do teste.

Adicionalmente, verificamos que na amostra estudada não houve uma distinção clara da pontuação em função da idade do participante, diferença esta que parece só ser determinante a partir de 8 anos de idade. Na amostra estudada, a faixa etária de 6 anos teve um rendimento abaixo do esperado, como pode ser visualizado na figura a seguir.

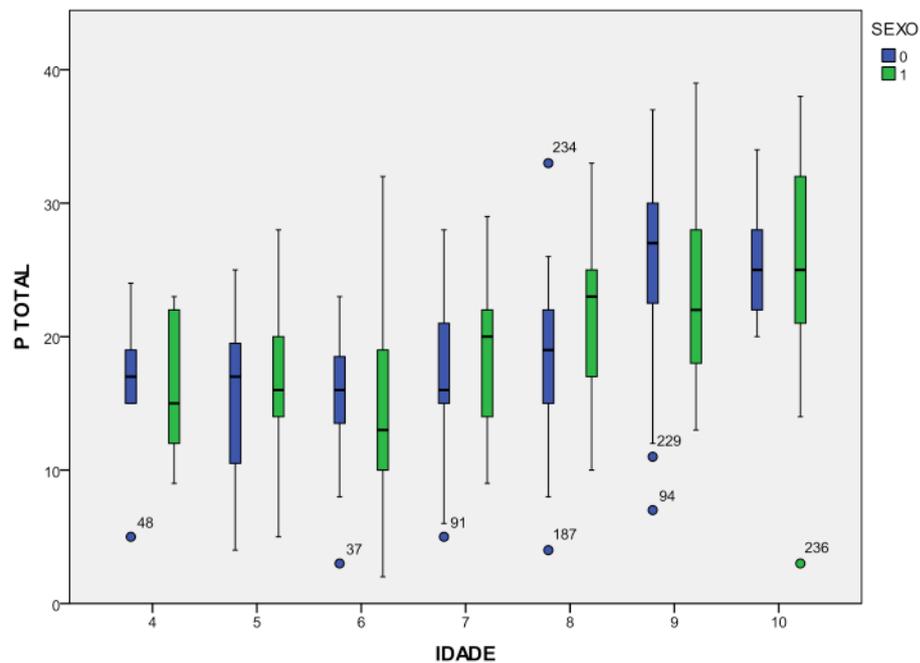


Figura 15 - Boxplot idade x pontuação total.

4.3.5. Conclusões

O uso de uma ferramenta computacional para aplicação de testes com crianças se mostrou um recurso atraente, simples e ágil. Os alunos não tiveram qualquer dificuldade em compreender e realizar o teste, haja vista que a maioria dos participantes já tinha boa ou, pelo menos, razoável proximidade com computadores. Além disso, o uso de tela sensível ao toque permitiu também às crianças menores utilizar o recurso computacional sem problema algum, mesmo aquelas (duas) que convivem com algum tipo de limitação motora.

Sem dúvida, uma grande vantagem em aplicar uma prova automatizada é que a base de dados já estará montada ao final da aplicação do teste, não requerendo a transposição dos dados registrados em papel para o computador, processo este que pode inserir erros nos dados, além de demandar tempo do pesquisador.

Também consideramos muito relevante a capacidade de aplicar a prova para grupos de alunos ao invés de apenas individualmente. Em dois momentos organizamos grupos de quatro alunos entre sete e dez anos e aplicamos o teste simultaneamente através de quatro computadores. Não constatamos nenhum tipo de prejuízo nessa modalidade de aplicação do teste, seja do ponto de vista da aplicação em si ou do desempenho dos alunos, o que indica que a aplicação do teste pode ser realizada de forma muito mais ágil do que no método tradicional. A ressalva fica para grupos na faixa etária mais baixa, de 4 e 5 anos, onde a necessidade de atendimento personalizado era bem mais evidente e optamos por não submetê-los a testes em grupos.

De negativo, percebemos uma incidência que julgamos relativamente alta de pessoas cujas medições apresentaram desajustes nos extremos (outliers), ou seja, pessoas com nível de habilidade alta errando itens com baixo nível de dificuldade ou pessoas com nível de habilidade baixa acertando itens com nível alto de dificuldade. Devido a esse fato, 32

participantes que fizeram o teste foram excluídos na primeira parte da análise, o que pode significar que diante de um teste no computador que se assemelha a um jogo de perguntas e respostas alguns alunos se sintam menos comprometidos e produzam respostas com a provisoriade que o computador os acostumou através de outras tarefas. Por outro lado, é preciso lembrar que todos os participantes foram submetidos a todas as questões, apenas sendo interrompidos em casos de desistências explícitas ou quando o pesquisador considerou necessário. O manual da prova original Figuras Incompletas, em papel, recomenda que para idades a partir de 6 anos, inclusive, a prova seja iniciada no item 13 (assinalando sucesso de 1 a 12) e que para qualquer idade a prova seja interrompida após quatro insucessos consecutivos. Essa prática adaptativa é saudável porque reduz a quantidade de itens aplicados aos alunos, ajudando a controlar os falso-positivos (os falso-positivos, quando detectados pelo processamento, são factíveis de eliminação). Estimamos que regras como essas, implantadas à luz do estudo aqui produzido, proporcionariam resultados mais conservadores, porém, caso utilizássemos essas regras, perderíamos parte da informação sobre os itens e estudantes, principalmente daqueles fora das expectativas mais conservadoras de inteligência, o que não era desejado.

Quanto aos itens, constatamos o primor com que foram elaborados durante o trabalho de produção da ECCOs 4/10. Ao aproveitá-los para a versão informatizada, apenas utilizando traços um pouco mais definidos, acabamos não tendo, praticamente, nenhum esforço para calibrá-los. A exceção ficou por conta do item i10, que acabou sendo excluído. Por outro lado, a existência de um item desajustado permitiu demonstrar a capacidade de softwares como o Winsteps identificarem esses desajustes e apresentá-los ao investigador para que sejam devidamente tratados.

A comparação das medidas obtidas com as versões em papel e eletrônica da prova Figuras Incompletas da ECCOs 4/10 apresentaram diversas diferenças, seja no sentido da

discriminação de acordo com a idade do aluno, na própria ordem de dificuldade dos itens ou nos escores por faixa etária. Isso não foi uma surpresa para nós, pois adotamos condutas diferentes para aplicação da prova, tal como submeter todos os itens a todos os alunos enquanto na ECCOs 4/10 há critérios que determinam o início e fim da prova de acordo com a idade do participante e com a quantidade de erros consecutivos. O simples fato da prova ser realizada no computador e controlada pelo aluno já modifica a relação deste com a prova. Além disso, a prova informatizada foi submetida à população brasileira enquanto a original foi trabalhada com a população portuguesa.

Para uma futura nova versão deste teste havemos de considerar a possibilidade de retirar o item i01 e, talvez, o i02, pois estes pouco serviram para discriminar as pessoas. Por outro lado, inserir itens com capacidade de aumentar a discriminação das habilidades superiores, principalmente, e das habilidades medianas, em torno de zero logits poderia deixar a escala melhor distribuída. Também seria oportuno modelar os pontos de entrada e saída no teste de acordo com as idades dos participantes.

Na verdade, como concluímos, o que acabamos fazendo não foi exatamente informatizar a prova Figuras Incompletas da ECCOs 4/10, mas, sim, aproveitar sua estrutura consistente de itens bem elaborados e testados para viabilizar a construção de uma prova eletrônica sob um novo ponto de vista, o da teoria da resposta ao item. Dessa forma, obtivemos uma escala abrangendo praticamente todo o *continuum* do construto sem ter tido a necessidade de rever a construção dos itens várias vezes, como o foi feito para a versão original, em papel, embora tenhamos detectado que a inclusão de alguns novos itens poderão melhorar a discriminação da escala em algumas localizações. Esse mesmo conceito pode ser estendido para outras provas.

Capítulo V

Estudo 3: informatização das provas sequenciais

5. ESTUDO 3: INFORMATIZAÇÃO DAS PROVAS SEQUENCIAIS

Esta parte do estudo pretendeu ampliar a oferta de provas informatizadas da ECCOs 4/10 no formato sequencial de apresentação dos itens, seguindo, sempre que possível, as mesmas regras utilizadas nas provas de lápis e papel. Foram informatizadas três provas: Comparação de Figuras, que avalia o processo psicológico da percepção, Elementos em Árvores, que avalia o processo de memória e Desenhos Absurdos, que avalia a percepção. Além dessas três provas, para compatibilizar a linguagem de programação de todas as provas visando integrá-las em um mesmo programa de computador, foi necessário fazer a migração para Java da prova de raciocínio Figuras Incompletas, descrita no capítulo IV e que já fora fruto de um estudo anterior.

Dessa forma, a ECCOs 4/10 passou a dispor de um total de quatro provas informatizadas no modelo sequencial. Adicionalmente, foi inserida uma tela de consulta gráfica das atividades realizadas pelas crianças, que pode ser vista no Anexo A, o que permitiu uma visualização integrada da performance dos participantes em cada prova em conjunto com os dados de identificação e indicadores escolares e socioeconômicos.

Devido às singularidades de cada prova, os critérios de aplicação e os resultados de cada uma delas estão descritos em tópicos específicos apresentados mais adiante.

5.1. Desenvolvimento dos programas de computador

Os programas de computador que foram construídos para as versões informatizadas das provas seguiram um padrão único, compreendendo os seguintes principais componentes: gráficos e ilustrações, banco de itens e banco de dados das coletas das respostas, além dos próprios códigos dos programas.

Considerou-se a existência de duas tarefas bem distintas no processo: a aplicação do teste e a coleta dos dados, que são realizadas juntas, sempre que possível, e a interpretação

dos dados coletados. Enquanto a aplicação e coleta constitui uma rotina, ou seja, uma tarefa regular, padrão e repetitiva, a interpretação dos dados requer bastante flexibilidade, pois depende do que o pesquisador estiver interessado em analisar a cada momento. Dessa forma, o instrumento que construímos no âmbito deste trabalho pode ser considerado uma ferramenta computadorizada para realizar a aplicação de testes e prover a coleta de dados. Já a tarefa de interpretação dos dados foi feita sempre com a ajuda de programas estatísticos consagrados para essa função, como SPSS e Winsteps, a partir dos dados exportados pela ferramenta de coleta de dados que foi criada.

Conforme nos disse Brito (2009), nos casos de avaliações da inteligência a informação recolhida através dos testes deve ser complementada por informações dos pais e dos professores. Dessa forma, conforme pode ser observado na estrutura de dados apresentada no Anexo K, foi idealizada uma base de dados que além de registrar as informações pertinentes à realização e desempenho no próprio teste, também contemplou campos para registro de informações adicionais que fossem relevantes para a interpretação dos resultados no tocante aos aspectos socioeconômicos, acadêmicos e de eventuais comprometimentos físicos, sensoriais, psicológicos ou mentais da criança. Esses campos são preenchidos através de informações fornecidas pelas escolas participantes.

5.1.1. Interface

Como ocorre em muitos programas de computador, havemos de considerar que o programa terá sempre dois perfis distintos de usuários: o da criança que estiver participando do teste e o do próprio aplicador do teste. Se para o aplicador a necessidade de preenchimento de um pequeno cadastro não é uma tarefa complexa, para as crianças, principalmente as de menor idade, o programa deve ser atraente, mas, ao mesmo tempo, robusto o suficiente para evitar que operações indesejadas ou erros espúrios ocorram, visto que as crianças estão

passíveis de produzir cliques aleatórios do mouse ou se antecipar às instruções e tentar interagir com o programa de alguma forma. Também devemos considerar que o manuseio do teclado ou do mouse por crianças a partir de quatro anos de idade poderia, por si só, representar um fator de inibição, limitação e mesmo inserção de erros.

Nesse sentido, concebemos uma interface limpa para realização dos testes, com poucos detalhes, privilegiando a utilização de botões e figuras grandes, conforme pode ser observado nos exemplos apresentados no Anexo A. Adicionalmente à simplicidade da interface, utilizamos monitores com telas sensíveis ao toque (touch screen) para a administração do teste, possibilitando às crianças interagirem com o instrumento simplesmente apontando com o dedo na tela a opção desejada. Na verdade, este novo programa ampliou e melhorou o instrumento produzido para informatizar a prova Figuras Incompletas, conforme descrito no capítulo IV.

5.1.2. Linguagem de programação

Desta vez, para o desenvolvimento dos programas foi utilizada a linguagem de programação Java e bibliotecas auxiliares, principalmente a interface JavaFX, uma API (application program interface) que compreende uma verdadeira plataforma para criação de aplicativos para internet sob a bandeira das tecnologias rich internet applications (RIA). A JavaFX expande o poder do Java, permitindo criar experiências visuais envolventes, através da integração de gráficos vetoriais, recursos Web de animação, áudio e vídeo em aplicações ricas, interativas e imersivas para diversos dispositivos e sistemas operacionais.

As imagens dos itens, ilustrações e artes dos componentes adicionais das telas dos programas foram desenhadas no Adobe Flash e exportadas como imagens nos formatos de arquivos PNG ou JPG, dependendo da sua utilização pelo programa.

5.1.3. Banco de Itens

Os itens foram armazenados separadamente dos dados decorrentes da realização dos testes. Para a construção do banco de itens, utilizou-se uma estrutura padronizada pela eXtensible Markup Language (XML), um formato para a criação de documentos com dados organizados de forma hierárquica, que mantém a separação entre conteúdo e formatação, apresenta alta portabilidade e é de fácil integração entre programas, linguagens de programação e sistemas operacionais. Um exemplo da estrutura XML para um item da prova Figuras Incompletas pode ser visto na Figura 16.

```
<?xml version="1.0" encoding="UTF-8"?>
<Eccos>
  <paginas>
    <pagina>
      <numeracao numero="1"/>
      <pergunta>
        <item>perguntas/01_N/B.png</item>
        <item>perguntas/01_N/D.png</item>
        <item>perguntas/01_N/E.png</item>
      </pergunta>
      <alternativas>
        <alt>perguntas/01_N/A.png</alt>
        <alt>perguntas/01_N/B.png</alt>
        <alt>perguntas/01_N/C.png</alt>
        <alt>perguntas/01_N/D.png</alt>
        <alt>perguntas/01_N/E.png</alt>
      </alternativas>
      <resposta letra="A" />
    </pagina>
  </paginas>
</Eccos>
```

Figura 16 - Código XML que descreve o item 1 da prova Figuras Incompletas. O bloco <pergunta> define os arquivos de imagem que devem ser apresentados, enquanto o bloco <alternativas> designam as imagens das respostas. A resposta correta é a opção A.

O banco de itens foi construído de forma fiel aos itens que fazem parte das respectivas provas da ECCOs 4/10 em lápis e papel, com suave reestilização, como pode ser percebido na Figura 17.

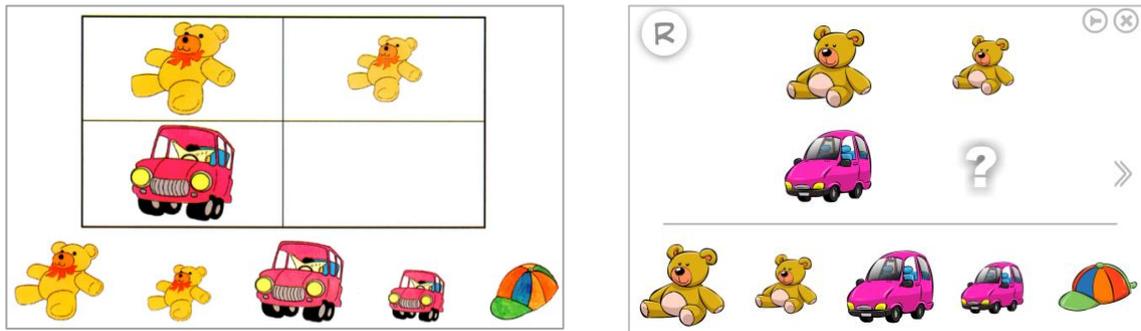


Figura 17 - Representação da diferença de traço nas ilustrações do teste com lápis e papel (à esquerda) e no teste informatizado (à direita) na prova Figuras Incompletas

5.1.4. Banco de dados

Para armazenamento dos dados cadastrais e informativos sobre as crianças, bem como os dados de respostas e realização dos testes (tempos, escores, etc.), foram utilizados bancos de dados padrão SQLite, uma biblioteca gratuita de código aberto escrita em linguagem C que implementa um banco de dados relacional do tipo Structured Query Language (SQL) que fica embutido no corpo dos próprios programas. Assim, os programas que usam a biblioteca SQLite têm acesso a um banco de dados no padrão SQL sem necessidade de utilizar um servidor de banco de dados em separado. Desta maneira, o SQLite não é uma biblioteca cliente usada para se conectar com um grande servidor de banco de dados, mas sim o próprio serviço de banco de dados. As estruturas dos bancos de dados estão demonstradas no Anexo K.

5.2. Versão informatizada sequencial da prova de memória Elementos em Árvore

Como vimos no capítulo II, o objetivo da prova não verbal Elementos em Árvores é avaliar a capacidade para reter e evocar elementos simples apresentados anteriormente à criança. Desta forma, a prova consiste em mostrar para a criança uma árvore com alguns elementos (enfeites). Após alguns segundos, os elementos desaparecerão e será solicitado à criança para indicar onde eles estavam. Procurou-se que os desenhos tivessem algum significado para as crianças, tornando-o agradável e lúdico, apresentando uma certa uniformidade de item para item. Todos os itens estão representados no Anexo L.

5.2.1. Objetivo

O objetivo do estudo para informatizar a prova Elementos em Árvore da ECCOs 4/10 foi oferecer uma alternativa à sua congênere em formato papel, procurando validar o uso da modalidade eletrônica dessa prova pelo psicólogo e identificar os eventuais ganhos permitidos por esse tipo de abordagem.

De antemão, sabia-se que devido à baixa quantidade de itens disponíveis nessa prova ela não se prestaria para incorporação dos procedimentos metodológicos da teoria da resposta ao item nem mesmo poderia ser transformada em teste adaptativo computadorizado, mas que, por outro lado, existia um conjunto de benefícios que se poderia oferecer ao psicólogo, tais como realizar a cotação dos pontos automaticamente e controlar o tempo de exibição dos elementos na tela.

5.2.2. Método

Participantes

Participaram deste estudo 28 crianças de uma escola particular do município de Jaboatão dos Guararapes, Pernambuco, classificados por idade e sexo, conforme Tabela 7.

Tabela 7 - Classificação dos participantes da Prova Elementos em Árvores

Classificação por sexo	Total de Alunos	Classificação por Idade						
		4 anos	5 anos	6 anos	7 anos	8 anos	9 anos	10 anos
Masculino	15	2	3	3	1	3	2	1
Feminino	13	2	2	2	2	2	1	2
Total:	28	4	5	5	3	5	3	3

Instrumento

Foi utilizado o programa de computador descrito no tópico 5.1 como instrumento suficiente para realizar as seguintes funções:

- Apresentar, registrar e armazenar no banco de dados as informações cadastrais e os indicadores acadêmicos e socioeconômicos do aluno
- Apresentar os itens na tela do computador, coletando as respostas fornecidas pelo aluno
- Registrar os indicadores auxiliares referentes à realização do teste, tais como data e hora, tempos para cada item e tempo acumulado
- Controlar os tempos de exibição dos elementos das árvores em cada item, conforme o manual de aplicação da prova
- Realizar automaticamente a cotação dos pontos obtidos pela criança
- Finalizar o teste quando alcançado o critério de parada definido no manual.

Procedimentos

Para a aplicação do teste foi utilizado em momentos alternados um tablet de 10 polegadas e um computador que também dispunha de tela sensível ao toque (touch screen), visando facilitar a aplicação e aumentar o interesse por parte da criança.

A aplicação da prova foi realizada conforme os procedimentos descritos no manual de aplicação da ECCOs 4/10, ajustados para a versão eletrônica, de forma assistida, com o pesquisador interagindo com a criança durante todo o teste, tal qual a aplicação do teste em

papel, conforme os procedimentos descritos no Anexo H. A administração da prova era encerrada automaticamente quando nas duas árvores de uma mesma série a criança não conseguia um total de pontos, no mínimo, igual ao número de elementos presentes em uma só das árvores dessa série.

Assim como na versão original, dois itens de treino eram oferecidos no início do teste, visando a completa compreensão pela criança do funcionamento do teste antes de começar a fase de pontuação.

Frente às respostas produzidas pelas crianças, traduzidas pelos toques na tela nos locais onde acreditam que os elementos estavam, o único feedback dado é uma sinalização visual temporária da posição onde ocorreu o toque. A sinalização não permanecia na tela, uma vez que se deixássemos o registro dos toques demarcados provocaria uma diferenciação importante frente à prova original, em papel, onde a criança apenas aponta para as posições onde os elementos estariam.

5.2.3. Resultados

Em geral, após as devidas explicações e a administração dos dois itens de treino, as crianças conseguiram utilizar o programa de forma autônoma, mesmo as mais jovens, fosse a prova realizada no tablet ou no computador.

