UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM
PSICOLOGIA COGNITIVA

RENATO GUEDES DOS SANTOS

CONSTRUÇÃO E VALIDAÇÃO DE VERSÃO INFORMATIZADA DA PROVA DE RACIOCÍNIO DA ECCOS 4/10

RECIFE

2011

RENATO GUEDES DOS SANTOS

CONSTRUÇÃO E VALIDAÇÃO DE VERSÃO INFORMATIZADA DA PROVA DE RACIOCÍNIO DA ECCOS 4/10

Dissertação apresentada ao Programa de Pós-Graduação do Departamento de Psicologia da Universidade Federal de Pernambuco como requisito parcial para obtenção do título de Mestre em Psicologia.

Linha de Pesquisa: Psicologia Cognitiva

Orientador: Prof. Antonio Roazzi

RECIFE
2011

Catalogação na fonte
Bibliotecária: Maria Janeide Pereira da Silva, CRB4-1262

S237c Santos, Renato Guedes dos.
Construção e validação de versão informatizada da prova de raciocínio da ECCOs 4/10 / Renato Guedes dos Santos. – 2011.
73 f. : il. ; 30 cm.

Orientador : Prof. Dr. Antonio Roazzi.
Dissertação (mestrado) - Universidade Federal de Pernambuco, CFCH. Programa de Pós-Graduação em Psicologia Cognitiva, Recife, 2011.
Inclui Referências e anexos.

1. Psicologia cognitiva. 2. Psicometria. 3. Raciocínio. 4. Testes psicológicos para crianças. 5. Testes psicológicos informatizados. 6. Avaliação. 7. Teoria da resposta ao item. I. Roazzi, Antonio (Orientador). II. Título.

153 CDD (22. ed.) UFPE (BCFCH2018-079)

RENATO GUEDES DOS SANTOS

CONSTRUÇÃO E VALIDAÇÃO DE VERSÃO INFORMATIZADA DA PROVA DE RACIOCÍNIO DA ECCOS 4/10

Dissertação apresentada ao Programa de Pós-Graduação do Departamento de Psicologia da Universidade Federal de Pernambuco como requisito parcial para obtenção do título de Mestre em Psicologia Cognitiva.

Aprovado em: 25/02/2011

BANCA EXAMINADORA

_____

Prof. Dr. Antonio Roazzi (Orientador)
Universidade Federal de Pernambuco

_____

Profa. Dra. Suely de Melo Santana (Examinador Externo)
Universidade Católica de Pernambuco

_____

Prof. Dr. Bruno Campello de Souza (Examinador Externo)
Universidade Federal de Pernambuco

RESUMO

A utilização dos recursos de multimídia e interatividade permitidos pelos computadores 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 isso representar um fator de validade dos testes ao torná-los mais atraentes aos sujeitos que os realizam. Por sua vez, a Teoria da Resposta ao Item vem se tornando a técnica predominante para a montagem e interpretação de testes, superando limitações da Teoria Clássica dos Testes e permitindo estabelecer correspondências entre variáveis latentes e suas manifestações. Nesse contexto, este trabalho relata o processo de construção e validação de um teste voltado à avaliação do raciocínio abstrato baseado na Escala de Avaliação de Competências Cognitivas para Crianças dos 4 aos 10 anos de Idade – ECCOs 4/10, uma bateria de testes voltada para avaliação cognitiva na infância. Vantagens, facilidades, limites e dificuldades foram percebidos na construção e aplicação desse instrumento e estão aqui relatados.

Palavras-chave: Psicometria. Raciocínio. Avaliação. Testes psicológicos informatizados. Teoria de Resposta ao Item. ECCOs.

ABSTRACT

The use of multimedia and interactivity allowed by computers promises to revolutionize the reality of psychological tests, particularly those focused on the cognitive assessment in childhood. In the logic of tests that evolve and adapt to technological innovation, we would say that the electronic test represents a attention of psychology to the growth of new technologies, and this could represent a factor of validity of the tests to make them more attractive to individuals who perform them. On the other hand, the Item Response Theory has become the predominant technique for the creation and interpretation of tests, overcoming the limitations of Classical Test Theory and allowing to establish connections between latent variables and their manifestations. In this context, this paper describes the process of constructing and validating a test aimed at assessing abstract reasoning based on the Assessment Scale Cognitive Skills for Children from 4 to 10 years of age - ECCOs 4 / 10, a battery of tests aimed at cognitive assessment in childhood. Advantages, facilities, limits and difficulties were perceived in the construction and application of this instrument and are reported here.

Keywords: Psychometrics. Reasoning. Evaluation. Computerized psychological testing. Item Response Theory. ECCOs.