Do ponto de vista do pesquisador, administrador do teste, além dos benefícios de portabilidade e coleta automática de dados, inerentes à qualquer teste informatizado, a versão eletrônica da prova Elementos em Árvore apresentou vantagens adicionais bastante importantes em relação à sua similar, em papel, quer sejam dispensar o manuseio de cartões e de cronômetros, bem como calcular a pontuação de forma automática, uma vez que o programa foi suficiente para dar conta de todo o processo de controle de tempo de exposição dos itens e da cotação da pontuação dos participantes, liberando o pesquisador da forte assistência requerida na versão não automatizada. Ou seja, a aplicação da prova se tornou bem mais confortável para o psicólogo ou pesquisador e pode ajudar a eliminar eventuais diferenças entre estilos ou mesmo qualidade da aplicação entre profissionais diferentes, colaborando na uniformização do padrão dos resultados. Ressalta-se que o processo de pontuação manual dessa prova, como apresentado no Anexo H, é relativamente complexo e sujeito a eventuais equívocos, pois envolve temporização, adição e diminuição de pontos, conforme os acertos e erros dos participantes vão acontecendo ao longo da prova.

Os escores obtidos pelas crianças no teste eletrônico se mostraram ligeiramente superiores e com menores desvios do que os achados nas provas em papel na amostra brasileira do estudo 1 (capítulo III) e de Portugal, como pode ser visto na Tabela 8. Porém, não é possível generalizar e afirmar que o desempenho dos estudantes melhora na prova eletrônica porque esses indicadores podem ter sido beneficiados pela baixa quantidade de participantes no estudo com teste eletrônico (n=28) frente aos estudos com testes em papel (n=604 no brasileiro, n=539 no português). Além do mais, as condições diferenciadas de

aplicação também podem ter influenciado nos resultados, uma vez que no estudo 1 e na aplicação em Portugal os alunos estavam sendo submetidos a todas as onze provas da ECCOs 4/10 em sequência, ao contrário do estudo com a prova informatizada.

É importante notar que a relação entre os desempenhos e as idades dos participantes manteve-se coerente em todos os resultados, crescendo o escore na medida do crescimento da idade, indicando que a prova eletrônica está funcionando a contento e pode ser utilizada, com vantagens, em substituição à prova em papel.

Tabela 8 - Médias e desvios da prova Elementos em Árvores.

Idade	BRASIL				PORTUGAL	
	Versão Digital (n=28)		Versão Papel (n=604)		Versão Papel (n=539)	
	Média	Desvio	Média	Desvio	Média	Desvio
4	22,75	2,22	20,40	7,03	9,64	5,89
5	24,40	4,77	23,85	7,08	15,57	7,96
6	26,20	5,40	25,53	7,12	21,26	9,47
7	29,67	2,08	29,10	5,77	25,23	11,49
8	33,80	3,77	31,87	7,13	32,35	11,19
9	36,33	4,04	33,40	6,21	34,81	10,63
10	43,00	8,54	35,20	6,68	36,65	10,72

Nenhuma desvantagem pode ser observada ante a utilização deste instrumento no formato eletrônico em relação à aplicação tradicional, com cartões, cronômetros, fichas e lápis.

5.3. Versão informatizada sequencial da prova de compreensão Desenhos Absurdos

Como vimos no capítulo II, na ECCOs 4/10 o processo de compreensão é abordado frente à realidade social, na forma como a criança conceitualiza as outras pessoas e como aborda os pensamentos, emoções, intenções e pontos de vista dos outros (Shantz, 1975). Conforme Brito e Almeida (2009), a prova Desenhos Absurdos "é constituída por imagens relativas a acontecimentos familiares à criança, apresentando incongruências de elementos-contextos, através dos quais se pretende avaliar competências relativas ao conhecimento e compreensão do meio e à organização semântica ou contextual dos diversos elementos constantes das figuras". Em outras palavras, são apresentadas imagens com situações que denotam algum absurdo que deverá ser percebido e indicado pela criança. Todos os itens estão demonstrados no Anexo O.

5.3.1. Objetivo

O objetivo do estudo para informatizar a prova Desenhos Absurdos da ECCOs 4/10 foi oferecer uma alternativa à sua congênere em formato papel, procurando validar o uso da modalidade eletrônica dessa prova pelo psicólogo e identificar os eventuais ganhos permitidos por esse tipo de abordagem.

De antemão, sabia-se que devido à necessidade constante de diálogo entre o aplicador e a criança nessa prova ela não se prestaria para incorporação dos procedimentos metodológicos da teoria da resposta ao item nem ser transformada em teste adaptativo computadorizado. A análise sobre as características dos itens e do sistema de pontuação indicava que também não seria possível determinar os escores das crianças automaticamente pelo computador.

5.3.2. Método

Participantes

Participou desse ensaio o mesmo grupo de 28 alunos apresentados na Tabela 7 referente à prova Elementos em Árvore.

Instrumentos

Foi utilizado o programa de computador descrito no tópico 5.1 como instrumento auxiliar à aplicação da prova, dando conta das seguintes funções:

- Apresentar, registrar e armazenar no banco de dados as informações cadastrais e os indicadores acadêmicos e socioeconômicos do aluno
- Apresentar os itens na tela do computador, coletando as respostas fornecidas pelo aluno
- Registrar os indicadores auxiliares referentes à realização do teste, tais como data e hora, tempos para cada item e tempo acumulado
- Finalizar o teste quando alcançado o critério de parada definido no manual.

Procedimentos

A aplicação desta prova seguiu os procedimentos descritos no manual de aplicação da ECCOs 4/10, ajustados para a versão eletrônica, conforme Anexo I. Para todas as crianças, a prova tem início no item 1 e termina após quatro respostas mal sucedidas ou quando todos os itens tiverem sido apresentados.

Assim como na versão original, dois itens de treino eram oferecidos no início do teste, momento em que o pesquisador procurava dar à criança uma completa compreensão do funcionamento do teste antes de começar a fase de pontuação.

5.3.3. Resultados

Devido às características de cotação para pontuação, que depende da fala da criança, e da construção de alguns itens, que permitem múltiplas interpretações, a versão informatizada dessa prova não foi suficiente para dispensar a assistência continuada do aplicador do teste, não trazendo benefícios inovadores ao processo de teste e representando pouco mais do que a simples substituição do meio papel pelo eletrônico.

Anota-se que o processo de pontuação implica em ouvir o que a criança fala sobre o absurdo que ela foi capaz de observar nas figuras, podendo ser atribuído um, dois ou três pontos para cada item, de acordo com o que ela descreveu. Além do mais, alguns itens admitem que a localização do absurdo possa variar dependendo da interpretação da criança sobre a cena. Por exemplo, o item em que a figura de um menino e um macaco estão em posições trocadas, o menino dentro da jaula e o macaco fora dela, dá margem para a criança dizer que o absurdo é o menino dentro da jaula ou que o absurdo seja o macaco estar fora da jaula. As duas respostas podem ser consideradas certas. Em um outro item semelhante, o gato está dentro do aquário, enquanto o peixe está fora dele. Em outro, um menino está surfando com um skate, o que poderia indicar que o menino está surfando com trajes e equipamentos errados ou que o mar está no lugar da rua... Em itens concebidos dessa forma não é possível a um programa de computador reconhecer se a criança percebeu ou não o absurdo apenas com um simples apontar na tela ou com o clicar do mouse. Para que a criança pudesse realizar essa prova com autonomia, sem a assistência continuada do aplicador do teste, seria necessário promover alterações importantes no banco de itens tendo isso em mente, de forma que todos os itens admitissem apenas uma interpretação e que essa interpretação única pudesse ser traduzida por um clique que o indicasse, ou seja, o absurdo precisaria estar localizado em uma determinada área da tela. Exemplos de itens com esse formato podem ser obtidos na própria prova Desenhos Absurdos, tal como no item em que uma criança está

uniformizada, correndo em um campo e chutando uma bola, mas calçando um sapato de salto alto - esse item admite que o clique do mouse no sapato é a única resposta correta, nenhuma outra.

Assim, dentre as quatro provas que foram informatizadas, essa foi a única que requereu marcação de escores externos, não sendo possível fazer a coleta dos dados das respostas e pontuação automaticamente. Dessa forma, os benefícios da prova informatizada em relação à original foi apenas dispensar o uso de papel, apresentando os itens na tela do computador ou tablet e dar conta dos registros cadastrais do aluno. De qualquer forma, utilizamos técnicas da linguagem de programação que permitiram separar virtualmente cada elemento das figuras, sendo possível identificar e registrar qual foi a parte da imagem que o aluno indicou como absurdo em cada item. Por exemplo, em uma figura em que aparece um hipopótamo de óculos escuros na praia, é possível saber se o aluno clicou (com o dedo ou mouse) na região dos óculos ou se clicou em outra região diferente.

Vale destacar que uma prova como essa, onde se busca identificar a percepção pela criança de absurdos dentro de contextos sociais, tem um enorme potencial para ser totalmente automatizada e utilizar itens multimídia, com movimentos e sons, estratégias que estão ganhando força no seio da psicometria e que visam ser mais envolventes e naturais do que os itens estáticos, mas necessitaria de um trabalho à parte.

5.4. Versão informatizada sequencial da prova de percepção Comparação de Figuras

Como vimos no capítulo II, na ECCOs 4/10 a prova perceptiva está centrada na avaliação da percepção visual, articulando a capacidade de discriminação e de acuidade visuais com a velocidade desse processamento. As tarefas presentes nesta prova requerem concentração e atenção por parte da criança para análise e captação de detalhes que diferenciam ou aproximam uma figura da outra. Segundo Brito e Almeida (2009), "a prova de Comparação de Figuras apresenta cinco imagens que diferem apenas em pequenos pormenores, devendo a criança indicar as duas imagens que são exatamente iguais". Todos os itens estão representados no Anexo N.

5.4.1. Objetivo

O objetivo do estudo para informatizar a prova Comparação de Figuras da ECCOs 4/10 foi oferecer uma alternativa à sua congênere em formato papel, procurando validar o uso da modalidade eletrônica dessa prova pelo psicólogo e identificar os eventuais ganhos permitidos por esse tipo de abordagem. Além disso, a informatização da versão sequencial da prova de percepção faz parte da preparação para o estudo seguinte, descrito no capítulo VI, que irá desenvolver e validar uma versão especial dessa prova no modelo de teste adaptativo computadorizado.

Antecipadamente, já sabíamos que a versão eletrônica dessa prova permitiria um grande grau de automação, seja permitindo o controle do avanço no teste diretamente pelo aluno, seja pela cotação automática da pontuação.

5.4.2. Método

Participantes

Participaram desse estudo 63 alunos de uma escola particular do município de Caruaru, em Pernambuco, conforme caracterizado na Tabela 9.

Tabela 9 - Classificação dos participantes da prova Comparação de Figuras.

Classificação por sexo	Total de Alunos	Classificação por Idade						
		4 anos	5 anos	6 anos	7 anos	8 anos	9 anos	10 anos
Masculino	39	2	2	8	7	5	10	5
Feminino	24	2	2	5	5	4	3	3
Total:	63	4	4	13	12	9	13	8

Instrumento

Foi utilizado o programa de computador descrito no tópico 5.1 como instrumento de apoio para realizar as seguintes funções:

- Apresentar, registrar e armazenar no banco de dados as informações cadastrais e os indicadores acadêmicos e socioeconômicos do aluno
- Apresentar os itens na tela do computador, coletando as respostas fornecidas pelo aluno
- Registrar os indicadores auxiliares referentes à realização do teste, tais como data e hora, tempos para cada item e tempo acumulado
- Realizar automaticamente a cotação dos pontos obtidos pela criança
- Finalizar o teste quando alcançado qualquer dos critérios de parada definidos no manual.

Procedimentos

Para a aplicação do teste foi utilizado em momentos alternados um tablet de 10 polegadas e um computador que também dispunha de tela sensível ao toque (touch screen), visando facilitar a aplicação e aumentar o interesse por parte da criança.

Assim como na versão original, dois itens de treino eram oferecidos no início do teste, momento em que o pesquisador procurava dar à criança uma completa compreensão do funcionamento do teste antes de começar a fase de pontuação.

A aplicação da prova seguiu os procedimentos descritos no manual de aplicação da ECCOs 4/10, ajustados para a versão eletrônica, conforme Anexo G, de forma assistida pelo pesquisador, mas procurando que o aluno se tornasse autônomo para a condução do teste, interagindo com o computador para responder e avançar para o próximo item. A administração da prova era encerrada automaticamente quando a criança não fosse bem sucedida em quatro itens consecutivos ou pelo esgotamento do tempo máximo previsto para execução do teste que varia em função da idade da criança, tal como descrito no Anexo G.

Através de toques com o dedo na tela ou com o clicar do mouse, a criança podia selecionar o par de figuras que identificasse como sendo exatamente iguais. Para cada toque o programa desenhava uma moldura vermelha ao redor da imagem selecionada como forma de demarcar as imagens que foram indicadas. Um novo toque sobre a figura já demarcada fazia desaparecer a moldura, liberando para a seleção de outra figura.

Para definir a sequência de apresentação dos itens, as respostas dos 1.145 participantes do estudo 1 (descrito no capítulo III), realizado no Brasil e do estudo feito em Portugal, ambos no formato tradicional, em papel, foram submetidas ao programa Winsteps, gerando uma escala avalizada pela teoria da resposta ao item (TRI) com estimativas das dificuldades dos itens, variando de -3,64 até 3,64 logits.

5.4.3. Resultados

A escala definida pelo programa Winsteps serviu de base orientadora para a definição da sequência de apresentação dos itens do teste, ou seja, ao invés de serem apresentados os itens na mesma sequência utilizada no teste em papel, sequenciamos a apresentação dos itens de acordo com a dificuldade demonstrada pelos participantes nos estudos anteriores. Essa modificação no sequenciamento era necessária por dois motivos: primeiro, para justificar o próprio encerramento do teste após quatro erros sucessivos, o que só faz sentido se os itens estiverem ordenados de acordo com a sua dificuldade; segundo, para que seja possível utilizar e comparar a prova de Comparação de Figuras no estudo com testes adaptativos, que será demonstrado adiante.

Além das vantagens óbvias de cotação dos pontos e coleta automática dos dados, a versão eletrônica da prova de percepção também eliminou a necessidade de se utilizar e controlar o cronômetro, uma vez que o programa já considerava o tempo máximo de administração da prova de acordo com a idade do aluno, encerrando-a, automaticamente, ao término do tempo.

Capítulo VI

Estudo 4 - Teste Adaptativo Computadorizado

6. ESTUDO 4: TESTE ADAPTATIVO COMPUTADORIZADO (TAC)

6.1. Introdução

Como vimos, o teste adaptativo computadorizado (TAC) ou teste adaptativo informatizado (TAI), em inglês computer-adaptive testing (CAT), é um padrão para o desenvolvimento e aplicação de testes que representa o estado da arte no campo da psicometria moderna. Nessa abordagem, o computador seleciona cada item que será apresentado ao participante em função do nível de habilidade que ele vai demonstrando ter durante a realização do próprio teste, resultando em provas personalizadas, adaptadas à capacidade cognitiva de cada indivíduo. O TAC simboliza o desenho atual de instrumentos de medições psicológicas e educacionais usando computadores, sendo indicado para mensuração de habilidades, medidas de personalidade e de variáveis atitudinais.

A utilização de TAC é crescente em todo o mundo, já sendo largamente utilizado nos países desenvolvidos, notadamente no monitoramento da aprendizagem escolar, área onde vem se tornando recurso essencial das mais recentes propostas educativas, tais como nos modelos híbridos de aprendizagem, onde os métodos das aulas presenciais são combinados com atividades mediadas por computadores (“Blended_learning,” n.d.).

Basicamente, a implementação de um teste adaptativo computadorizado requer um banco de itens calibrado pela teoria da resposta ao item. Com a ajuda da TRI podemos extrair a dificuldade e a quantidade de informação presente em cada item, de forma a escalonar os itens de acordo com suas dificuldades para projetar formas paralelas de testes e para prover dados para os testes adaptativos. Uma vez de posse do banco de itens, durante a aplicação do teste, cada nova resposta do participante faz com que seja gerada uma nova estimativa de habilidade que leve em consideração todas as suas respostas anteriores. Diante de cada nova estimativa de habilidade, o computador seleciona o item mais adequado para ser apresentado

como próximo item, que poderá ser mais fácil ou mais difícil do que o anterior a depender da estimativa gerada. Esse processo continua em ciclo até que um critério de parada previamente estabelecido seja satisfeito, quando a estimativa de habilidade final será gerada. Considerando que o banco de itens seja adequado, os pontos essenciais para o bom desempenho do TAC são os cálculos das estimativas de habilidades e os critérios de parada.

Dando prosseguimento às pesquisas realizadas sobre a ECCOs 4/10, neste capítulo iremos apresentar um estudo onde foram desenvolvidos testes adaptativos computadorizados a partir das provas Figuras Incompletas e Comparação de Figuras, que avaliam os processos de raciocínio e percepção, respectivamente, na ECCOs 4/10. Relatamos todo o caminho percorrido, desde a escolha das provas e dos recursos necessários para a produção do programa de computador adaptativo até a própria aplicação dos testes, passando pelas dificuldades encontradas e as soluções apresentadas, culminando com as comparações com as versões não adaptativas do teste e o relato das conclusões. Como se trata de um campo ainda pouco explorado no Brasil, esperamos que os ganhos decorrentes do domínio das teorias e tecnologias envolvidas no processo transcendam este trabalho e possam ser replicados para outros estudos e aplicações.

6.2. A escolha das provas

Antes de mais nada, precisamos considerar que o desenvolvimento de provas no modelo de TAC baseadas na TRI pressupõe a existência de um banco de itens devidamente analisado de maneira que seja conhecido para cada item o índice de dificuldade, a quantidade de informação que ele carrega e se todos os itens apontam em uma única direção. Além disso, o teste adaptativo também irá requerer uma quantidade suficiente de itens para formar um conjunto hierarquicamente ordenado pela dificuldade estimada, de maneira a permitir

localizar a cada instante um determinado item dentro do *continuum* do construto que se deseja avaliar.

Como quando da concepção das provas da ECCOs 4/10 não se previa a utilização de seus itens em testes adaptativos computadorizados, nossa primeira tarefa foi determinar quais seriam as provas que ofereciam um mínimo de condições para a construção de testes adaptativos computadorizados. Como sabemos, a ECCOs 4/10 é composta por onze provas que avaliam seis processos cognitivos nas dimensões verbal e não verbal. Em teoria, as seis provas de natureza não verbal seriam as candidatas para serem informatizadas, pois, diferentemente das provas de natureza verbal, elas não requererem a leitura de textos nem mesmo exigem a realização de muitos diálogos entre o pesquisador e o participante durante sua aplicação. Essas provas poderiam ser estruturadas e programadas mais facilmente para utilização diretamente pela criança participante, apenas com a orientação e acompanhamento do pesquisador.

No entanto, olhando mais de perto as necessidades inerentes aos procedimentos de aplicação de cada prova não verbal e recorrendo às características essenciais para elaboração de um teste adaptativo, constatamos pontos inconvenientes e até impossibilidades diante de algumas provas para suportar um estudo de teste adaptativo computadorizado no modelo pretendido. As seguintes análises foram consideradas sobre a possibilidade de construção das provas como teste adaptativo computadorizado (TAC):

Prova Elementos em Árvores (descartada para TAC)

- Pontos desfavoráveis: não dispõe de itens em quantidade suficiente e utiliza critério de pontuação inadequado para TAC.
- Não há pontos favoráveis.

Prova Composição de Padrões (descartada para TAC)

- Pontos desfavoráveis: não dispõe de itens em quantidade suficiente; critério de pontuação inadequado para TAC.
- Não há pontos favoráveis.

Prova Construção de Desenhos (descartada para TAC)

- Pontos desfavoráveis: prevê dinâmica de interação continuada com o aplicador do teste.
- Não há pontos favoráveis.

Prova Figuras Absurdas (descartada para TAC)

- Pontos favoráveis: dispõe de pequena quantidade de itens, mas ainda assim seria razoável para experimentação.
- Pontos desfavoráveis: requer interação continuada com o pesquisador, inclusive para realizar a cotação de pontos.

Prova Comparação de Figuras

- Pontos favoráveis: dispõe de quantidade razoável de itens para o estudo; os procedimentos de aplicação e pontuação não exigem interação continuada com os alunos.
- Pontos desfavoráveis: na prova com lápis e papel utiliza-se o tempo de execução da prova como um critério de parada e como indicativo para a medida de desempenho. No caso do teste adaptativo, como se pressupõe a apresentação de um número menor e variável de itens, o tempo não pode ser utilizado como critério de parada.

Prova Figuras Incompletas

- Pontos favoráveis: dispõe de quantidade razoável de itens para o estudo; os procedimentos de aplicação e pontuação não exigem interação continuada com os alunos; já estava informatizada no modelo sequencial e seus itens já haviam sido testados sob a ótica da TRI.
- Pontos desfavoráveis: na prova com lápis e papel utiliza o tempo como critério de parada e como indicativo para a medida de desempenho.

Diante das análises aqui expostas, acabamos optando por realizar o estudo de testes adaptativos computadorizados com as provas Figuras Incompletas, que se presta à avaliação do raciocínio abstrato, e Comparação de Figuras, que avalia o processo de percepção, haja vista serem as únicas que apresentaram características mínimas válidas para um teste adaptativo. No caso da prova de raciocínio, Figuras Incompletas, já havia uma versão computadorizada que fora produzida em conformidade com a teoria da resposta ao item (ver o estudo 2 no Capítulo IV) e cujas propriedades, conhecidas, indicavam ser possível desenvolver um teste adaptativo sobre ela. No caso da prova de percepção também já existia uma análise preliminar dos itens baseada na TRI, conforme estudo 3 apresentado no Capítulo V, que se utilizou dos dados da aplicação do teste no formato lápis e papel como ponto de partida para compreender a dificuldade dos itens. As características dos itens da prova Comparação de Figuras, no entanto, não se mostraram tão favoráveis quanto as da prova Figuras Incompletas, haja vista a existência de lacunas e sobreposições na escala.

6.3. Construção do programa de computador adaptativo

Programa

O programa adaptativo foi construído tendo por base o programa que fora construído para aplicação da versão sequencial dos testes, descrito no item 5.1, incluindo o banco de itens e a interface com o aluno. Para a construção da versão adaptativa, integramos ao Java a linguagem de programação R para executar funções especializadas nas estatísticas inerentes aos testes adaptativos. Segundo o site do projeto R (<http://www.r-project.org>), o R é, ao mesmo tempo, uma linguagem de programação e um ambiente computacional para criação e manipulação de estatísticas e gráficos que foi desenvolvida no Bell Laboratories por John Chambers e colegas. A linguagem R provê grande variedade de estatísticas e técnicas gráficas, sendo altamente extensível.

Para possibilitar a implementação das funções adaptativas associado ao R pelo programa, utilizamos a biblioteca catR, desenvolvida por David Magis e amigos. Apesar de existirem outras bibliotecas para auxiliar o programador na criação de testes adaptativos na linguagem R, tal como a catIRT (<http://cran.r-project.org/web/packages/catIrt/catIrt.pdf>), a catR foi aquela que consideramos mais completa no momento por contemplar diferentes métodos para as funções de TAC em cada passo do programa e, além disso, possuir grande flexibilidade e facilidade para gerar padrões de respostas. Através das funções disponíveis na biblioteca catR, obtivemos um controle absoluto sobre todos os passos necessários para a execução do algoritmo do teste adaptativo, como, por exemplo, para seleção do primeiro e do próximo item ou para calcular as estimativas de habilidades parcial e final, entre outros.

Banco de Itens

No caso da prova de raciocínio, Figuras Incompletas, foi utilizado o banco de itens oriundo da prova informatizada sequencial (ver estudo 2 do Capítulo IV), uma vez que ele já estava completamente validado e calibrado pela teoria da resposta ao item (TRI), como consequência daquele estudo. Esse banco de itens é composto por 43 itens que representam um conjunto equilibrado de itens, com boas características psicométricas. A distribuição é suficiente para exibir a média e dois desvios padrão, tanto para itens quanto para pessoas, podendo representar, praticamente, toda faixa de habilidade das pessoas. A medida de dificuldade dos itens se situou entre -4,48 e 2,41 logits e a estimativa de habilidade na amostragem de 304 pessoas do estudo 2 variou de -4,34 a 3,39 logits.

Já para a prova de percepção, Comparação de Figuras, foi utilizado o banco de dados construído no estudo 3, descrito no Capítulo V, onde foi realizada uma avaliação inicial da dificuldade e propriedades dos itens tendo por base a coleta de dados realizada na aplicação da prova no formato lápis e papel no Brasil (ver Capítulo III). Através do software estatístico Winsteps foi gerada uma distribuição de dificuldade dos 38 itens disponíveis e produzida uma escala cuja medida de dificuldade dos itens variou de -3,64 a 3,64 logits.

Banco de Dados

Já a base de dados referente à realização dos testes precisou ter sua estrutura ampliada para conter as especificidades dos testes adaptativos, passando a registrar os dados das estatísticas parciais e finais produzidas pelo programa. O Anexo K descreve a estrutura da base de dados para o teste adaptativo.

6.4. O estudo empírico

6.4.1. Objetivos

6.4.1.1. *Geral*

O objetivo geral desse estudo 4 foi construir uma versão protótipo de teste adaptativo computadorizado para avaliação dos processos cognitivos de raciocínio e percepção, tendo por base os itens das provas da ECCOs 4/10 denominadas Figuras Incompletas e Comparação de Figuras, respectivamente.

6.4.1.2. *Específicos*

Interessou a este estudo verificar a correlação entre as habilidades estimadas no teste adaptativo com a prova original, em lápis e papel, e com as provas computadorizadas no modelo sequencial (estudo 2 no capítulo IV e estudo 3 no capítulo V). Poderia o TAC ser mais eficiente do que o teste sequencial, seja em lápis e papel ou informatizado?

Também havemos de compreender as vantagens e desvantagens da utilização desse tipo de abordagem em testes para avaliação de crianças com idade variando entre 4 e 10 anos, identificando os eventuais problemas e limites processuais ou tecnológicos para aplicação de testes adaptativos computadorizados com esse público.

Do ponto de vista da aquisição de conhecimentos, nos interessou apreender os conceitos, teorias e procedimentos metodológicos que ancoram os testes adaptativos computadorizados, haja vista seu potencial para utilização em outros estudos em benefício da Educação e da Psicologia. No mesmo sentido, queríamos mapear as linguagens de programação, bibliotecas e recursos computacionais que orientam a construção de abordagens dessa natureza.

6.4.2. Método

6.4.2.1. Participantes

Estudo piloto: participaram do piloto 63 crianças de duas escolas particulares para classe média (notadamente classes C e B) do município de Jaboatão dos Guararapes e Caruaru, em Pernambuco, sendo 8 crianças matriculadas na Educação Infantil (pré-escolar) e 55 crianças no Ensino Fundamental I (1º ciclo do ensino básico). A amostra está distribuída por idade e sexo na Tabela 10 a seguir.

Tabela 10 - Distribuição dos participantes no estudo piloto por idade e sexo

Classificação por sexo	Total de Alunos	Classificação por Idade						
		4 anos	5 anos	6 anos	7 anos	8 anos	9 anos	10 anos
Masculino	39	2	2	8	7	5	10	5
Feminino	24	2	2	5	5	4	3	3
Total:	63	4	4	13	12	9	13	8

Teste formal: participaram do teste formal 53 crianças que haviam participado do estudo piloto e mais 79 crianças das mesmas escolas, totalizando 132 crianças. Destas, 29 estavam matriculadas na Educação Infantil e 103 no Ensino Fundamental I. A Tabela 11 a seguir sumariza a caracterização do grupo.

Tabela 11 - Distribuição dos participantes do teste formal por idade e sexo

Classificação por sexo	Total de Alunos	Classificação por Idade						
		4 anos	5 anos	6 anos	7 anos	8 anos	9 anos	10 anos
Masculino	73	7	7	14	10	9	15	11
Feminino	59	8	7	8	11	9	8	8
Total:	132	15	14	22	21	18	23	19

6.4.2.2. *Instrumentos e procedimentos*

Para a administração do teste e a respectiva coleta de dados foi utilizado o programa de computador descrito no tópico 6.3, utilizando a linguagem Java / JavaFX, integrada à linguagem R versão 3.0 com a adição do pacote catR. O programa tem a capacidade de rodar ambos os testes, de raciocínio e percepção, em dois modelos distintos: sequencial ou adaptativo. No caso do teste em modelo sequencial, todos os itens disponíveis no banco de itens vão sendo apresentados ordenados pelo seu nível de dificuldade até que o banco de itens seja exaurido ou que o aluno erre quatro itens consecutivos. No caso do teste adaptativo, o programa utiliza as estratégias delineadas a seguir para cada ponto do algoritmo de teste adaptativo.

O modelo de Rasch (1PL) foi utilizado como modelo logístico da TRI, servindo tanto na fase de calibração dos itens quanto para base da análise dos dados coletados.

O procedimento de teste-reteste foi utilizado para avaliar a fidedignidade das medidas do teste, observando intervalo médio de trinta dias entre as coletas. As correlações entre as versões sequencial (estudo 2, anterior) e adaptativa (estudo atual) do teste também foram verificadas.

Para a análise dos resultados dos testes foram empregados os programas IBM SPSS versão 21 e Winsteps Rasch Measurement versão 3.81, este destinado a análises de aplicações baseadas na TRI utilizando o modelo de Rasch.

A respeito dos procedimento em relação as testes adaptativos computadorizados (TAC), os seguintes conceitos e critérios foram utilizados:

Seleção do primeiro item

A seleção do primeiro item pelo programa foi feita através de sorteio de itens cuja dificuldade estivesse próxima de 30% da capacidade da escala, abaixo, portanto,

do seu ponto central, permitindo variar o primeiro item a cada aplicação do programa.

Estimativa de habilidade

As estimativas de habilidades foram calculadas utilizando-se as funções `tethaEst()` e `semTheta()` disponíveis no pacote `catR`. Essas funções implementam o método Bayesiano *expected a posteriori* (EAP), sendo o nível de precisão fornecido através do desvio-padrão a posteriori (*posteriori standard deviation* – PSD).

Seleção do próximo item

Para seleção do próximo item a cada passo do teste foi utilizado o método de máxima informação de Fisher (*maximum Fisher information*), obtido através da função `nextItem()` do `catR`.

Critérios de parada

Como critério principal de parada para o programa adaptativo do teste foi utilizada a estimativa do erro-padrão presente na medida. Considerou-se que o teste disporia de uma medida aceitável e poderia ser finalizado se o erro estimado nas quatro últimas medições fosse aceitável e estivesse estabilizado com variação menor que 0,001. Garantiu-se que um mínimo de 10 e um máximo de 25 itens fossem administrados em todos os testes. O tempo de duração do teste não foi utilizado como critério de parada.

6.4.2.3. *Procedimentos*

Este estudo está em conformidade com a resolução CNS 196/96 e suas complementares do Conselho Nacional de Saúde. O termo de consentimento está reproduzido no Anexo E.

A administração dos testes ocorreu dentro do ambiente escolar utilizando computadores com telas sensíveis ao toque (touch screen), onde os alunos podiam indicar as suas escolhas apontando na tela a opção desejada para cada item. Os indicadores de desempenho escolar e de situação socioeconômica dos alunos foram solicitados previamente às escolas de forma que no momento do registro do aluno na base de dados o cadastro já fosse completamente preenchido.

Os alunos foram submetidos aos testes de raciocínio e de percepção nas modalidades sequencial (estudos 2 e 3) e adaptativa em semanas alternadas.

6.4.3. Resultados

O estudo piloto serviu para ajustar alguns detalhes da programação, para perceber o comportamento dos testes adaptativos e para ajudar a validar a estabilidade da medida em situações distintas, mas não houve necessidade de muitas alterações. Na verdade, como havíamos desenvolvido a prova informatizada de Figuras Incompletas durante o estudo 2, descrito no Capítulo IV, já dispúnhamos das artes gráficas e da lógica de funcionamento para o teste de raciocínio. A versão sequencial da prova de Comparação de Figuras também foi desenvolvida na sequência, imediatamente após a conclusão do estudo 2.

Os testes adaptativos se comportaram dentro do padrão programado, sem surpresas. Como esperado na modalidade adaptativa para ambos os testes, o algoritmo do programa foi capaz de refinar as estimativas de habilidades na medida em que os itens foram sendo administrados até satisfazer o critério de parada. As Figuras 18 e 19 adiante retratam duas situações reais, demonstrando visualmente esse comportamento de forma bastante clara.

Também é possível perceber visualmente nas Figuras 18 e 19 que o critério de parada baseado no tamanho do erro padrão foi determinante para finalizar quase todos os testes realizados, assim como era desejado. Em apenas dois casos, ambos no teste de Comparação de Figuras, a finalização ocorreu em função de ser alcançado o número máximo de 25 itens previstos para administração. Em outras palavras, na grande maioria dos testes realizados o programa conseguiu construir a confiança prevista para a medida da estimativa de habilidade do aluno participante e deu o teste por concluído.

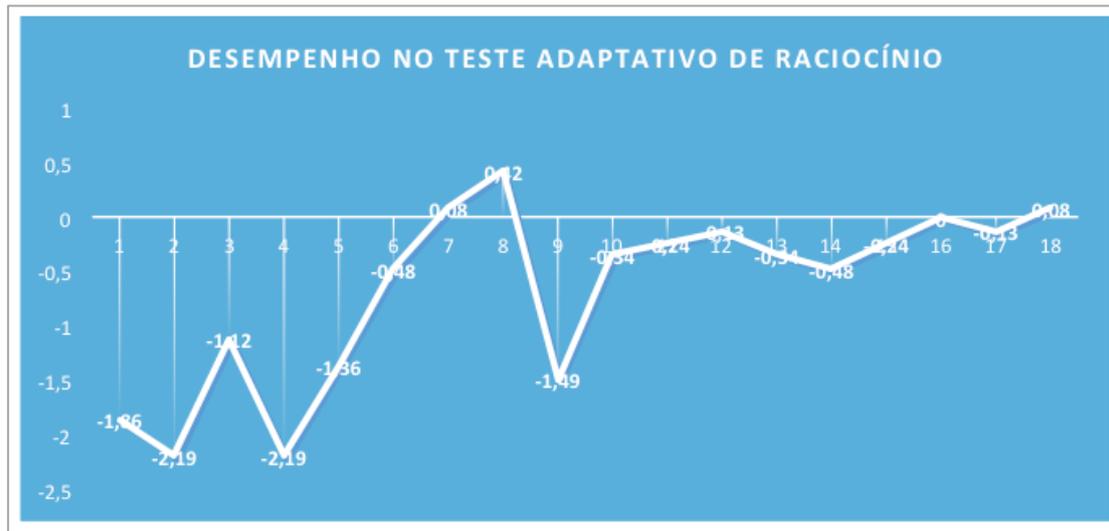


Figura 18 - Exemplo real de desempenho no teste adaptativo de raciocínio - prova Figuras Incompletas: o programa vai ajustando a estimativa de habilidade do aluno na medida em que mais itens são administrados.

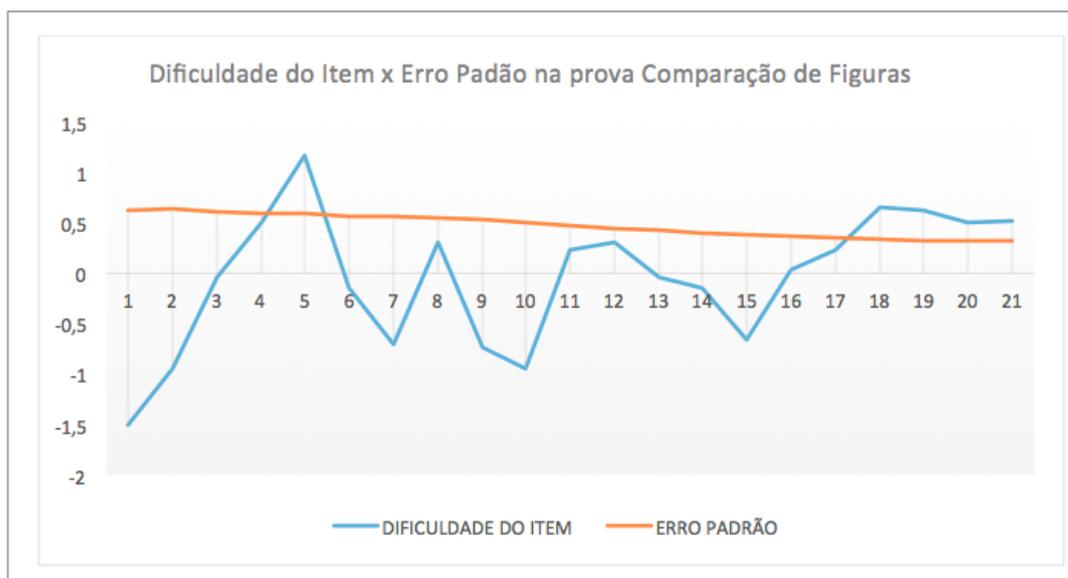


Figura 19 - Exemplo real de desempenho no teste adaptativo de percepção – prova Comparação de Figuras. Quanto mais itens são administrados, menor o erro estimado para a medida. No caso, foi necessário administrar 21 itens para se obter uma estimativa de erro aceitável para a estimativa de habilidade.

Para refletir sobre o que acontece durante a execução de um teste adaptativo vamos olhar para os dados na Tabela 12, a seguir, que remonta passo a passo a administração do teste e as estimativas de habilidades que foram calculadas pelo programa. Lembramos que, tal como foi definido pelo método, nesse teste o primeiro item a ser administrado apresenta dificuldade em

torno de 30% da capacidade da escala, abaixo, portanto, do seu ponto central. O principal critério de parada determina que o erro padrão presente na medida deveria estar estabilizado com variação menor que 0,001 e tendo um mínimo de 10 itens administrados.

Tabela 12 - Itens administrados versus estimativas de habilidade

Qtd de Itens Administrados	Acerto	Estimativa de Habilidade	Erro Padrão
1	0	-3,516762	0,58
2	1	0,061605	0,57
3	0	-0,858074	0,57
4	1	0,061605	0,54
5	1	0,551377	0,53
6	1	0,671605	0,52
7	1	0,809769	0,48
8	0	-0,058229	0,44
9	1	0,718522	0,42
10	1	0,927385	0,40
11	1	1,148805	0,39
12	0	1,111999	0,38
13	0	0,876980	0,37
14	1	0,964563	0,35
15	1	1,385940	0,34
16	0	1,298730	0,33
17	1	1,320951	0,33
18	1	1,347822	0,33

É fácil perceber que o registro resultante da administração do teste exemplificado na Tabela 12 ilustra vários dos princípios dos testes adaptativos baseados na teoria da resposta ao item. Vejamos:

1. A estimativa de habilidade (teta) aumenta a cada resposta correta e diminui a cada resposta incorreta do participante.
2. As diferenças absolutas entre as sucessivas estimativas de habilidade foram diminuindo na medida em que mais itens foram sendo administrados, indicando

que o teste estava convergindo para um nível de habilidade adequado ao participante.

3. Quanto mais itens eram administrados, mais o erro padrão da medida tendia a diminuir. Como a estimativa de habilidade representa o escore verdadeiro adicionado do erro padrão da medida, de fato, geralmente, uma quantidade maior de respostas leva ao aprimoramento da estimativa de habilidade, ou seja, a porção do escore verdadeiro presente na medida aumenta enquanto a porção do erro diminui.
4. Com o progresso do teste o participante tendeu a alternar entre respostas corretas e incorretas, como pode ser observado com mais clareza a partir do sétimo item. Esse comportamento reflete o processo de convergência que está subjacente ao TAC. Como resultado, normalmente, cada participante irá responder a um conjunto de itens sobre os quais ele dará respostas corretas a cerca de 50% dos itens, mesmo que cada participante responda a conjuntos de itens com diferentes dificuldades. Em certo sentido, esta característica dos testes adaptativos computadorizados tende a igualar o "ambiente psicológico" do teste através de participantes com diferentes níveis de habilidade do traço latente avaliado. Já no teste convencional (não adaptativo) os participantes mais habilidosos responderão à maioria dos itens corretamente e os menos habilidosos responderão à maioria dos itens de forma incorreta.

Comparemos, então, as perspectivas da medição clássica com a medição adaptativa. A medição clássica fixa a quantidade de itens que serão administrados, admitindo que a incerteza (o erro) sobre a estimativa de habilidade poderá variar para cada participante. Em contraste, no teste adaptativo se busca fixar a incerteza que queremos admitir para a estimativa da medida, admitindo que a quantidade de itens administrados pode variar. Como

resultado teremos uma redução significativa no número de itens necessários para produzir a medida e um aumento da precisão de medição.

Outro comportamento esperado quando da aplicação de testes adaptativos é a redução da quantidade de itens administrados até que o programa tenha uma confiança suficientemente forte quanto à medida de habilidade do participante. A Tabela 13 apresenta a quantidade de itens que foram administrados por tipo de prova para produzir a estimativa de habilidade, demonstrando claramente maior eficiência nos testes adaptativos. Em média, para o programa determinar a estimativa final de habilidade nos testes adaptativos foi necessário administrar cerca de 18 itens no caso da prova de raciocínio e de 15 itens no caso da prova de percepção. Como a prova de raciocínio dispõe de 43 itens no total e a prova de percepção 38, conclui-se que houve uma redução significativa, bastando administrar cerca de 40% dos itens para se produzir a estimativa final de habilidade.

Tabela 13 - Quantidade de itens administrados por tipo de prova

Estatísticas descritivas					
	N	Mínimo	Máximo	Média	Desvio padrão
Figuras Incompletas - Linear	129	4	42	34,09	11,827
Comparação de Figuras - Linear	106	9	38	27,95	9,720
Figuras Incompletas - Adaptativa	132	10	24	17,59	4,842
Comparação de Figuras - Adaptativa	121	10	25	14,85	5,118
N válido (de lista)	103				

Em função da aplicação de uma quantidade menor de itens, os tempos para a aplicação das provas também foram reduzidos. O tempo médio de realização da prova de percepção caiu de 06min25s na versão sequencial para 03min43s na versão adaptativa, representando uma redução de 42%. Na prova de raciocínio a redução foi menos significativa, caindo de 05min27s na versão sequencial para 04min41s na versão adaptativa ou 14%. Lembramos que quando aplicados os testes na modalidade sequencial adota-se

como critério de parada quatro erros consecutivos por parte do aluno. No entanto, na prova com lápis e papel existem outros critérios para determinar o início e a parada dos testes que não foram reproduzidos nessa versão informatizada. Por exemplo, no teste com lápis e papel a prova de percepção pode ser finalizada pelo tempo decorrido e no teste de raciocínio a idade determina a dificuldade do item inicial a ser administrado.

Apesar da redução da quantidade dos itens administrados nos testes adaptativos, em geral, os resultados mantiveram-se coerentes com aqueles obtidos nos testes sequenciais. A correlação de Pearson obtida pelo programa SPSS indicou correlações positivas fortes entre as provas sequenciais e suas equivalentes adaptativas, como podemos observar na Tabela 14 adiante. Também se percebe a existência de correlações entre os testes que avaliam processos cognitivos diferentes, mas estas não foram muito significativas.