LISTA DE FIGURAS

Figura 1 – Curvas Características do Item – CCI.....24

Figura 2 – Mapa comparativo entre pessoas e itens no continuum da variável latente – pré-teste.....47

Figura 3 – Mapa comparativo entre pessoas e itens no continuum da variável latente.....52

Figura 4 – Boxplot idade x pontuação total.....55

Figura 5 - Menu Principal.....62

Figura 6 - Cadastro: identificação e indicadores socioeconômicos.....62

Figura 7 - Cadastro: dados escolares e médicos.....63

Figura 8 - Exemplo de exercício.....63

Figura 9 - Exemplo de exercício.....64

Figura 10 - Exemplo de exercício.....64

LISTA DE QUADROS

Quadro 1 – Estrutura e provas da ECCOs 4/10.....16

Quadro 2 - Transformações subjacentes aos itens da prova de Imagens Incompletas.....18

Quadro 3 – Distribuição da população por idade e sexo – pré-teste.....45

Quadro 4 – Estatísticas dos itens.....48

Quadro 5 – Classificação dos participantes por idade e sexo.....51

LISTA DE TABELAS

Tabela 1 – Sumário da medida de 272 pessoas.....53

Tabela 2 – Sumário da medida de 42 itens.....53

Tabela 3 – Saída da Análise dos Principais Componentes (em eigenvalue).....54

LISTA DE ABREVIATURAS E SIGLAS

APC	Análise dos Principais Componentes
CCI	Curva Característica do Item
CNS	Conselho Nacional de Saúde
DIF	Índice de Disfunção Cultural
ECCOs 4/10	Escala de Avaliação de Competências Cognitivas para Crianças dos 4 aos 10 Anos de Idade
MNSQ	Média-Quadrada
TCT	Teoria Clássica dos Testes
TRI	Teoria da Resposta ao Item
UFPE	Universidade Federal de Pernambuco
UMinho	Instituto de Educação e Psicologia da Universidade do Minho
ZSTD	Estatísticas Padronizadas

SUMÁRIO

1 INTRODUÇÃO 12

2 ARCABOUÇO TEÓRICO 15

2.1 A ECCOs 4/10 E A PROVA DE RACIOCÍNIO NÃO VERBAL 15

2.2 A MEDIDA EM PSICOLOGIA 20

2.3 A TEORIA DA RESPOSTA AO ITEM – TRI 23

2.3.1 Unidimensionalidade e independência local 26

2.3.2 Modelos 27

2.4 O MODELO DE RASCH 28

2.4.1 A análise dos itens – validade e fidedignidade 30

2.4.2 Separação 33

2.4.3 As estatísticas de aderência ao modelo 34

2.4.4 Unidimensionalidade em Rasch 35

2.5 RELAÇÃO ENTRE O COMPUTADOR E A TESTAGEM 37

3 CONSTRUÇÃO DO INSTRUMENTO DE COLETA DE DADOS 40

3.1 ELABORAÇÃO DO INSTRUMENTO 40

3.1.1 Linguagem de programação 40

3.1.2 Modelagem da base de dados 41

3.1.3 Interface 41

4 A PESQUISA DE CAMPO 43

4.1 OBJETIVOS 43

4.2 PROCEDIMENTOS E INSTRUMENTOS UTILIZADOS 44

4.3 PRÉ-TESTE 45

4.3.1 Participantes 45

4.3.2 Observações e recomendações oriundas do pré-teste 46

4.4 A APLICAÇÃO DO TESTE FORMAL 50

4.4.1 Participantes 50

4.4.2 Análise dos dados 51

4.5 CONCLUSÕES 55

4.6 CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES 57

REFERÊNCIAS 59

APÊNDICE A – TELAS DE EXEMPLO DA PROVA FIGURAS INCOMPLETAS DA ECCOs INFORMATIZADA 62

APÊNDICE B – ESTRUTURA DE DADOS DA PROVA DE FIGURAS INCOMPLETAS DA ECCOs 65

APÊNDICE C – ESTATÍSTICAS DOS ITENS DO SPSS 66

APÊNDICE D – ESTATÍSTICA DOS ITENS DO WINSTEPS 67

APÊNDICE E – CURVAS DO TESTE 68

APÊNDICE F – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO 69

APÊNDICE G - PROCEDIMENTOS PARA APLICAÇÃO DO TESTE (1/2) 71

APÊNDICE H – PROCEDIMENTOS PARA APLICAÇÃO DO TESTE (2/2) 72

APÊNDICE I – FICHA DE INFORMAÇÃO DO ALUNO 73

1 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. Na revisão liberada em 2009, fruto do trabalho da Dra. Maria de Lurdes Dias Brito sob a orientação do Prof. Dr. Leandro Almeida, ela foi denominada *Escala de Avaliação de Competências Cognitivas para Crianças dos 4 aos 10 Anos de Idade – ECCOs 4/10*. Ao final do trabalho, entre os objetivos que os autores estabeleceram para dar continuidade a esse grande projeto constavam a adaptação da escala à população de outros países de expressão portuguesa e a elaboração de uma versão computadorizada.

Nesse contexto, o Departamento de Psicologia da Universidade Federal de Pernambuco – UFPE, parceiro brasileiro da UMinho em pesquisas no campo da psicologia cognitiva, em 2009 começou a validar essa escala para a população brasileira, o que permitirá ampliar o público usuário potencial da escala, bem como sua importância para psicólogos de ambos os países.