Adicionalmente, utilizamos o núcleo do programa construído neste estudo para produzir estimativas de habilidades sobre as bases de dados dos estudos 1 e 2, descritos nos Capítulos III e IV. Fizemos dois tipos de estimativas para cada entrada do banco de dados: na primeira calculamos a proficiência do respondente pelo teste total, ou seja, considerando todas as suas respostas; na segunda estimativa simulamos funcionamento normal do teste adaptativo, mas as respostas para cada item eram fornecidas ao programa de acordo com as respostas existentes nas bases de dados produzidas durante aqueles estudos. Ao final, utilizamos o SPSS para calcular a correlação de Pearson entre as estimativas e encontramos correlações positivas muito fortes. No caso da base de dados do estudo 2, que produziu um instrumento informatizado sequencial baseado no TRI, a correlação ficou acima de 900, num sinal de que o teste adaptativo é capaz de produzir escores comparáveis tanto para o modelo sequencial quanto para o modelo de lápis e papel.

Tabela 14 – Determinação da correlação de Pearson entre as provas sequenciais e adaptativas de raciocínio e percepção.

Correlações entre as provas adaptativas e lineares					
		Percepção Linear	Percepção Adaptativa	Raciocínio Linear	Raciocínio Adaptativa
Percepção Linear	Correlação de Pearson	1	,784**	,281**	,285**
	Sig. (1 extremidade)		,000	,002	,002
	N	106	106	103	106
Percepção Adaptativa	Correlação de Pearson	,784**	1	,258**	,320**
	Sig. (1 extremidade)	,000		,002	,000
	N	106	121	118	121
Raciocínio Linear	Correlação de Pearson	,281**	,258**	1	,879**
	Sig. (1 extremidade)	,002	,002		,000
	N	103	118	129	129
Raciocínio Adaptativa	Correlação de Pearson	,285**	,320**	,879**	1
	Sig. (1 extremidade)	,002	,000	,000	
	N	106	121	129	132

** A correlação é significativa no nível 0,01 (1 extremidade).

As principais estatísticas globais sobre itens e pessoas do teste de percepção estão apresentadas na Tabela 15, demonstrando um teste razoável, com índices de fidedignidade elevados, infits e outfits médios aceitáveis, medidas em torno de zero e erro padrão baixo.

Tabela 15 - Sumário das medidas de alunos e itens na prova de percepção.

SUMÁRIO DA MEDIDA DE 132 ALUNOS - PERCEPÇÃO								
	ESCORE TOTAL	CONTAGEM	MEDIDA	ERRO DO MODELO	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MÉDIA	17.3	38.0	-.23	.43	1.00	.0	.97	.0
DESVIO PD	6.8	0.0	1.26	.06	.26	1.2	.43	.9
MÁXIMO	32.0	38.0	2.36	.81	1.86	3.8	2.47	2.5
MÍNIMO	2.0	38.0	-4.13	.39	.46	-2.8	.29	-2.2
REAL RMSE	.46	DESVIO VERD	1.17	SEPARAÇÃO	2.54	CONFIABILIDADE	.87	
MODELO RMSE	.44	DESVIO VERD	1.18	SEPARAÇÃO	2.69	CONFIABILIDADE	.87	
ERRO PADRÃO DA MÉDIA DO ALUNO = .10								
CORRELAÇÃO ENTRE ESCORE BRUTO E MEDIDA DO ALUNO = .99								
ALFA DE CRONBACH (KR-20) ESCORE BRUTO FIDEDIGNIDADE DO TESTE = .87								
SUMÁRIO DA MEDIDA DE 38 ITENS - PERCEPÇÃO								
	ESCORE TOTAL	CONTAGEM	MEDIDA	ERRO DO MODELO	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MÉDIA	53.9	132.0	-.00	.24	1.00	.0	.96	-.1
DESVIO PD	35.1	0.0	1.65	.05	.12	1.1	.19	.9
MÁXIMO	125.0	132.0	2.27	.41	1.38	2.9	1.56	2.7
MÍNIMO	11.0	132.0	-4.01	.19	.79	-2.3	.57	-2.0
REAL RMSE	.25	DESVIO VERD	1.63	SEPARAÇÃO	6.49	CONFIABILIDADE	.98	
MODELO RMSE	.25	DESVIO VERD	1.63	SEPARAÇÃO	6.629	CONFIABILIDADE	.98	
ERRO PADRÃO DA MÉDIA DO ITEM = .27								
CORRELAÇÃO ENTRE ESCORE BRUTO E MEDIDA DO ITEM = .99								

Quanto à fidedignidade e validade do teste, nenhum dos itens apresentou correlação point-biserial negativa, ressaltando a qualidade dos itens, conforme defendido por Linacre (2008). A consistência do teste também foi confirmada pelo alfa de Cronbach com valor de 0,87. A função de informação do teste junto com as curvas características dos itens e do teste, apresentadas no Anexo D, permitem uma visão gráfica complementar do teste.

A unidimensionalidade dos dados foi verificada através do método da análise dos principais componentes – APC, exemplificado na Tabela 16. Como o propósito da APC de resíduos é explicar a variância ao invés de construir variáveis (como na análise fatorial), primeiro de tudo deve-se verificar se o contraste nos residuais é grande o suficiente para explicar muito da variância. Como o primeiro contraste ficou no considerado nível de ruído

(até 2 eigenvalues), podemos concluir que não há uma segunda dimensão com força suficiente para interferir nas medidas e que podemos prosseguir com a análise.

Tabela 16 - Mapa de dimensionalidade.

TABELA 23.0 Raciocínio		Fev 12 20:28 2015	
ENTRADA: 132 PESSOAS 43 ITENS		WINSTEPS 3.81.0	

Tabela da variância RESIDUAL PADRONIZADA (em unidades de Eigenvalue)			
		-- Observado --	-- Esperado
Total da variância bruta nas observações	= 74.9	100.0%	100.0%
Variância bruta explicada pelas medidas	= 46,2	61.7%	61.4%
Variância bruta explicada pelas pessoas	= 26.4	35.2%	35.4%
Variância bruta explicada pelos itens	= 19.8	26.4%	26.0%
Variância bruta inexplicada (total)	= 28.7	38.3%	38.6%
Variância inexplicada no primeiro contraste	= 1.9	2.5%	6.8%
Variância inexplicada no segundo contraste	= 1.8	2.4%	6.4%
Variância inexplicada no terceiro contraste	= 1.5	2.0%	5.8%
Variância inexplicada no quarto contraste	= 1.2	1.6%	5.1%
Variância inexplicada no quinto contraste	= 1.1	1.3%	4.9%

Olhando para o mapa apresentado na Figura 20, adiante, que compara a dificuldade dos itens do teste de raciocínio e percepção com a estimativa de habilidade dos alunos, podemos observar que o teste de raciocínio dispõe de características psicométricas satisfatórias, apresentando um certo equilíbrio entre as habilidades dos alunos, exibidas no lado esquerdo do mapa, e as dificuldades dos itens, exibidas no lado direito do mapa. A distribuição é suficiente para exibir a média e dois desvios padrão, tanto para os itens quanto para os alunos. Os itens dão cobertura a, praticamente, toda faixa de habilidade das pessoas na amostra, apresentando medidas de dificuldade estimadas entre 2,48 a -4,41 logits. No entanto, apesar da distribuição aceitável dos itens, temos um problema bastante relevante para o caso dos testes adaptativos que é a pequena quantidade de itens. Reparem que não há muitas alternativas para itens com medidas de dificuldade próximas para as diferentes faixas de

dificuldade. A situação se agrava para o teste de percepção, pois, além da baixa quantidade de itens, existe maior irregularidade na distribuição dos itens ao longo da escala, apresentando lacunas em algumas faixas de dificuldade. A reduzida quantidade de itens disponíveis para as duas provas reflete situações indesejadas, que somente poderiam ser contornadas com a ampliação do banco de itens.

É claro que o padrão de respostas dos participantes também é fator decisivo para a convergência do cálculo da estimativa de habilidade para próximo do escore verdadeiro. Porém, em situações como as observadas aqui, o teste adaptativo poderá ter dificuldade para selecionar os itens apropriados e até mesmo não ser capaz de convergir a estimativa de habilidade para um ponto próximo do escore verdadeiro da pessoa. Além disso, fatalmente, haverá grande exposição dos itens para os participantes, tornando-os conhecidos do público, o que não é desejável.

começar a se ajustar à capacidade do aluno. Ressalta-se que a única diferença entre os dois tipos de testes é o banco de itens, pois ambos utilizam o mesmo programa de computador, baseados no mesmo algoritmo de teste adaptativo.

Porém, devemos estar cientes que o estudo com testes adaptativos baseados nas provas da ECCOs 4/10 não pode ter seus resultados confrontados cegamente com aqueles produzidos pela versão tradicional, em lápis e papel, pois existem diferenças conceituais, estruturais e operacionais entre essas versões. Dessa forma, os testes produzidos neste estudo de teste adaptativo também não têm a pretensão de substituir as provas originais, mas demonstrar a viabilidade de fazê-lo, no futuro, com a ampliação e calibração adequada do banco de itens. No entanto, como ambas as versões apontam para uma mesma direção, acreditamos que os procedimentos metodológicos e os achados deste trabalho possam colaborar com a modelagem de testes cada vez mais eficientes e atraentes.

Na prática, durante a aplicação dos testes sentimos que mesmo um teste que é percebido pelas crianças como um jogo no computador é capaz de gerar um certo estresse, uma espécie de desconforto, em várias delas. Diversas crianças externaram ansiedade, fosse através de atitudes pouco naturais, pela passividade inicial ou pelas mãos frias ou humedecidas.

Assim como verificamos no teste produzido no estudo 2, em 2011, as crianças não tiveram qualquer dificuldade em se relacionar com os computadores, até porque, como trabalhamos apenas com crianças de escolas particulares destinadas às classes A, B e C, o computador já é utilizado por quase todas elas, em casa e na escola. Mesmo assim, o uso de computadores com telas sensíveis ao toque foi um diferencial que representou uma decisão acertada. Sem dúvida, poder “brincar” com a tela motivou a participação das crianças e eliminou qualquer eventual dificuldade de operação do mouse ou de outros apontadores.

O teste adaptativo se mostrou bastante eficiente, conseguindo convergir para estimativas de habilidades próximas dos escores verdadeiros dos alunos, reduzindo em mais de 50% a quantidade de itens administrados, o tempo de aplicação do teste e, conseqüentemente, evitando que o aluno ficasse cansado ou entediado durante o teste. No entanto, algumas preocupações e questionamentos quanto à validade de sua utilização em certos casos são inevitáveis.

Em primeiro lugar, havemos de considerar que o estudo contempla crianças, muitas de tenra idade. Durante a administração dos testes podíamos perceber a frustração de algumas delas, principalmente as de menor idade, quando, ao passar pelos itens dos testes, se deparavam com itens com nível de dificuldade muito acima da sua capacidade. Não falamos só da capacidade de conseguir entender e resolver a tarefa que ora se apresentava, mas, também, da capacidade de compreender o que significavam aquelas imagens que apareceram na tela, cujas formas elas desconhecem e, sem saberem do que se trata, na maioria das vezes não conseguem nem sequer esboçar uma tentativa de acertar.

Ora, vejamos a situação: o teste acabou de começar e a criança já não sabe o que fazer com ele... A criança olha para o pesquisador na tentativa de receber alguma ajuda, mas o apoio que lhe chega é insuficiente para sua compreensão do problema, de forma que ela acaba por responder, timidamente, qualquer coisa e passa para o próximo item. O pior é que talvez ela também não entenderá bem o segundo ou o terceiro item. Claro, teste adaptativo é assim mesmo, ele vai percebendo que o participante está errando e vai procurando adaptar o nível de dificuldade dos itens à estimativa de habilidade da pessoa. A questão é que se dispomos de um banco de itens muito pequeno, como nestes casos, com itens muito separados em dificuldade, ou com irregularidades na representação na escala, antes mesmo que as estimativas de habilidades consigam se aproximar do escore verdadeiro daquela pequena criança, talvez ela já não queira mais “brincar”.

Então, tal como suposto pelos procedimentos metodológicos do TAC, havemos de utilizar sempre bancos de itens especialmente otimizados, em particular no caso de aplicações com crianças. Isso significa dizer que precisamos de bancos de itens maiores, contendo tarefas que possam ser distribuídas uniformemente ao longo de toda a escala de dificuldade. De fato, isso ajudaria muito, mas, no caso da ECCOs 4/10, temos uma escala que se destina a avaliar crianças cujas idades variam dos 4 aos 10 anos. Portanto, temos pessoas em fases bem distintas de desenvolvimento, contemplando os estágios pré-operatório e operatório-concreto descritos por Piaget. A recomendação mais comum para seleção do primeiro item de um teste adaptativo é que ele esteja próximo do centro da escala. No nosso teste, optamos por usar um nível menor, em torno de 30% da dificuldade da escala, mas, se considerarmos, por exemplo, que na prova de raciocínio temos 43 itens ordenados pela dificuldade, os itens próximos a 30% da escala serão os itens de 11 a 14 – na verdade, tratam-se de itens que podem ser muito difíceis para crianças com 4 ou 5 anos de idade. Óbvio que poderíamos adotar critérios baseados na idade do aluno para estabelecer o item inicial, como alguns pesquisadores sugerem, mas, mesmo assim, o método *expected a posteriori* (EAP) irá, provavelmente, determinar um nível de dificuldade bem mais alto para o segundo item, pois no início do teste a quantidade de itens administrados ainda é muito pequena para produzir boas estimativas de habilidades. Nesse caso, nossa avaliação é que precisaríamos, realmente, dispor de bancos de itens separados por faixa etária ou utilizar estratégias multidimensionais, com subescalas de habilidades compatíveis com a idade, o que também envolveria uma quantidade maior de itens.

Outro ponto a considerar que envolve crianças pequenas, entre 4 e 6 anos, principalmente, mas que também pode alcançar as crianças que tenham um desempenho muito alto ou um desempenho muito baixo, é que o teste, tal como está, pode ser rápido demais, realizado em muito pouco tempo. Em muitos casos, vimos que o teste estava

concluído após a administração de apenas dez itens, mesmo assim por causa de uma exigência técnica imposta ao programa. Assim, não foi raro o teste ser encerrado pelo programa após pouco mais de um minuto de duração. Ótimo, novamente, essa era mesmo a ideia para o teste adaptativo, ou seja, encerrar quando atingir um nível de segurança suficiente para a estimativa da medida. Mas, nessas horas, para o participante que se prontificou a colaborar com a pesquisa, pareceu-nos que restava uma sensação de dúvida, de incredulidade, de que ele fez ou deixou de fazer alguma coisa. Como? Já acabou? Era só isso? Eu fui bem no teste? Essas perguntas eram inevitáveis. Nesse sentido, após as primeiras aplicações, sentimos a necessidade de explicar aos participantes antes do início dos testes que não ficassem surpresos se o teste fosse concluído em um tempo pequeno.

Considerações gerais e Recomendações

7. CONSIDERAÇÕES GERAIS E RECOMENDAÇÕES

Na esteira da evolução dos testes em benefício da psicometria verificamos que o atual estado da arte está representado pelos testes adaptativos computadorizados baseados na teoria da resposta ao item. Também vimos que itens multimídia e interativos poderão se tornar uma prática regular num futuro breve. Por sua vez, destaca-se que os testes em lápis e papel vêm cumprindo sua missão com galhardia há muito tempo. Só olhando para a história dos testes adaptativos, já se passaram cerca de cento e dez anos desde que Binet aplicou seus testes para avaliação da inteligência.

De fato, justifica-se, os testes em lápis e papel podem ser muito flexíveis. Embora, normalmente, eles sejam estruturados para uma quantidade fixa de itens que são apresentados sequencialmente, em muitos casos são utilizadas estratégias que os tornam adaptativos combinando critérios de sequenciamento, de início, de finalização ou de temporização, como no caso da própria ECCOs 4/10 original, por exemplo. Além disso, mesmo os testes aplicados no modelo lápis e papel costumam ter os dados das respostas analisados com auxílio dos computadores. Assim, eles podem estar fundamentados tanto na teoria clássica dos testes, caso em que importa o desempenho global no teste, quanto na teoria da resposta ao item, quando o foco recai sobre cada item do teste e nos permite fazer previsões, estimativas ou inferências sobre as habilidades dos participantes.

Por sua vez, os testes informatizados, aqueles aplicados com o uso do computador, começaram reproduzindo grande parte dos testes no formato lápis e papel e foram incorporando vantagens importantes, tal como a automação do registro das respostas e dos dados cadastrais dos participantes diretamente em bases de dados eletrônica. Mas isso era só o começo. A evolução dos testes informatizados pegou carona no grande avanço da indústria de computadores verificado a partir dos anos 1970 em função da chegada dos circuitos

integrados com larga escala de integração (LSI) que viriam a revolucionar as aplicações da microeletrônica na sociedade global (Famm, 1988, p. 237). O aumento da capacidade de processamento e armazenamento, a redução do tamanho dos equipamentos e dos seus custos de produção fizeram despencar o preço de comercialização dos computadores, tornando-os recursos de fácil acesso e grande portabilidade. Para se ter uma ideia, os computadores, tablets e smartphones disponíveis nas residências e instituições atualmente são capazes de realizar bilhões de operações por segundo - apenas como referência, em 2014, o supercomputador Tianhe-2 foi considerado o computador mais rápido do mundo, podendo executar hipotéticas 33,86 quatrilhões de operações por segundo ("Flops," n.d.).

Essas condições viabilizaram a produção de testes computadorizados inteligentes, capazes de dar conta de algoritmos matemáticos complexos, como os da TRI, tratar grandes massas de dados, administrar e até mesmo produzir itens relevantes para os participantes durante o teste, calcular estimativas de habilidades a cada resposta produzida, avaliando continuamente o desempenho do indivíduo, tomar decisões em funções dessas respostas, verificar adequação de modelos estatísticos, enfim, aprimorar o teste para ajustá-lo à capacidade e ao comportamento que o seu usuário vai demonstrando ao longo da execução do próprio teste.

Um ponto relevante em todas as situações foi que considerando a utilização de notebooks ou tablets para aplicação dos testes informatizados, sejam eles sequenciais ou adaptativos, a portabilidade dos testes passou a ser muito grande, permitindo que o teste vá aonde as crianças estejam. O fato de o teste poder ser itinerante, acompanhando o profissional, pode ser muito útil quando pensamos na aplicação dos testes em escolas, principalmente nas grandes redes públicas, ou mesmo com enfermos em hospitais.

No primeiro capítulo ficou demonstrado que a utilização de testes informatizados baseados na TRI para auxiliar na elaboração, análise e pontuação de testes é crescente em

todo o mundo, inclusive no Brasil, já sendo, atualmente, o paradigma mais utilizado para testes de certificação de proficiência e avaliação do aprendizado, principalmente em provas orientadas para grandes grupos. A Prova Brasil, o Exame Nacional do Ensino Médio (ENEM) e o Exame Nacional de Avaliação do Desempenho dos Estudantes (ENADE), no Brasil, o Scholastic Assessment Test (SAT), o Graduate Record Examination (GRE) e o Trends in International Mathematics and Science Study (TIMSS), nos Estados Unidos, assim como o Programme for International Student Assessment (PISA) mantido pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE), na Europa, são alguns exemplos nacionais e internacionais muito bem sucedidos de aplicação da TRI.

Já quanto ao uso de testes adaptativos no Brasil, este ainda é incipiente, em especial em benefício da psicologia, mas se trata de um caminho sem volta. Aos poucos, iniciativas nesse sentido estão aparecendo, quer através de empresas estrangeiras, quer através de desenvolvimento nacional público e privado, como vimos no primeiro capítulo. Dentro das ciências humanas, as áreas ligadas a certificações de proficiências, de seleção de candidatos e de monitoramento da aprendizagem, incluindo os de auto avaliação, concentram grande parte dos testes adaptativos computadorizados existentes ou em desenvolvimento baseados na TRI.

De uma forma geral, os testes informatizados têm mostrado que podem prover portabilidade, melhor e maior controle sobre os instrumentos, facilidade de uso, redução da fadiga, autonomia para o participante e até motivação adicional para as pessoas que os realizam. Aliás, este vem sendo um ponto de interesse em muitas pesquisas atualmente. A influência de forças interdisciplinares decorrentes da evolução da ciência cognitiva, estatística matemática, psicologia educacional e de ciências da computação estão começando a aparecer para beneficiar a avaliação educacional e psicológica. Sem alarde, os testes informatizados estão evoluindo e ganhando abordagens inovadoras, novas roupagens que prometem transformar a forma tal como os conhecemos atualmente.

Os testes adaptativos são exemplos vivos dessa evolução. Na forma não informatizada eles existem desde o início do século XX. Ao longo do tempo, diversas estratégias foram utilizadas para tentar adaptar os testes às capacidades individuais. Mesmo nas provas da ECCOs 4/10 realizadas em lápis e papel procurou-se dar algum tratamento adaptativo, sempre que possível, tal como utilizar a idade do participante para determinar o item inicial ou encerrar o teste após uma sucessão de quatro erros. De fato, aplicar um mesmo teste para pessoas com níveis diferentes de habilidades, interesses ou até comprometimentos físicos ou psicológicos pode ser injusto e prejudicar o resultado das avaliações, além de cansativo, desmotivante ou até mesmo estressante para quem os realiza. O uso de computadores revolucionou o percurso dos testes adaptativos, propiciando o redesenho de instrumentos de medições psicológicas e educacionais e sendo rapidamente indicado para testes de mensuração de habilidade ou realização, assim como para medidas de personalidade e variáveis atitudinais.

Tudo indica que a aprendizagem adaptativa terá um grande impacto na forma como estudamos e aprendemos, pois através das avaliações produzidas nos computadores cada aluno pode compreender onde está posicionado na rota da aprendizagem desejada e identificar o que ele precisa fazer para atingir seus objetivos, sendo levado a estudar aquilo que realmente importa para complementar ou aprofundar a sua formação. Modelos para aprendizagem adaptativa são componentes essenciais para o ensino a distância e modelos b-learning (blended learning) de ensino, no qual parte da aprendizagem é feita a distância e parte presencial, sendo os testes adaptativos computadorizados um dos pilares desses modelos.

Acredita-se que testes mais envolventes e representativos levem a estimativas mais precisas das habilidades dos participantes, haja vista que, em teoria, eles realizarão os testes mais concentrados e produzirão respostas mais adequadas. Nesse sentido, algumas propostas

são de fácil implementação e baixo impacto, quase invisíveis, como a de alterar a dinâmica dos testes para incluir pseudo-itens fáceis visando aumentar a quantidade de acertos apenas para não deixar cair a motivação do participante. Outras propostas são mais arrojadas, como a de produção de itens multimídia que se utilizem de vídeos e gráficos interativos, como os estudos que estão sendo desenvolvidos pelo Smarter Balanced Assessment Consortium, descrito no capítulo I, que buscam o engajamento de estudantes com um olho no flow, o estado psicológico caracterizado por intensa concentração e prazer que representa uma ótima experiência para o participante (Csikszentmihalyi, 1990) e norteia a construção dos jogos eletrônicos.

Na University of Alberta, no Canadá, o Centre for Research in Applied Measurement and Evaluation vem trabalhando em um modelo de teste que denominou Computer Adaptive-Attribute Testing (CA-AT) (Gierl & Zhou, 2008), cujos conceitos e procedimentos nos permitem fundir os benefícios da administração de testes adaptativos computadorizados com os benefícios psicológicos da avaliação cognitiva diagnóstica para produzir um método inovador de teste adaptativo psicologicamente amparado (psychologically-based adaptive testing) que pode ser usado para promover a avaliação formativa em sala de aula usando o computador.

Embretson (2005), em artigo intitulado *Measuring Human Intelligence with Artificial Intelligence*, propõe utilizar um método remanescente dos programas eletrônicos de xadrez para propiciar a geração adaptativa de itens. Nesse modelo, os itens ideais para cada participante são construídos durante o próprio teste a partir de fragmentos existentes em um banco de dados, requerendo a integração de duas linhas de pesquisas: os métodos psicométricos para os testes adaptativos e a análise cognitiva dos itens.

Alinhado com esses estudos, queremos destacar que a prova *Desenhos Absurdos*, onde se busca identificar a percepção pela criança de absurdos dentro de contextos sociais,

tem um enorme potencial para ser totalmente automatizada e utilizar itens multimídia, com movimentos e sons, estratégias que, como vimos, estão ganhando força no seio da psicometria e que visam ser mais envolventes e naturais do que os itens estáticos.

Os estudos aqui relatados demonstraram que é inegável que o teste adaptativo computadorizado funciona, que pode entregar o que promete e ser muito útil e eficiente para muitas aplicações. No entanto, as observações levantadas nos resultados do capítulo V, sobre o desenvolvimento e aplicação do teste adaptativo, nos levam a reflexões sobre o quão cuidadoso devemos ser no planejamento de um teste adaptativo, pois essa atenção será preditora do fracasso ou do sucesso e aceitação do teste. Precisamos considerar a sua finalidade, o público alvo e até definições específicas do algoritmo para cada caso.

Algumas vantagens dos testes adaptativos computadorizados (TAC) sobre os testes de tamanho fixo, sejam eles informatizados ou não, são evidentes e devem ser salientadas.

Vejamos:

- No TAC, podemos fixar a margem de erro que desejamos aceitar para a medida e admitir que a quantidade de itens administrados irá variar a cada participante. Já nos testes de tamanho fixo, ao contrário, o número de itens administrados é o mesmo para todos os participantes, admitindo-se a variação da certeza sobre a medida produzida. Essa diferença de perspectiva representa uma grande vantagem para o TAC, pois significa que ele produzirá medidas mais ajustadas para todos os participantes, independentemente dos níveis de habilidade.
- No TAC o computador apresentará itens que são mais relevantes para cada participante, produzindo medidas ajustadas para todas as faixas de habilidade, enquanto o teste de tamanho fixo tenderá a produzir medidas com pouco erro para pessoas que tenham nível mediano de habilidade, aumentando o erro da medida

quanto mais o escore verdadeiro do participante se aproxime dos extremos da escala.

- Em geral, o TAC produz estimativas de medidas confiáveis administrando-se menos itens do que nos testes de tamanho fixo, levando-nos a um pensamento imediato de que os testes serão aplicados em um tempo menor. Isso é verdade, mas a questão principal é o que essa redução do tempo significa para as pessoas envolvidas na realização e aplicação do teste, pois pode estar associada à redução da fadiga ou mesmo dor, principalmente na clínica psicológica ou médica, levando a resultados mais confiáveis, justamente por serem obtidos de maneira mais confortável para pacientes e profissionais da saúde.

Por outro lado, a impossibilidade de revisar os itens e até alterar respostas já dadas aos itens é considerado um problema significativo do TAC, constituindo uma desvantagem significativa em relação às provas em lápis e papel. Acontece que como cada item é selecionado de acordo com a habilidade estimada e que o cálculo dessa estimativa considera todas as respostas anteriores a cada momento, qualquer resposta modificada implicaria em recalculas as estimativas de habilidades para apresentar novos itens, levando instabilidade às estimativas.

A oferta de ferramentas para criação de testes adaptativos também vem aumentando paulatinamente. A biblioteca catR, utilizada no estudo 3 deste trabalho, por exemplo, dispõe de diversas funções que facilitam criação de testes adaptativos usando a linguagem de programação R, sendo um exemplo de recurso útil. Também já estão disponíveis estruturas complexas de código aberto, tal como a plataforma online Concerto (<http://www.psychometrics.cam.ac.uk/newconcerto>), desenvolvida, disponibilizada e mantida pelo The Psychometrics Centre da University of Cambridge.

Uma rigorosa avaliação para seleção das ferramentas mais adequadas para o desenvolvimento de testes adaptativos deve ser feita para cada caso, pois não existe uma única “melhor alternativa”, haja vista envolver diversas competências e necessidades técnicas e de aplicabilidade.

Também ressaltamos a percepção de Bugbee e Bernt (1990), abordada durante o aprofundamento teórico deste trabalho, de que os testes computadorizados constroem mais os examinandos do que os testes de papel e lápis. Estudos também sugerem que demoramos mais para ler ou detectar erros na tela do computador do que em materiais impressos. Além disso, de uma forma geral, versões adaptativas computadorizadas não admitem a utilização de estratégias simples por parte dos examinandos, como sublinhar partes do texto, riscar opções eliminadas ou pular questões para voltar a elas mais tarde.

Outro ponto importante que deve ser analisado com muito cuidado no planejamento de um teste adaptativo refere-se à finalidade do teste e às estratégias para sua aplicação. Testes para seleção de vagas, para certificação de proficiências ou para avaliação rotineira da aprendizagem escolar têm características muito diferentes e envolvem momentos psicológicos bem distintos, merecendo, da mesma forma, tratamento diferenciados. Note-se, por exemplo, que testes cujo tempo de realização possa ser muito longo para alguns e muito curto para outros podem fazer com que os examinandos prevejam os resultados em função desse tempo, de maneira acertada ou não, fato que pode ser indesejável em muitas situações.

Voltando-nos à ECCOs 4/10, frente às possibilidades que os testes computadorizados proporcionam, em especial empregando a teoria da resposta ao item e recursos adaptativos, podemos dizer que ainda temos um longo caminho pela frente no sentido da modernização dessa escala. Na análise inicial que fizemos para definir quais das provas da ECCOs 4/10 poderíamos informatizar, vimos que para algumas provas o recurso computacional poderia ser muito limitante enquanto para outras extremamente libertador. Algumas provas não

verbais admitem grande autonomia por parte do participante na sua realização, podendo, praticamente, dispensar a interação com o administrador dos testes e sendo aquelas mais indicadas para utilização do computador, tais como as provas Figuras Incompletas, Composição de Padrões, Comparação de Figuras ou mesmo Elementos em Árvores. Percebemos, ainda, que a prova Desenhos Absurdos tem potencial para utilização de itens multimídia e itens interativos, mas precisaria ser remodelada. Já em relação às provas verbais, onde se requer maior interação com o pesquisador, não se pode dizer o mesmo, a menos que o recurso de informática fosse destinado apenas ao registro das anotações do administrador dos testes ou para a exibição de situações. Para esses casos parece não ser possível garantir que as condições ideais de realização do teste de forma autônoma pelo examinando fossem respeitadas, uma vez que o aluno precisaria ouvir frases geradas pelo computador e o computador precisaria compreender a resposta falada pelo aluno e até interpretar suas reações. De fato, já dispomos de tecnologias que permitiriam ao computador interagir com o aluno dessa forma, “falando”, “ouvindo” e “interpretando”, mas o ambiente teria que ser extremamente controlado e o teste sujeito a fortes intervenções por parte do pesquisador.

Adicionalmente, pudemos perceber que na ECCOs 4/10 alguns itens das provas Figuras Incompletas, que avalia o raciocínio, e também na prova Desenhos Absurdos, que avalia o processo de compreensão, usam imagens representativas do cotidiano que estão, de certa forma, “envelhecendo”, sendo relativamente comum que a criança não consiga fazer a associação das imagens de alguns itens com o raciocínio pretendido. Por exemplo, percebemos certa dificuldade para muitos participantes distinguirem linhas de modernidade ou de antiguidade em figuras de automóveis ou aviões que foram desenhadas há anos e não se parecem muito condizentes com o que temos no cotidiano dessas crianças, nascidas somente a partir de 2003. Da mesma forma que os desenhos animados, super-heróis e até mesmo alguns brinquedos de hoje não guardam semelhança com os de 15 ou 20 anos atrás,

entendemos que a ECCOs precisará atualizar seu estilo gráfico em pouco tempo, principalmente porque em muitos casos ela está interessada em promover relações com o contexto social da criança e precisa oferecer condições para que essas relações possam ser reconhecidas e estabelecidas.

Nesse sentido, ficam como sugestões para continuidade e ampliação desse trabalho em benefício da ECCOs 4/10:

- A revisão dos gráficos de todos os itens no sentido de atualizá-los para os tempos atuais, sempre que se notar necessário;
- Realizar um estudo voltado ao aumento de motivação das crianças com base na prova Desenhos Absurdos, através da revisão, criação e introdução de itens multimídia interativos;
- Produzir uma versão adaptativa da prova Desenhos Absurdos reestruturada;
- Consolidar o uso dos testes adaptativos, através da ampliação do banco de itens das provas Comparação de Figuras e Desenhos Incompletos e da introdução de estratégias multidimensionais.

Por fim, queremos registrar que uma página na internet, especificamente para suporte ao projeto de informatização da escala, está sendo construída a partir do endereço <http://www.renatoguedes.com.br> e reafirmar que os conhecimentos adquiridos referentes aos estudos realizados ao longo dessa trajetória, que envolveu a construção e validação de escalas psicométricas e a teoria da resposta ao item, culminando, agora, na construção de testes adaptativos computadorizados, podem ser replicados rapidamente para outros estudos e aplicações, vindo a colaborar com a modelagem de testes cada vez mais eficientes e atraentes para as pessoas.

Vale também registrar que o tema ora abordado tem feito com que as pesquisas e os resultados, mesmo parciais, dos estudos aqui relatados venham gerando interesse na comunidade acadêmica e estejam sendo aceitos e apresentados em congressos nacionais e internacionais, tais como no VIII Congresso Ibero-americano de Avaliação Psicológica

(2011), em Lisboa, Portugal, no XIV Encuentro Internacional Virtual Educa (2013), em Medellín, Colômbia, e no IX Congresso Brasileiro de Psicologia do Desenvolvimento (2013), em João Pessoa, Brasil.

Referências

REFERÊNCIAS

- Almeida, L. S. (1988). Teorias da inteligência. Porto: Edições Jornal de Psicologia.
- Almeida, L. S. (1994). Inteligência: Definição e medida. Aveiro: Centro de Investigação, Difusão e Intervenção Educacional.
- Almeida, L. S. (2008). Constructo e medida da inteligência: Contributos da abordagem factorial. In Candeias, A.; Almeida, L.; Roazzi, A. & Primi, R. (Org.). Inteligência: Definição e medida na confluência de múltiplas concepções. São Paulo: Casa do Psicólogo.
- Aldrich, John (1997). R. A. Fisher and the Making of Maximum Likelihood 1912 – 1922. *Statistical Science*, Vol. 12, No. 3, 162-176.
- Andrich, D. (1978) Relationships between the Thurstone and Rasch approaches to item scaling. *Applied Psychological Measurement*, 2, 449-460.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory*. New York: Academic Press.
- Ayala, R. J. (2009). *The Theory and practice of item response theory*. Nova York, NY : The Guilford Press.
- Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Recent advances in learning and motivation*. New York: Academic Press.
- Baker, F. D. (1992) *Item response theory: parameter estimation techniques*. New York, NY: Marcel Dekker.
- Bayes' theorem. (n.d.). In Wikipedia. Acessado em 16 de dezembro de 2014, em http://en.wikipedia.org/wiki/Bayes%27_theorem
- Bayliss, D. M., Jarrold, C., Gunn, D., & Baddeley, A (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, 132 (1), 71-92.
- Binet, A., & Simon, Th. A. (1905). Méthode nouvelle pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique*, 11, 191-244.
- Blended_learning. (n.d.). In Wikipedia. Acessado em 12 de dezembro de 2014, em http://en.wikipedia.org/wiki/Blended_learning
- Bond, Trevor G., & Fox, Christine M. (2007). *Applying The Rasch Model: Fundamental Measurement in the Human Sciences*. New Jersey – EUA: Lawrence Erlbaum Associates, Inc.

- Bracken, B. A., & Walker, K. C. (1997). The utility of intelligence tests for preschool children. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues*. New York: Guilford Press, pp. 484-502.
- Bridgman, P. W. (1928). *The logic of modern physics*. New York: Macmillan.
- Brito, L. (2009). *Desenvolvimento e realização cognitiva na infância: Construção e validação da ECCOs 4\10*. 244 f. Tese (Doutorado em Psicologia). Orientador Prof. Dr. Leandro da Silva Almeida. Universidade do Minho, Instituto de Educação e Psicologia, Braga, Portugal.
- Brito, L., & Almeida, L. (2009). *Manual da Escala de Competências Cognitivas para Crianças (ECCOs 4/10)*.
- Brito, L., Almeida, S. L., Viola, L., & Chaves, S. (2006). *Avaliação da memória: Estudo comparativo com provas de duas baterias de avaliação compósita da inteligência para crianças*. Actas da 11ª Conferência Internacional de Avaliação Psicológica: Formas e Contextos. Braga: Psiquilibrios
- Brody, E. B., & Brody, N. (1976). *Intelligence. Nature, determinants and consequences*. New York: Academic Press.
- Bugbee, A. C., Jr., & Bernt, F. M. (1990). Testing by computer: Findings in six years of use. *Journal of Research on Computing Education*, 23, 87-100.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 629-637.
- Cattell, R. B. (1971). *Intelligence: Its structure, growth and action*. Boston: Houghton Mifflin.
- Chi, M. T. H., & Glaser, R. (1992). A capacidade para a solução de problemas. In R. Sternberg (Ed.), *As Capacidades Intelectuais Humanas*. Porto Alegre: Artes Médicas.
- Csikszentmihalyi, Mihaly. (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper and Row.
- Cohen, E. H., & Amar, R. (1999). External variables as points in SSA: A comparison with the unfolding techniques. In R. Meyer Schweizer, D. Hänzi, B. Jann, E. Peier-Kläntschi & H. J.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- da Silva, A., Mourão, L., Meneses, P. P. M., Zerbini, T., Borges-Andrade, J. E., & Vilas-Boas, R. V. G. (2012). *Medidas de Avaliação em Treinamento, Desenvolvimento e Educação: Ferramentas para gestão de pessoas*. Artmed Editora

- Davis, Michelle R. (2012, outubro). Adaptive Testing Evolves to Assess Common-Core Skills. *Education Week*, 06, 01, 12-16. Princeton, New Jersey. Disponível em <http://www.edweek.org/dd/articles/2012/10/17/01adaptive.h06.html>
- DeVellis, R. F. (2003). *Scale development: theory and applications* (second edition). Estados Unidos: Sage Publications, Inc.
- Embretson, S. E. (2005). Measuring Human Intelligence with Artificial Intelligence: Adaptive Item Generation. in Sternberg, R. J., & Pretz, J. E. (2005). *Cognition and Intelligence: Identifying the Mechanisms of the Mind*. (pp. 251-267). New York. Cambridge University Press.
- Field, A. (2009). *Discovering statistics using SPSS* (third edition). Dubai: Oriental Press.
- Fliege, H., Becker, J., Walter, O. B., Rose, M., Bjorner, J. B., & Klapp, B. F. (2009). Evaluation of a computer-adaptive test for the assessment of depression (D-CAT) in clinical application. *Int J Methods Psychiatr Res.*;18(1):23-36. doi: 10.1002/mpr.274. PMID: 19194856
- Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45(1), 30-45.
- Flops. (n.d.). In Wikipedia. Acessado em 16 de dezembro de 2014, em <http://pt.wikipedia.org/wiki/FLOPS>
- French, C. C., & Colman. A. M. (1995). *Cognitive Psychology*. London: Longman.
- Garcia-Martinez, J., Rodrigues-Gonzalez, J. M., & Blanco-Picabia, A. (1997). Evaluacion de la Memoria y la Atencion. In *Manual de Evaluación Psicológica: Fundamentos, técnicas y aplicaciones*. Madrid: Siglo Veintiuno de España Editores, S. A., pp. 631-657.
- Gibbons, R. D., Weiss, D. J., Kupfer, D. J., Frank, E., Fagiolini, A., Grochocinski, V. J., ... Immekus, J. C. (2008). Using Computerized Adaptive Testing to Reduce the Burden of Mental Health Assessment. *Psychiatric Services* (Washington, D.C.), 59(4), 361–368. doi:10.1176/appi.ps.59.4.361
- Gibbons, R. D., Weiss, D. J., Pilkonis, P. A., Frank, E., Moore, T., Kim, J. B., & Kupfer, D. J. (2014). Development of the CAT-ANX: A Computerized Adaptive Test for Anxiety. *The American Journal of Psychiatry*, 171(2), 187–194. doi:10.1176/appi.ajp.2013.13020178
- Gierl, J. Mark., & Zhou, J. (2008) Computer Adaptive-Attribute Testing: A New Approach to Cognitive Diagnostic Assessment *Zeitschrift für Psychologie / Journal of Psychology* 2008; Vol. 216(1):29–39. Hogrefe & Huber Publishers.
- Green, B. F. (2000). System design and operation. In Wainer, H. (Ed.) *Computerized Adaptive Testing: A Primer*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Grieve, J. (1995). *Neuropsicologia: avaliação de la percepção y de la cognición*. Buenos Aires: Editorial Médica Internacional.
- Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 33, 469-504.
- Guttman, L. (1991). *Louis Guttman: In memoriam - Chapters from an unfinished textbook on facet theory*. Jerusalém, Israel: The Israel Academy of Sciences and Humanities.
- Habib, M. (2003). *Bases neurológicas dos comportamentos*. Lisboa: CLIMEPSI, 2ª ed.
- Hambleton, R. K., Swaminathan, H. & Rogers, J. (1991). *Fundamentals of item response theory*. Beverly Hills, CA: SAGE.
- Hart, D. L., Mioduski, J. E., Werneke, M. W., Stratford, P. W. (2006). Simulated computerized adaptive test for patients with lumbar spine impairments was efficient and produced valid measures of function. *J Clin Epidemiol*. 2006 Sep; 59(9):947-56. Epub 2006 Jul 11. PubMed PMID: 16895818.
- Henning, G. (1987). *A guide to language testing: development, evaluation, research*. Cambridge, MA: Newbury House.
- Hicks, Marilyn M. (1989). *The TOEFL Computerized Placement Test: Adaptive Conventional Measurement*. Research Report 31. Educational Testing Service (ETS).
- Hogan, Thomas P. (2006). *Introdução à prática de testes psicológicos; Tradução de Luís Antônio Fajardo Pontes ; revisão técnica Francisco Donizetti Mendes Takahashi – Rio de Janeiro : LTC*.
- Hsueh, I. P., Chen, J. H., Wang, C. H., Chen, C. T., Sheu, C. F., Wang, W. C., Hou, W. H., & Hsieh, C. L. (2010). Development of a computerized adaptive test for assessing balance function in patients with stroke. *Phys Ther*. 90(9):1336-44. doi:10.2522/ptj.20090395. Epub 2010 Jun 30. PubMed PMID: 20592269.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Kaufman, A. S. (1979). *Intelligence testing: Principles, applications, and issues*. Belmont: Brooks/Cole Publishing, 2nd ed.
- Kingsbury, G. G., & Zara, A. R. (1989). Procedures for selecting items for computerized adaptive tests, In: *Applied measurement in education*, 1989, pp. 16.
- Kintsch, W. (1998). *Comprehension: a paradigm for cognition*. Cambridge: Cambridge University Press.

- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.;
- Lazarsfeld, Paul F. (1959) "Latent Structure Analysis", in *Psychology: A Study of a Science*, Vol. 3, S.
- Linacre, J. M. (1998). Detecting multidimensionality: which residual data-type works best? *Journal of Outcome Measurement*, 2, 3, 266-283.
- Linacre, J. M. (2000). *Computer-adaptive testing: A methodology whose time has come* (MESA Memorandum No. 69), MESA Psychometric Laboratory, University of Chicago.
- Linacre, J. M. (2002). *Rasch Measurement Transactions*, 1 Volume 16:2, Autumn 2002 6:2
- Linacre, J. M. (2008). The Expected Value of a point-Biserial (or Similar) Correlation, *Rasch Measurement Transactions*, 22(1): 1154.
- Lord, F. (1952). *A Theory of Test Scores* (Psychometric Monograph No. 7). Richmond, VA: Psychometric Corporation. Acessado em 02/02/2015 em <http://www.psychometrika.org/journal/online/MN07.pdf>
- Lord, F. M. (1980). *Application of item response theory to practical testing problems*, Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lumsden, J. (1978). Tests are perfectly reliable. *British Journal of Mathematical and Statistical Psychology*, 31, 19-26.
- Ma, S.-C., Chien, T.-W., Wang, H.-H., Li, Y.-C., & Yui, M.-S. (2014). Applying Computerized Adaptive Testing to the Negative Acts Questionnaire-Revised: Rasch Analysis of Workplace Bullying. *Journal of Medical Internet Research*, 16(2), e50. doi:10.2196/jmir.2819
- McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment*. Boston: Allyn & Bacon.
- Millsap, Claudette M. (2000). *Comparison of Computer Testing versus Traditional Paper and Pencil Testing*. (Doctoral dissertation, University of North Texas). Disponível em http://digital.library.unt.edu/ark:/67531/metadc2621/m2/1/high_res_d/Dissertation.pdf
- Nota Técnica do Ministério da Educação (2012). Disponível em http://download.inep.gov.br/educacao_basica/enem/nota_tecnica/2011/nota_tecnica_tri_enem_18012012.pdf
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice Hall.
- Olea J., Ponsoda, V., & Prieto, G. (1999) "Tests Informatizados: Fundamentos y Aplicaciones", Ediciones Pirámide, 1999.

- Primi, R., & Almeida, L. S. (2002). Inteligência geral ou fluida: Desenvolvimentos recentes na sua concepção. *Sobredotação*, 3 (2), 127-144.
- Pasquali, L. (1997). O investimento em Testes Psicológicos. *Anais do I Congresso Ibero-Americano de Avaliação Psicológica* (pp. 59-60). Porto Alegre: PUCRS.
- Pasquali, L. (2003). *Psicometria : teoria dos testes na psicologia e na educação*. Petrópolis, RJ : Vozes.
- Pasquali, L. (2009). *Psicometria*. *Rev. esc. enferm. USP* [online]. Vol.43, pp. 992-999 . Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0080-62342009000500002&lng=en&nrm=iso>
- Pasquali, L., & Colaboradores. (2010). *Instrumentação Psicológica: Fundamentos e práticas*. Porto Alegre. Artmed
- Pasquali, L., & Primi, R. (2003). Fundamentos da Teoria da Resposta ao Item –TRI. *Avaliação Psicológica*, p.99-110.
- Piton-Gonçalves, Jean., Monzón, A. J. B., & Aluisio, S. M. (2009). Métodos de avaliação informatizada que tratam o conhecimento parcial do aluno e geram provas individualizadas. In: *XX Simpósio Brasileiro de Informática na Educação*, Florianópolis. *Anais do XX Simpósio Brasileiro de Informática na Educação - 17-20 de novembro de 2009 - Porto Alegre: Sociedade Brasileira de Computação*, 2009.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Danish Institute for Educational Research, Copenhagen.
- Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests (Expanded ed.). Chicago: University of Chicago Press.
- Reckase, M. D. (1974). An interactive computer program for tailored testing based on the one parameter logistic model. *Behavior Research Methods and Instrumentation*, 6, 208-212.
- Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, 9, 401-412.
- Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, 21, 25-36.
- Richard, J-F. (1995) *Les activités mentales: comprendre, raisonner, trouver des solutions*. Paris: Armand Colin.
- Roazzi, A., & Dias, M. G. B. B. (2001). Teoria das facetas e avaliação na pesquisa social transcultural: Explorações no estudo do juízo moral. In *Conselho Regional de Psicologia – 13a Região PB/RN (Ed.), A diversidade da avaliação psicológica: Considerações teóricas e práticas* (pp. 157-190). João Pessoa, PB: Ideia.

- Roazzi, A. (2008). O que nos torna uma espécie inteligente: A Inteligência em uma perspectiva epistemológica. Em L. Almeida; A. Candeias; A. Roazzi & R. Primi (Orgs.), *Inteligência: Definição e Medida na Confluência de Múltiplas Concepções* (Cap. 1, pp.13-48). São Paulo: Casa do Psicólogo.
- Roazzi, A. (1995). Categorização, Formação de Conceitos e Processos de Construção de Mundo: Procedimento de Classificações Múltiplas para o Estudo de Temas Conceituais e sua Forma de Análise Através de Métodos Multidimensionais. *Cadernos de Psicologia*, 1, 1-27.
- Rudner, Lawrence M. (1998). An On-line, Interactive, Computer Adaptive Testing Tutorial, Disponível em <http://edres.org/scripts/cat>. Acessado em 12 de Dezembro de 2014.
- Santos, R. G., & Almeida, C. S. F. (2007). Modelagem de Software para Auxiliar na Educação de Pessoas com Autismo. 63 f. Tese (Especialização em Educação Especial Inclusiva). Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, Brasil.
- Sandoval, J. A. C. (1993). *La creatividad: teoria básica y implicaciones pedagógicas*. Salamanca: Librería Cervantes.
- Seabra-Santos, M. J. (1998). WPPSI-R: Estudos de adaptação e validação em crianças portuguesas. Dissertação de doutoramento em Psicologia. Coimbra: Universidade de Coimbra.
- Seabra-Santos, M. J. (2000). Avaliação psicológica em idade pré-escolar: O caso da avaliação da inteligência. *Psychologica*, 25, 143-162.
- Shantz, C. U. (1975). The development of social cognition. In E. M. Hetherington (Ed.), *Review of Child Development Research*. Chicago: The University of Chicago Press, Vol. 5.
- Smits, N., Cuijpers, P., van Straten, A. (2011). Applying computerized adaptive testing to the CES-D scale: a simulation study. *Psychiatry Res.* 2011 Jun 30;188(1):147-55. doi: 10.1016/j.psychres.2010.12.001. Epub 2011 Jan 3. PubMed PMID: 21208660.
- Springer, S. P., & Deutsch, G. (1994). *Cerebro izquierdo, cerebro derecho*. Barcelona: Editorial Gedisa.
- Snijders-Oomen, N., Tellegen, P. J., & Laros, J. A. (1989). *S. O. N.-R 5 ½ -17*. The Netherlands: Wolters-Noordhoff.
- Sternberg, R. J. (1992). *As capacidades intelectuais humanas: uma abordagem em processamento de informações*. Porto Alegre: Artes Médicas.
- Sternberg, R. J., & Pretz, J. E. (2005). *Cognition and Intelligence: Identifying the Mechanisms of the Mind*. New York. Cambridge University Press.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680.

- Stone, E. Davey, T. (2011). *Computer-Adaptive Testing for Students With Disabilities: A Review of the Literature*. Research Report ETS RR-11-32. Educational Testing Service (ETS).
- Tarrida, A. C. (2002). *La inteligencia en acción*. Barcelona: Masson.
- Tiberghien, G. (1999). *Mémoire ou mémoires*. In J-F Dortier (Coord.), *Le cerveau et la pensée: La révolution des sciences cognitives*. Auxerre: Presses Universitaires de France.
- Tirapu-Ustárroz, J., & Muñoz-Céspedes, J. M. (2005). *Memoria y funciones ejecutivas*. *Revista de Neurologia*, 41 (8), 475-484.
- Thissen, D., & Mislevy, R. J. (2000). *Testing Algorithms*. In Wainer, H. (Ed.) *Computerized Adaptive Testing: A Primer*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Thompson, Nathan A., & Weiss, David A. (2011). *A Framework for the Development of Computerized Adaptive Tests*. *Practical Assessment, Research & Evaluation*, 16(1). Available online: <http://pareonline.net/getvn.asp?v=16&n=1>
- Thorpe, G. L., McMillan, E., Sigmon, S. T., Owings, L. R., Dawson, R., & Bouman, P. (2007). *Latent trait modeling with the Common Beliefs Survey: Using item response theory to evaluate an irrational beliefs inventory*. *Journal of Rational- Emotive & Cognitive- Behavior Therapy*, 25, 175-189. doi: 10.1007/s10942-006-0039-9
- Thurstone, L. L. (1925). *A method of scaling psychological and educational tests*. *Psychometrika*, 47, 397-412.
- Thurstone, L. L. (1928). *Attitudes can be measure*. *American Journal of Sociology*, 33, 529-554.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Torrecilla, F. J. M. (2006). *Estudios sobre eficacia escolar en Iberoamérica: 15 buenas investigaciones*. 1. ed. Bogotá: Convenio Andrés Bello.
- van der Linden, W. J., & Pashley, P. J. (2010). *Item selection and ability estimation adaptive testing*. In W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 3-30). New York: Springer.
- Varma, S. (2010). *Preliminary Item Statistics Using Point-Biserial Correlation and P-Values*. Educational Data Systems, Inc. Disponível em http://www.eddata.com/resources/publications/EDS_Point_Biserial.pdf
- Vernon, P. E. (1950). *The structure of human abilities*. London: Methuen.
- Wainer, H., & Kiely, G. L. (1987) "Item clusters and computerized adaptive testing: a case of testlets", In: *Journal of Educational Measurement*, 1987, pp. 16.

- Wainer, H., & Mislevy, R.J. (2000). Item response theory, calibration, and estimation. In Wainer, H. (Ed.) *Computerized Adaptive Testing: A Primer*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Walker, J., Böhnke, J. R., Cerny, T., Strasser, F. (2010) Development of symptom assessments utilising item response theory and computer-adaptive testing--a practical method based on a systematic review. *Crit Rev Oncol Hematol*. 2010 Jan; 73(1):47-67. doi: 10.1016/j.critrevonc.2009.03.007. Epub. Review. PubMed PMID: 19375939.
- Wang, Chun., & Chang, Hua-Hua. (2011). *Kullback–Leibler Information and Its Applications in Multi-Dimensional Adaptive Testing*. Springer-Verlag, *Psychometrika*, 76, 1, 13-39
- Ware, J. E. Jr., Kosinski, M., Bjorner, J. B., Bayliss, M. S., Batenhorst, A., Dahlöf, C. G., Tepper, S., & Dowson, A. (2003) Applications of computerized adaptive testing (CAT) to the assessment of headache impact. *Qual Life Res*. 2003 Dec;12 (8):935-52. PubMed PMID:14651413.
- Warm, T. A. (1989). Weighted Likelihood Estimation of Ability in Item Response Theory. *Psychometrika*, 54, 427-450.
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21, 361-375.
- Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence – third edition*. San Antonio, TX; The Psychological Corporation.
- Winsteps Help for Rasch Analysis. (2010). Disponível em http://www.winsteps.com/winman/multidimensionality.htm?zoom_highlightsub=unidimen
- Woodcock, R. W. (1997). The Woodcock-Johnson Tests of Cognitive Ability - revised. In D. P. Flanagan, J. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues*. New York: The Guilford Press, pp. 230-246.
- Woodcock, R. W. (2002). New looks in the assessment of cognitive ability. *Peabody Journal of Education*, 77 (2), 6-22.
- Wright, B. D. (1996). Local dependency, correlations and principal components. *Rasch Meas Trans*, 10, 3, 509-511.
- Yao, T. (1991) CAT with a poorly calibrated item bank. *Rasch Measurement Transactions* 5:2, p. 141.
- Yen, W.M. (1986). The choice of scale for educational measurement: An IRT perspective. *Journal of Educational Measure*, 23, 299-326.

Anexos

Anexos

ANEXO A – Telas da versão informatizada da escala ECCOs 4/10



Tela do Menu Principal

Cadastro

Identificação

Protocolo: 000002

Cidade: Recife Estado: PE País: Brasil

Escola: Sociedade Educacional Piedade

Nome: Renato Guedes dos Santos Data de Nascimento: 2002-01-21 Idade: 8

Sexo: M F

Série: Não estuda 1 Ano 2 Ano 3 Ano 4 Ano 5 Ano

Pré-escola: 0 1 2 3 4 5 ECCOs tradicional P Não Sim

Indicadores Sócio Econômicos

Escolaridade do Pai: Não escolarizado Fundamental Médio Superior Pós-graduação
 Mestrado Doutorado

Profissão do Pai: Analista de Sistemas

Escolaridade da Mãe: Não escolarizado Fundamental Médio Superior Pós-graduação
 Mestrado Doutorado

Profissão da Mãe: Professora

Tela de cadastro: identificação do aluno e indicadores socioeconômicos

Cadastro

Dados Escolares

Capacidade Cognitiva: Muito fraca Fraca Mediana Boa Muito boa

Rendimento Acadêmico: Muito fraco Fraco Mediano Bom Muito bom

Disciplina: Indisciplinado Disciplinado

Motivação: Pouca Mediana Muita

Assiduidade: Falta muito Falta às vezes Dificilmente falta

Repetente: Não Sim

Dados Médicos

Tem algum tipo de comprometimento? Sim Não

Motor

Visual

Auditivo

Fala

Neurológico

Cognitivo

Deficiência:

Cadastrar

Tela de cadastro: dados escolares e médicos

Escolha um teste para: Renato Guedes dos Santos Protocolo: 0000017

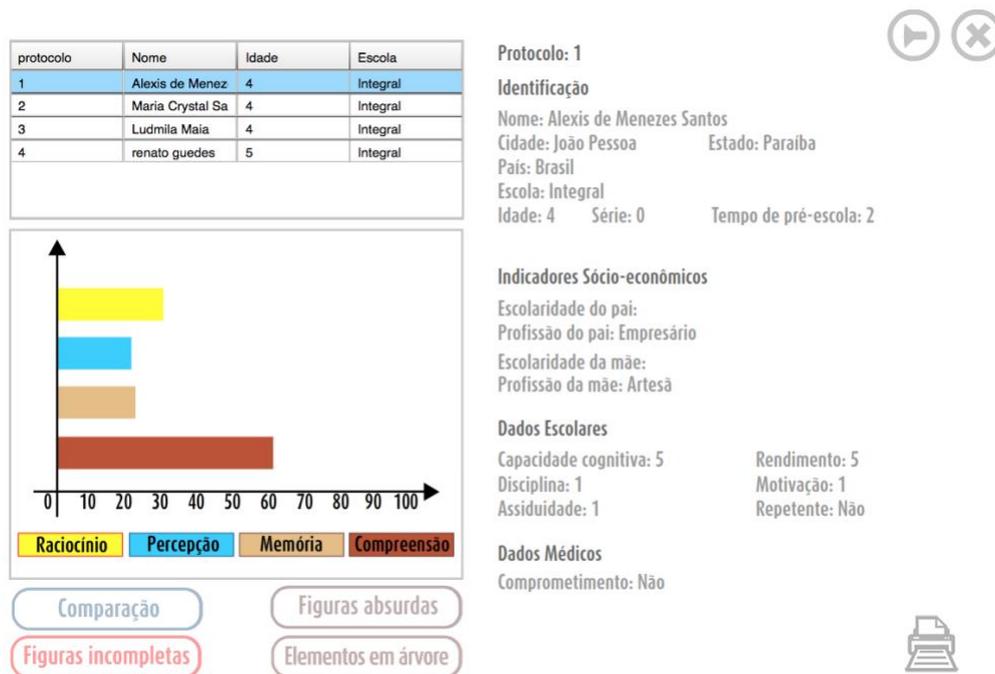
Comparação

Figuras incompletas

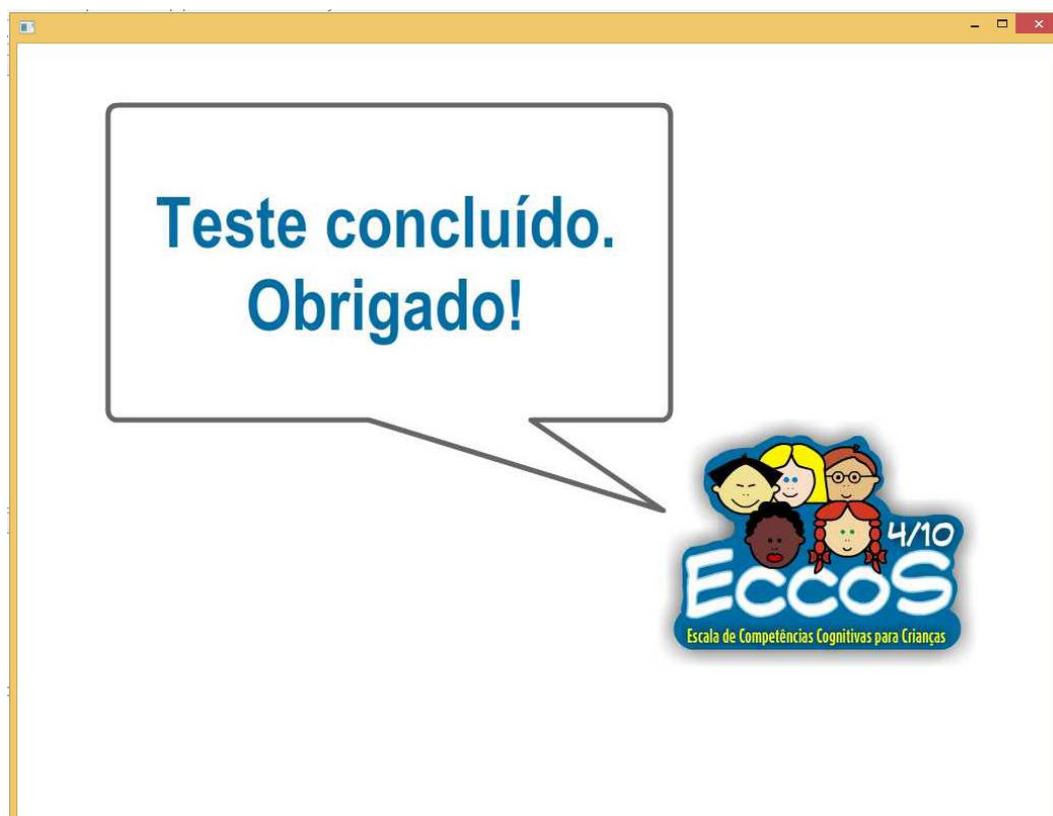
Figuras absurdas

Elementos em árvore

Tela de seleção do teste



Tela de resultados por aluno



Tela de finalização do teste

ANEXO B – Estatísticas dos Itens do SPSS da prova de Figuras Incompletas

Estatísticas dos Itens

	Média se o item for deletado	Variância se o item for deletado	Correlação Item-Total	Alfa de Cronbach se o Item for deletado		Média se o item for deletado	Variância se o item for deletado	Correlação Item-Total	Alfa de Cronbach se o Item for deletado
i1	19,54	57,954	,265	,887	i23	20,20	55,370	,439	,884
i2	19,56	57,680	,287	,887	i24	20,14	54,981	,473	,884
i3	19,66	57,442	,193	,888	i25	20,02	54,623	,506	,883
i4	19,69	55,940	,436	,885	i26	20,18	55,484	,415	,885
i5	19,71	55,438	,504	,884	i27	20,12	55,325	,421	,885
i6	19,70	56,337	,363	,886	i28	20,08	55,259	,423	,885
i7	19,61	56,712	,410	,885	i29	20,33	56,622	,322	,886
i8	19,80	55,053	,498	,884	i30	20,20	55,970	,352	,886
i9	19,66	55,870	,489	,884	i31	20,24	55,321	,470	,884
i11	19,75	56,751	,264	,887	i32	20,25	55,411	,458	,884
i12	19,65	56,485	,381	,886	i33	20,29	56,664	,285	,887
i13	19,90	55,591	,384	,885	i34	20,11	54,893	,478	,884
i14	19,98	55,372	,403	,885	i35	20,19	56,448	,277	,887
i15	19,94	54,789	,489	,884	i36	20,24	55,136	,498	,884
i16	19,94	54,339	,552	,882	i37	20,14	56,497	,260	,888
i17	19,96	54,832	,480	,884	i38	20,33	57,220	,217	,888
i18	19,91	54,863	,483	,884	i39	20,36	57,220	,234	,887
i19	19,76	57,634	,120	,890	i40	20,22	54,558	,576	,882
i20	20,13	55,446	,404	,885	i41	20,38	57,942	,111	,889
i21	20,15	56,675	,237	,888	i42	20,33	56,777	,295	,887
i22	20,23	57,160	,185	,889	i43	20,36	56,513	,371	,886

Estatísticas de consistência

Alfa de Cronbach	Parte 1	Valor	,810
		N de Itens	21 ^a
	Parte 2	Valor	,818
		N de Itens	21 ^a
Total N de Itens			42
Correlação entre as formas			,690
Coeficiente Spearman-Brown	Tamanho igual		,816
	Tamanho não igual		,816
Coeficiente Guttman Split-Half			,815
a. Os itens são: i1, i2, i3, i4, i5, i6, i7, i8, i9, i11, i12, i13, i14, i15, i16, i17, i18, i19, i20, i21, i22.			
b. Os itens são: i23, i24, i25, i26, i27, i28, i29, i30, i31, i32, i33, i34, i35, i36, i37, i38, i39, i40, i41, i42, i43.			

Estatísticas de Confiabilidade

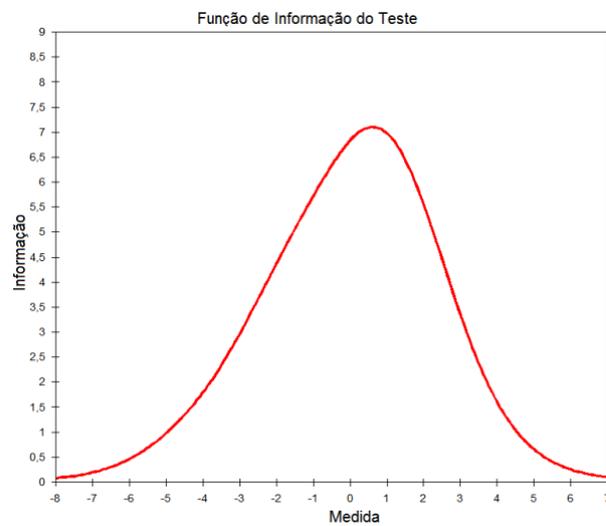
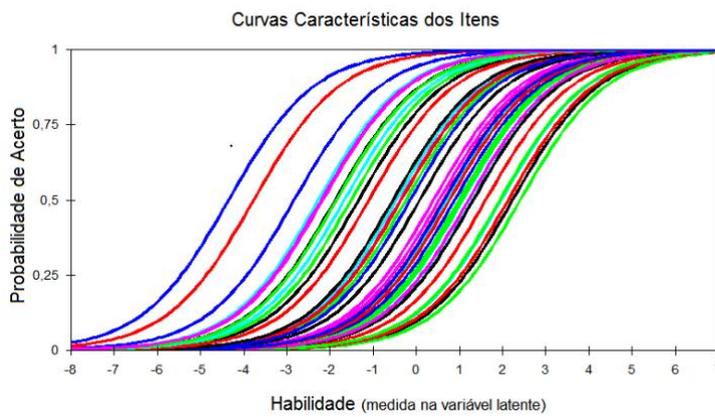
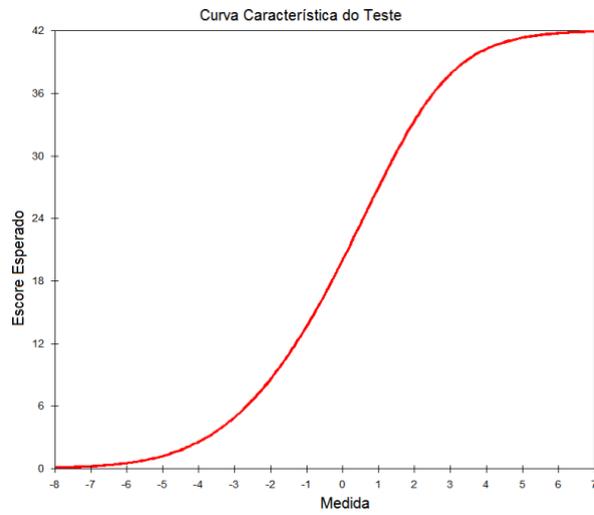
Alfa de Cronbach	Alfa de Cronbach baseado nos itens padronizados	Nr. de Itens
,888	,887	42

ANEXO C – Estatísticas dos Itens do Winsteps prova Figuras Incompletas

ESTATÍSTICAS DOS ITENS: ORDENADO PELA MEDIDA

NUMERO ENTRADA	ESCORE TOTAL	CONT. TOTAL	ESCORE MEDIDA	ERRO PADRÃO	INFIT MNSQ	OUTFIT ZSTD	PTBISERIAL CORR.	PREVISIBIL. EXP.	PREVISIBIL. OBS%	PREVISIBIL. EXP%	ITEM		
40	35	272	2.41	.20	1.20	1.6	1.42	1.4	.11	.30	86.8	87.6	141
42	40	272	2.23	.19	.89	-1.0	.90	-.3	.37	.31	87.5	86.0	143
38	41	272	2.20	.18	1.11	1.0	1.00	.1	.23	.31	84.2	85.6	139
28	48	272	1.97	.17	.99	-.1	1.08	.4	.32	.33	85.3	83.5	129
37	48	272	1.97	.17	1.16	1.5	1.09	.4	.22	.33	82.4	83.5	138
41	48	272	1.97	.17	1.05	.6	.94	-.2	.29	.33	82.4	83.5	142
32	60	272	1.64	.16	1.08	.9	1.20	1.0	.29	.35	77.6	80.2	133
31	72	272	1.34	.15	.90	-1.4	.82	-1.0	.46	.37	80.1	77.3	132
30	73	272	1.32	.15	.88	-1.7	.80	-1.2	.47	.37	82.7	77.1	131
35	73	272	1.32	.15	.84	-2.2	.75	-1.5	.50	.37	82.7	77.1	136
21	76	272	1.25	.15	1.23	3.0	1.37	2.0	.19	.38	71.3	76.5	122
39	78	272	1.20	.15	.77	-3.5	.64	-2.4	.58	.38	80.9	76.0	140
29	84	272	1.07	.15	1.03	.4	1.11	.7	.35	.39	75.0	74.8	130
22	85	272	1.05	.15	.95	-.7	.85	-1.0	.44	.39	74.6	74.6	123
34	88	272	.98	.15	1.15	2.3	1.15	1.0	.28	.39	68.8	74.0	135
25	90	272	.94	.14	.97	-.4	.89	-.7	.41	.39	74.6	73.7	126
20	98	272	.78	.14	1.20	3.2	1.24	1.7	.24	.40	65.4	72.4	121
36	101	272	.72	.14	1.17	2.8	1.29	2.1	.26	.40	65.1	71.9	137
23	102	272	.70	.14	.91	-1.5	.87	-1.0	.47	.40	77.2	71.8	124
19	105	272	.64	.14	1.00	.0	1.05	.5	.40	.41	71.0	71.4	120
26	106	272	.62	.14	.99	-.2	.92	-.7	.42	.41	72.8	71.3	127
33	110	272	.54	.14	.92	-1.5	.84	-1.4	.48	.41	74.3	70.9	134
27	116	272	.42	.14	.99	-.2	.96	-.3	.42	.41	71.7	70.4	128
24	134	272	.08	.14	.90	-2.0	.84	-1.6	.51	.42	75.4	69.7	125
13	145	272	-.13	.14	1.03	.5	.99	-.1	.40	.42	68.0	70.1	114
16	151	272	-.24	.14	.92	-1.4	.89	-1.1	.48	.42	76.5	70.6	117
14	156	272	-.34	.14	.92	-1.5	.87	-1.3	.49	.42	75.4	71.0	115
15	156	272	-.34	.14	.84	-2.9	.75	-2.7	.55	.42	77.6	71.0	116
17	163	272	-.48	.14	.92	-1.4	.92	-.8	.48	.42	74.3	71.7	118
12	167	272	-.56	.14	1.03	.5	1.07	.7	.38	.42	72.1	72.2	113
8	193	272	-1.12	.15	.87	-1.7	.82	-1.3	.50	.41	80.5	77.3	18
18	203	272	-1.36	.16	1.42	4.5	1.61	3.3	.12	.40	70.2	79.7	119
10	208	272	-1.49	.16	1.22	2.3	1.45	2.4	.26	.40	75.0	80.9	111
5	217	272	-1.74	.17	.81	-2.0	.71	-1.6	.50	.38	87.5	83.2	15
6	221	272	-1.86	.18	1.06	.6	.92	-.3	.36	.38	83.8	84.2	16
4	222	272	-1.89	.18	.94	-.5	.76	-1.2	.44	.38	84.9	84.5	14
3	231	272	-2.19	.19	1.29	2.2	1.35	1.4	.19	.36	83.5	87.0	13
9	231	272	-2.19	.19	.78	-1.8	.63	-1.6	.49	.36	90.1	87.0	19
11	233	272	-2.27	.20	.97	-.2	.84	-.5	.38	.36	87.9	87.6	112
7	246	272	-2.85	.23	.83	-1.0	.61	-1.2	.41	.32	91.9	91.3	17
2	260	272	-3.85	.32	.93	-.2	.57	-.8	.29	.24	96.0	95.7	12
1	265	272	-4.48	.40	.87	-.3	.35	-1.2	.27	.19	97.4	97.4	11
MÉDIA	132.8	272.0	.00	.17	1.00	-.1	.95	-.2			79.1	78.6	
D. PD.	69.2	.0	1.68	.05	.15	1.7	.26	1.3			7.7	7.3	

ANEXO D – Curvas dos itens, do teste e de informação do teste da prova Figuras Incompletas



ANEXO E - Termo de Consentimento Livre e Esclarecido

Projeto de pesquisa: Construção e validação de provas informatizadas integrante da Escala de Avaliação de Competências Cognitivas para Crianças dos 4 aos 10 Anos de Idade

Pesquisador Responsável: Renato Guedes dos Santos

Endereço: Rua Ribeiro de Brito 928, Boa Viagem, Recife / PE , CEP: 50.021-310

Tel: 81 3301-5711/ 81 9114-1994

Email: renato.gsantos@gmail.com

1. Natureza da pesquisa: Você está sendo convidado(a) a participar desta pesquisa que tem como finalidade compreender o desenvolvimento do raciocínio em crianças e pré-adolescentes de 4 a 10 anos. Com isto, procura-se produzir uma versão informatizada de uma escala para avaliação do desenvolvimento cognitivo denominada ECCOs, construída em Portugal e que também está sendo avaliada aqui no Brasil.
2. Participantes da pesquisa: Aproximadamente 250 crianças brasileiras de idades entre 4 e 10 anos.
3. Envolvimento na pesquisa: Ao aceitar participar deste estudo, você estará permitindo que seu(ua) filho(a) seja submetido a um teste não verbal, apresentado em computador. Este teste ocorrerá na escola. Você tem a liberdade de recusar a participação de seu filho(a) em qualquer fase da pesquisa, sem qualquer prejuízo para você ou para ele(a). Sempre que quiser você poderá pedir mais informações sobre a pesquisa ou entrar em contato com o coordenador da pesquisa através dos telefones 3301-5711 ou 9114-1994.
4. Sobre os testes: Ao participar deste estudo, seu(ua) filho(a) irá ser solicitado(a) a dar respostas a um teste no computador, apresentadas em ordem crescente de dificuldade, onde será acompanhado o ritmo de compreensão da criança.
5. Riscos e desconforto: A participação nesta pesquisa não traz complicações, talvez apenas um pequeno sentimento de timidez que algumas pessoas sentem quando estão sendo observadas ou avaliadas. Os procedimentos utilizados nesta pesquisa seguem as normas estabelecidas pela Resolução 196/96 do Conselho Nacional de Saúde e não oferecem risco a sua integridade física, psíquica e moral. Nenhum dos procedimentos utilizados oferece riscos a sua dignidade.
6. Confidencialidade: Todas as informações coletadas neste estudo são confidenciais. Os dados serão identificados com um código em substituição ao nome de quaisquer dos participantes. Apenas os pesquisadores envolvidos terão acesso integral aos dados. Os dados encontrados poderão ser divulgados em publicações, conferências ou em situações de ensino e seguirá a mesma conduta, ou seja, omitindo os dados pessoais dos participantes.
7. Benefícios: Ao participar desta pesquisa você não obterá nenhum benefício direto. Contudo, você estará colaborando com um estudo sobre o desenvolvimento de processos cognitivos em crianças que poderá beneficiar não apenas você ou a seu(ua) filho(a), mas a todo um grupo social. Posteriormente essas informações poderão ser usadas em benefício de futuras pesquisas.
8. Pagamento: Você não terá nenhum tipo de despesa por participar desta pesquisa. E nada será pago por sua participação.

Tendo em vista os itens acima apresentados, eu _____, de forma livre e esclarecida, manifesto meu interesse em participar da pesquisa.

Recife, / /

Seguem-se as assinaturas

ANEXO F - Procedimentos para aplicação do teste Figuras Incompletas

Material	Prova eletrônica composta por: - dois exercícios de treino - quarenta e três itens constituintes da prova
Critério de início	No teste sequencial, a prova inicia-se, para todas as crianças, no item 1. Na versão adaptativa do teste, o programa selecionará o item inicial e posteriores.
Critério de parada	No teste sequencial, a prova será interrompida no caso de 4 erros sucessivos No teste adaptativo, o programa irá interromper a prova quando o erro padrão estiver aceitável e a estimativa de habilidade se mantiver estável por 4 itens sucessivos

Instruções

Acomodar a criança adequadamente na frente do computador e apresentá-la à prova demonstrando como se dá a seleção da resposta e a passagem para um novo item. Instruir a criança, dizendo-lhe:

“Vou te mostrar umas figuras e vamos ver o que acontece quando passamos de uma para a outra.

Repara: temos aqui um urso de pelúcia grande. Ao passarmos para aqui (apontar a célula seguinte da tabela) passamos a ter um urso de pelúcia pequeno. Então, um urso grande transformou-se num urso pequeno. Agora vamos olhar para este carro (apontar a imagem inferior). É grande, não é? Ao passar para aqui (apontar a célula vazia) como irá ficar? ...”

Deixar a criança indicar a alternativa de resposta correta apontando na tela sensível ao toque ou clicando com o mouse. Caso a criança erre ou demonstre dificuldade, analisar cada alternativa de resposta com ela, explicando a razão pela qual não a resposta não está adequada.

Mostre para a criança o botão na tela que deve ser clicado para passar para o próximo item. Passar ao segundo exemplo, procurando verificar se a tarefa foi compreendida pela criança. Caso surjam dúvidas, deve se proceder a uma análise mais pormenorizada tal como atrás foi referido.

Antes de passar à aplicação dos itens constituintes da prova, é muito importante ter a certeza de que a criança percebeu o que lhe é pedido. Pode-se voltar para o item 1 novamente, a fim de garantir que a criança fixou a instrução.

No caso da aplicação do teste adaptativo, é importante alertar à criança que a prova irá alternar entre itens difíceis e fáceis, de forma a tentar reduzir algum eventual estresse.

Classificação	As respostas serão cotadas com um ou zero pontos consoante a alternativa de resposta selecionada pela criança for certa ou errada, respectivamente.
Pontuação	1 ponto para cada item respondido acertadamente - de zero a 42 pontos. No caso da testagem eletrônica, o sistema irá gerar um indicador da estimativa de habilidade calculada

Respostas corretas

Item	Resposta	Item	Resposta
1	1	23	3
2	3	24	5
3	3	25	3
4	1	26	4
5	2	27	5
6	2	28	2

7	3	29	5
8	1	30	3
9	2	31	1
excluído		32	4
11	5	33	1
12	2	34	3
13	5	35	3
14	3	36	2
15	5	37	4
16	1	38	5
17	4	39	3
18	1	40	1
19	4	41	3
20	5	42	3
21	4	43	2
22	2		

ANEXO G - Procedimentos para aplicação do teste Comparação de Figuras

Material	<p>Prova eletrônica composta por:</p> <ul style="list-style-type: none"> - dois exercícios de treino. - 38 conjuntos de cinco imagens. <p>O programa dispõe de cronômetro interno para medição do tempo de execução da prova..</p>
Critério de início de aplicação	<p>No teste sequencial, a prova inicia-se, para todas as crianças, no item 1.</p> <p>Na versão adaptativa do teste, o programa selecionará o item inicial e posteriores.</p>
Critério de paragem de aplicação	<p>O programa irá parar a administração da prova quando tiverem sido completados:</p> <ul style="list-style-type: none"> - 4 minutos, em crianças de 4 e de 5 anos; - 3 minutos e 30 segundos, em crianças de 6 e de 7 anos; - 3 minutos, em crianças de 8, de 9 e de 10 anos. <p>No teste sequencial, a prova também será interrompida no caso de 4 erros sucessivos</p> <p>No teste adaptativo, o programa irá interromper a prova quando o erro padrão estiver aceitável e a estimativa de habilidade se mantiver estável por 4 itens sucessivos</p>
Instruções	

A prova inicia-se dizendo à criança: “Vou te mostrar vários desenhos. Quero que olhe para todos com muita atenção e toque com o dedo (ou clique com o mouse) os dois desenhos que são exatamente iguais. Vamos experimentar.”

Apresentar o 1º exemplo à criança e dizer-lhe: *“Temos aqui vários desenhos. Quais são os que são iguais? Aponta com o dedo.”*

Se a resposta da criança for correta, passar ao segundo exemplo, seguindo o mesmo procedimento. Caso a criança dê uma resposta incorreta, procurar que compare as imagens entre si e chegue às alternativas corretas.

Aplicar, então, o exemplo 2, voltando a pedir-lhe que aponte as duas imagens que são mesmo iguais. Procurar que dê a resposta sem ajuda.

Apresentar o cronômetro existente no programa à criança explicando-lhe que serve para medir o tempo e que ele vai indicar quando devemos parar o teste.

Depois da aplicação dos exemplos e da apresentação do cronômetro, dizer à criança: *“Agora você vai me dizer quais são os desenhos iguais o mais depressa que puder.”*

A prova começa pelo item 1. O cronômetro começará a contar o tempo a partir do momento em que o primeiro item for apresentado à criança.

No caso da aplicação do teste adaptativo, é importante alertar à criança que a prova irá alternar entre itens difíceis e fáceis, de forma a tentar reduzir algum eventual estresse quando ela se deparar com itens acima da sua capacidade.

Classificação	Será atribuído um ponto para cada par de imagens iguais selecionado.
Pontuação	1 ponto para cada item respondido acertadamente - de 0 a 38 pontos. O programa também irá gerar e armazenar a estimativa de habilidade calculada

Respostas corretas:

Item	Par correto	Item	Par correto
1	2 - 4	20	1 - 4
2	1 - 5	21	2 - 5
3	2 - 5	22	3 - 5
4	1 - 3	23	2 - 4
5	1 - 4	24	1 - 3
6	2 - 4	25	2 - 5
7	2 - 5	26	1 - 5
8	1 - 4	27	2 - 4
9	1 - 3	28	1 - 5
10	2 - 4	29	1 - 5
11	1 - 5	30	3 - 5
12	1 - 3	31	2 - 5
13	3 - 5	32	1 - 5
14	1 - 5	33	1 - 4
15	3 - 5	34	2 - 5
16	3 - 4	35	1 - 3
17	1 - 4	36	2 - 5
18	2 - 5	37	2 - 4
19	3 - 5	38	3 - 5

ANEXO H - Procedimentos para aplicação do teste Elementos em Árvore

Material	Prova eletrônica composta por: - dois itens de treino. - seis pares de árvores com diferentes números de ramos, com enfeites que desaparecem após o tempo determinado O programa dispõe de cronômetro interno para medição do tempo de execução da prova.
Critério de início de aplicação	A prova se inicia na Árvore 1 (série 1) para todas as crianças.
Critério de paragem de aplicação	O programa para a administração da prova quando nas duas árvores de uma mesma série a criança não consegue um total de pontos, no mínimo, igual ao número de elementos presentes numa só das árvores dessa série.

A prova é iniciada dizendo à criança:

“Vou te apresentar uma árvore com bolas. Você vai olhar para ela com muita atenção para não esquecer em que ramos / galhos estão penduradas as bolas. Depois, vou te mostrar uma árvore sem as bolas e você vai me dizer onde estavam penduradas as bolas da árvore que você viu antes. Vamos experimentar.”

Apresentar para a criança a árvore do exemplo 1, cuja visualização na tela será mantida pelo programa por 15 segundos. Quando as bolas desaparecerem, o aplicador do teste deverá perguntar:

“Onde estavam penduradas as bolas na árvore?”

A criança deverá indicar com os dedos ou clicando com o mouse as posições onde as bolas estavam. Se a criança responder corretamente, pedir para a criança avançar para o exemplo 2, clicando no botão de avanço da tela, e seguir o mesmo procedimento. No caso de ela mostrar dificuldades no exemplo 2 voltar para o exemplo 1. Ao final dos dois exemplos será apresentada a árvore 1, do par 1.

O programa registrará na base de dados os locais (certos ou errados) onde a criança assinalou a presença de um elemento. O programa respeitará sempre o tempo total de apresentação indicado. Caso a criança diga que já sabe ou questionar se já pode retirar aquela árvore, peça-lhe que continue a olhar bem para o modelo até o fim do tempo.

Tempo limite de apresentação:

O tempo que o programa apresentará cada árvore será:

- Série 1 e 2 : Árvores 1, 2, 3 e 4 – 5 segundos
- Série 3 e 4 : Árvores 5, 6, 7 e 8 – 10 segundos
- Séries 5 e 6: Árvores 9, 10, 11 e 12 – 15 segundos

Pontuação:

De zero a 54 pontos.

Classificação

Atribuir **um ponto** por cada elemento colocado no local correto.

Serão ainda de considerar as seguintes situações:

- 1 - Nos casos em que a criança se referir a um número de elementos superior ao apresentado, **descontar-se-á um ponto** por cada elemento acrescentado;
- 2 - Sempre que, da aplicação do critério anteriormente referido, resultar um **total negativo na série** (isto é, em cada conjunto de duas árvores com igual número de ramos e de elementos desenhados), deve o mesmo ser reconvertido em zero pontos. Assim, o total negativo obtido numa qualquer série não afetará a pontuação obtida na outra série.
- 3 - o critério de parada terá em conta o número de pontos obtidos e não o número de elementos corretamente assinalados.

ANEXO I - Procedimentos para aplicação do teste Desenhos Absurdos

Material	Prova eletrônica composta por: - dois itens de treino - vinte e dois itens constituintes da prova
Critério de início	A prova inicia, para todas as crianças, no item 1.
Critério de parada	A prova será interrompida no caso de 4 erros sucessivos

Instruções:

A prova inicia se dizendo à criança:

“Vou te mostrar algumas figuras. Em cada uma delas há alguma coisa que está errada, alguma coisa absurda. Quero que aponte (ou clique) e me fale o que é que está errado em cada figura, que coisa estranha que você acha que tem em cada desenho.”

Apresentar a primeira figura à criança. Se ela não conseguir identificar o absurdo, perguntar o que ela vê na imagem. Após a descrição, explicar que as cenouras (apontar) não nascem em árvores, são raízes que ficam sob a terra. Apresentar, então, o segundo desenho e fazer de novo a pergunta:

“E aqui? O que é que está errado nesta figura? Qual é o absurdo que você pode notar?”

Se surgir dificuldade, proceder como foi referido para o primeiro item, explicando à criança a razão de ser do absurdo existente da imagem: os meninos não andam de skate na água.

Permitir, então, que a criança avance na apresentação dos itens. O programa registrará os elementos onde a criança clicar, mas ela deve ser orientada a sempre falar o que achou absurdo e deve-se anotar em separado qualquer informação que seja relevante. Se a criança não deixar claro a que absurdo ela se refere, pedir-lhe:

“O que é que você está vendo de errado na figura?” ou *“Porque é que você acha que isso é um absurdo?”*

Classificação	<p>O programa registrará o elemento que a criança indicar através de toque ou clique do mouse, mas o aplicador do teste deverá fazer sua própria cotação de pontos, haja vista que para pontuar é necessário ouvir o que a criança fala. Atribuir dois, um ou zero pontos de acordo com os seguintes critérios:</p> <p>Dois pontos: Quando a criança identifica (verbaliza) o absurdo de forma clara e correta, ou é capaz de justificar posteriormente a sua resposta (por exemplo: no item 8, “está errado porque a bailarina não fica dançando no meio da rua”);</p> <p>Um ponto: Quando a criança indica corretamente o absurdo, apontando-o com o dedo ou se referindo verbalmente ao elemento da imagem onde está situado o absurdo, mas não justifica a sua resposta (por exemplo: no item 12, aponta o hipopótamo sem explicar que está com óculos de sol);</p> <p>Zero pontos: Quando a resposta está errada, isto é, quando claramente a criança não compreendeu o absurdo da imagem (por exemplo, no item 5 afirma que o que está errado é o menino jogando bola porque a mãe não tinha deixado).</p>
Pontuações:	De zero a 46 pontos.

Sugestões para correção

Item	Absurdo a Identificar
1	O peixe fora do aquário e o gato no seu interior.
2	O macaco à solta e o menino na jaula.
3	Um carrinho passeando na água.
4	Uma raquete sendo encestada em vez da bola.
5	O jogador de futebol usa sapatos de mulher.
6	Um homem tomando banho vestido.
7	Um carro com chaminé.
8	Uma bailarina dançando no meio da rua.
9	Um pintor usando uma ferramenta no lugar do pincel.

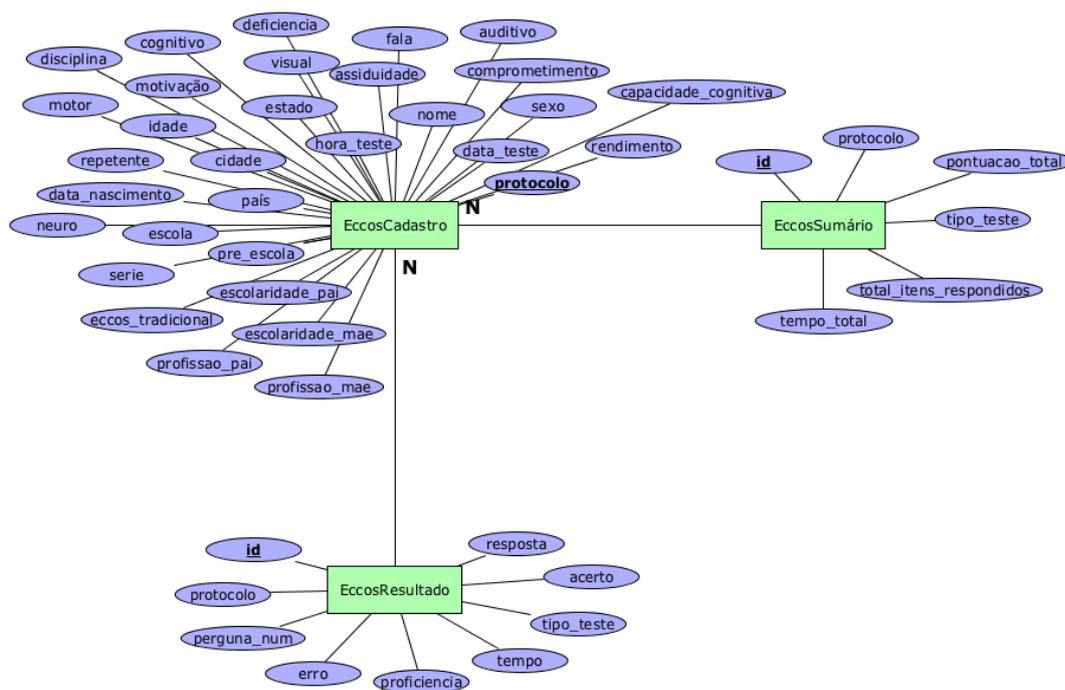
- 10 O cão com a coleira colocada na pata.
- 11 Uma mulher com bigode de homem.
- 12 Um hipopótamo com óculos.
- 13 Um homem sem cabelo se penteando.
- 14 Uma banda tocando na entrada do hospital.
- 15 Um carro circulando em sentido proibido.
- 16 Um policial e um ladrão se cumprimentando amigavelmente.
- 17 A posição do guarda-chuva não permite proteção da chuva.
- 18 O reflexo das casas não corresponde à imagem das mesmas.
- 19 A imagem do espelho está invertida.
- 20 O vento sopra em direções opostas.
- 21 Estão sendo pesadas, simultaneamente, frutas que têm preços diferentes.
- 22 A torcida está aplaudindo o gol do time adversário.

ANEXO K – Estrutura dos bancos de dados

Provas Adaptativas

O banco de dados foi composto por três tabelas, a seguir descritas:

- Tabela EccosCadastro: possui os atributos relacionados à criança no início do teste, quando o aplicador cadastra a mesma. As características presentes nesta tabela são: nome, sexo, cidade, estado, país, escola, data e hora que o teste foi realizado, data de nascimento, idade, pontuação na ECCOs tradicional, série, pré-escola, escolaridade e profissão do pai e da mãe, capacidade cognitiva, rendimento acadêmico, se a criança é repetente, se a criança é motivada, se tem comprometimento na capacidade de fala, audição, motora, visual, neurológica ou cognitiva, se é disciplinada e assídua na escola.
- A tabela EccosSumário: contém variáveis que armazenam características gerais dos testes finalizados, tais como o identificador da criança, a pontuação final, o tempo total necessário para o teste ser finalizado, quantos itens foram respondidos no total e qual foi o tipo de teste respondido.
- A tabela EccosResultado: contém detalhes específicos de cada teste finalizado, tais como: identificador da criança, qual o teste foi respondido, qual questão foi respondida, a resposta fornecida pelo aluno, se o aluno acertou ou não o item, quanto tempo a criança levou naquele item e qual o erro e proficiência calculados para esta questão respondida.

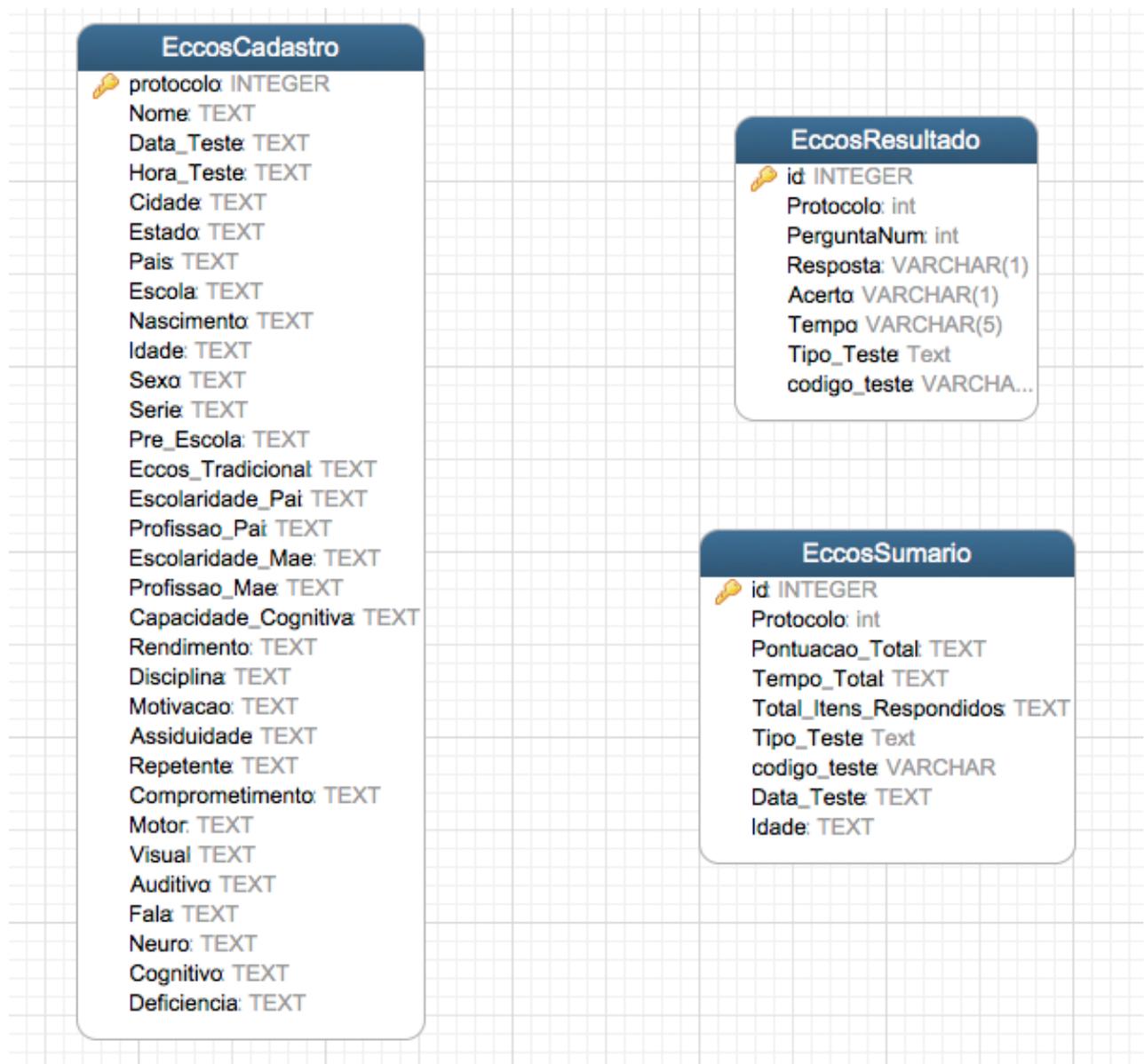


Provas sequenciais

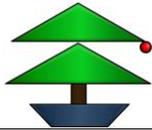
Tabela EccosCadastro - Registro dos dados de identificação da criança e dos indicadores acadêmicos e socioeconômicos

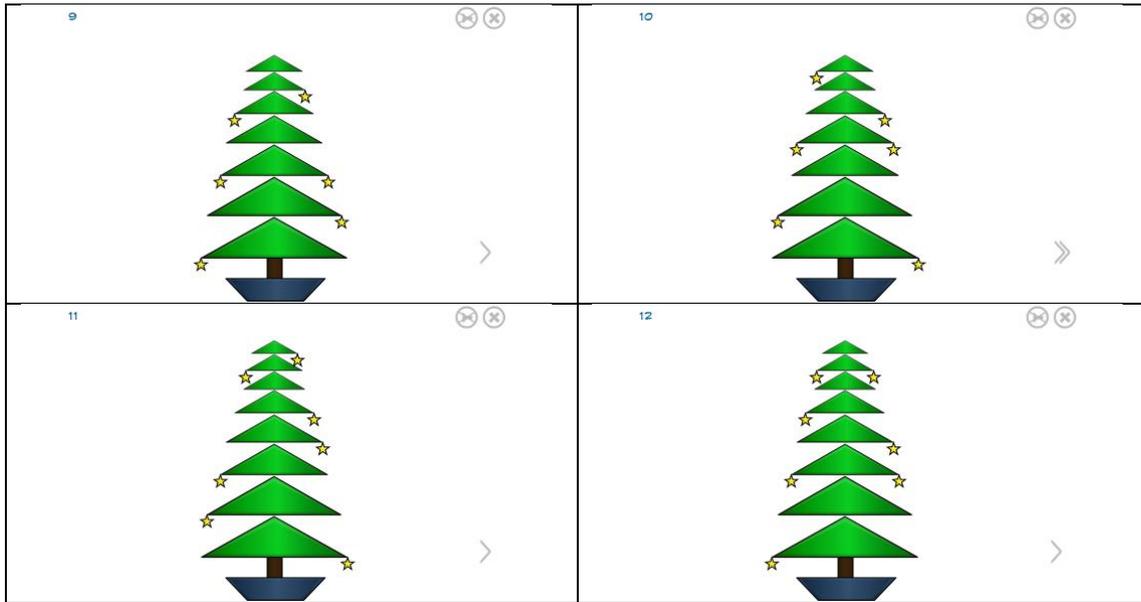
Tabela EccosResultado - É alimentada a cada resposta dada, toda vez que o aluno avança de um item para o próximo os dados são armazenados nessa tabela.

Tabela EccosSumario - Quando um teste é encerrado, independentemente da maneira como ele foi encerrado, essa tabela armazena os dados totalizados do teste que estava em curso.



ANEXO L - Itens da prova Elementos em Árvore - processo de memória

<p>(R) ⊗ ⊗</p>  <p>></p> <p>1</p>	<p>(R) ⊗ ⊗</p>  <p>< ></p> <p>2</p>
<p>⊗ ⊗</p>  <p>>></p> <p>3</p>	<p>⊗ ⊗</p>  <p>></p> <p>4</p>
<p>⊗ ⊗</p>  <p>></p> <p>5</p>	<p>⊗ ⊗</p>  <p>>></p> <p>6</p>
<p>⊗ ⊗</p>  <p>></p> <p>7</p>	<p>⊗ ⊗</p>  <p>>></p> <p>8</p>
<p>⊗ ⊗</p>  <p>>></p>	<p>⊗ ⊗</p>  <p>>></p>



ANEXO M - Itens da prova Figuras Incompletas - processo de raciocínio

<p>(R)  </p> <p> </p> <p></p>	<p>(R)  </p> <p> </p> <p></p>
<p>1  </p> <p> </p> <p></p>	<p>2  </p> <p> </p> <p></p>
<p>3  </p> <p> </p> <p></p>	<p>4  </p> <p> </p> <p></p>
<p>5  </p> <p> </p> <p></p>	<p>6  </p> <p> </p> <p></p>

<p>7</p> <p>⊗</p> <p>»</p>	<p>8</p> <p>⊗</p> <p>»</p>
<p>9</p> <p>⊗</p> <p>»</p>	<p>10</p> <p>⊗</p> <p>»</p>
<p>11</p> <p>⊗</p> <p>»</p>	<p>12</p> <p>⊗</p> <p>»</p>
<p>13</p> <p>⊗</p> <p>»</p>	<p>14</p> <p>⊗</p> <p>»</p>

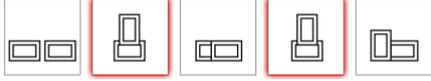
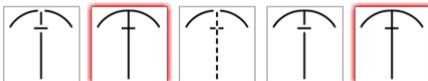
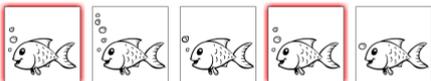
<p>15</p>	<p>16</p>
<p>17</p>	<p>18</p>
<p>19</p>	<p>20</p>
<p>21</p>	<p>22</p>

<p>23</p>	<p>24</p>
<p>25</p>	<p>26</p>
<p>27</p>	<p>28</p>
<p>29</p>	<p>30</p>

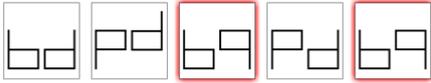
<p>31</p>	<p>32</p>
<p>33</p>	<p>34</p>
<p>35</p>	<p>36</p>
<p>37</p>	<p>38</p>

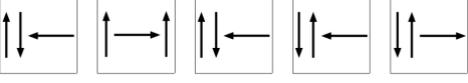
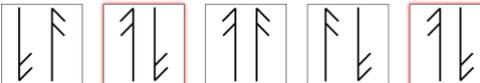
<p>39</p> <p>→ ↓ ↓ ? → ↑ ← ↓ ↑</p>	<p>40</p> <p>▤ ▥ ▴ ? ▾ ▴ ▾ ▴ ▾</p>
<p>41</p> <p>▴ ● ● ▴ ▴ ● ? ▴ ▴ ● ● ▴ ▴ ● ▴ ● ▴ ● ▴ ● ▴ ●</p>	<p>42</p> <p>▣ ● ▴ ● ▣ ▣ ● ? ▣ ▣ ● ▣ ● ▣ ● ▣ ● ▣ ● ▣ ●</p>
<p>43</p> <p>▭ ★ ★ ▭ ▴ ● ? ▴ ★ ● ● ▴ ● ● ● ★ ●</p>	

ANEXO N - Itens da prova Comparação de Figuras - processo de percepção

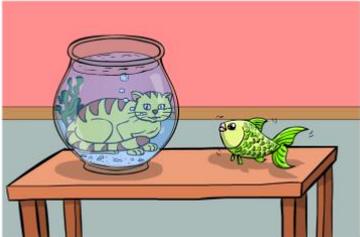
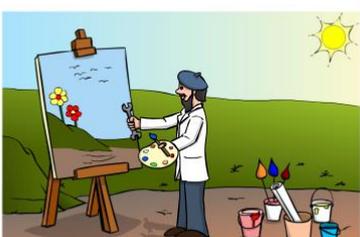
<p>R</p>  <p>< ></p>	<p>R</p>  <p>>></p>
<p>2</p>  <p>>></p>	<p>1</p>  <p>>></p>
<p>3</p>  <p>>></p>	<p>4</p>  <p>></p>
<p>5</p>  <p>></p>	<p>6</p>  <p>>></p>
<p>7</p>  <p>>></p>	<p>8</p>  <p>></p>

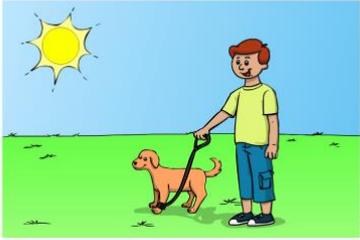
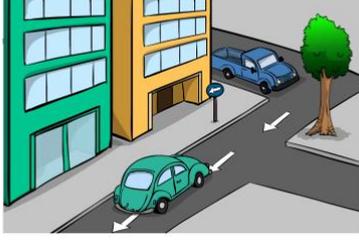
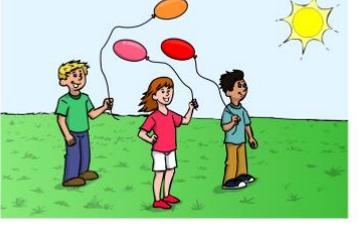
<p>9</p> <p>>></p>	<p>10</p> <p>>></p>
<p>11</p> <p>></p>	<p>12</p> <p>>></p>
<p>12</p> <p>>></p>	<p>13</p> <p>>></p>
<p>14</p> <p>>></p>	<p>15</p> <p>>></p>
<p>16</p> <p>>></p>	<p>17</p> <p>>></p>
<p>18</p> <p>>></p>	<p>19</p> <p>>></p>

<p>20</p>  <p>></p>	<p>21</p>  <p>></p>
<p>22</p>  <p>>></p>	<p>23</p>  <p>></p>
<p>24</p>  <p>></p>	<p>25</p>  <p>>></p>
<p>26</p>  <p>>></p>	<p>27</p>  <p>>></p>
<p>28</p>  <p>></p>	<p>29</p>  <p>>></p>
<p>30</p>  <p>></p>	<p>31</p>  <p>></p>

<p>32</p>  <p>>></p>	<p>33</p>  <p>>></p>
<p>35</p>  <p>>></p>	<p>38</p>  <p>>></p>
<p>34</p>  <p>>></p>	<p>37</p>  <p>>></p>

ANEXO O - Itens da prova Desenhos Absurdos - processo de compreensão

<p>1</p> 	<p>2</p> 
<p>3</p> 	<p>4</p> 
<p>5</p> 	<p>6</p> 
<p>8</p> 	<p>7</p> 
<p>8</p> 	<p>7</p> 

<p>9</p>  <p>></p>	<p>10</p>  <p>>></p>
<p>11</p>  <p>></p>	<p>12</p>  <p>></p>
<p>13</p>  <p>></p>	<p>14</p>  <p>>></p>
<p>15</p>  <p>>></p>	<p>16</p>  <p>>></p>
<p>17</p>  <p>>></p>	<p>18</p>  <p>>></p>
<p>19</p>  <p>>></p>	<p>20</p>  <p>></p>