No mesmo sentido, a missão deste trabalho foi construir e validar uma versão computadorizada da prova Figuras Incompletas da ECCOs 4/10, uma prova não verbal destinada à avaliação do raciocínio abstrato, indo ao encontro da recomendação de seus autores e colaboradores, do interesse dos dois centros acadêmicos envolvidos, bem como das necessidades de profissionais que se valem da mensuração de habilidades e processos cognitivos em crianças no seu dia-a-dia, como psicólogos e educadores.

De fato, os testes de raciocínio são considerados o núcleo da maioria dos testes de inteligência (Snijders, Tellegen, & Laros, 1989, p. 21). Os que se destinam à avaliação do raciocínio concreto procuram estabelecer uma conexão entre pessoas ou objetos que se encontram relacionados no tempo ou no espaço. Já os testes que avaliam o raciocínio abstrato tentam estabelecer relações entre conceitos abstratos de forma que o sujeito extraia um princípio de ordem a partir do material que lhe é apresentado para aplicá-lo a um novo material por analogia. Segundo Brito (2009), na prova Figuras Incompletas da ECCOs 4/10 o raciocínio abstrato também é avaliado através de tarefas analógicas à semelhança do que acontece com outras provas de avaliação cognitiva.

Espera-se que as versões computadorizadas dos testes psicológicos aumentem a atenção e a motivação das crianças na medida em que utilizem recursos multimídia (produzindo itens

até então impossíveis), permitam a elas terem a sensação de controle do teste ao sequenciar a apresentação dos itens no seu ritmo ou desencadeando movimentos na tela a partir das suas ações, como se elas estivessem brincando com o computador. Esse pode ser um caminho viável para a produção de resultados mais verdadeiros nos testes com crianças, resultados menos influenciados pelo desânimo consequente da passividade muitas vezes presente nos testes tradicionais com papel. Paralelamente, são óbvios os ganhos de produtividade no processo informatizado face à automação da coleta de dados e geração de relatórios.

Quando delineamos o projeto para informatização da prova Figuras Incompletas da ECCOs 4/10, concluímos que todos os itens presentes na prova original, em papel, deveriam ser revalidados para a versão informatizada, visto que as crianças poderiam se comportar de maneira diferente diante do computador. Dessa forma, mais do que comparar os resultados obtidos na prova original com aqueles que viessem a ser obtidos na versão eletrônica, haveríamos de validar os próprios itens, seus níveis de dificuldade e sua ordem de apresentação. Uma vez que todos os itens seriam reavaliados, decidimos introduzir no trabalho mais uma inovação: a Teoria da Resposta ao Item (TRI).

Segundo Pasquali (2003), a Teoria da Resposta ao Item 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. Superando algumas limitações da Teoria Clássica dos Testes (TCT), padrão vigente para construção de testes até então, a TRI reúne modelos estatísticos usados para fazer predições, estimativas ou inferências sobre as habilidades (competências, aptidões) medidas em um teste, permitindo-nos estabelecer uma correspondência entre variáveis latentes e suas manifestações (Ayala, 2009).

De uma forma geral, o objetivo deste trabalho foi construir uma versão informatizada da prova de avaliação do raciocínio não verbal denominada Figuras Incompletas da ECCOs 4/10 à luz da Teoria da Resposta ao Item. Importou validar seus resultados em comparação com a prova original, procurando identificar se as características psicométricas foram mantidas e se a utilização da versão informatizada se traduziu em benefícios para as crianças e psicólogos, detectando eventuais problemas e limites processuais ou tecnológicos.

E assim foi que a partir de uma prova tradicional, concebida na perspectiva da Teoria Clássica dos Testes e validada para a população portuguesa, construímos uma versão computadorizada dessa prova à luz da Teoria da Resposta ao Item validando-a com representantes da população brasileira. A pesquisa constatou vantagens na aplicação de testes

informatizados e maior engajamento das crianças. A utilização da TRI contribuiu para ampliar o entendimento sobre a prova e alunos, inclusive demonstrando onde se pode mexer na prova para se obter uma escala mais aprimorada.

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 estão aqui relatados.

Esta dissertação está estruturada em três seções de forma a conduzir o leitor através do roteiro que foi seguido para a produção dos resultados aqui descritos. As duas primeiras seções estão voltadas para o arcabouço teórico do trabalho e para a produção da ferramenta informatizada, que representa a própria prova Figuras Incompletas. A última seção descreve a parte mais empírica do projeto, concentrando a aplicação dos testes com as crianças e a análise e interpretação dos seus resultados.

Assim, inicialmente apresentaremos a ECCOs 4/10 cuja prova de raciocínio abstrato, denominada Figuras Incompletas, foi o alvo do nosso trabalho de informatização e validação, seguindo-se da conceituação de outros elementos necessários para a compreensão e execução do trabalho: a medida em psicologia, a Teoria da Resposta ao Item, o Modelo de Rasch e a relação da testagem com os computadores.

Em seguida, descrevemos o processo de construção da pesquisa, desde a explicitação de seus objetivos até a elaboração da prova informatizada, passando pelos procedimentos, abordando as características da base de dados, os aspectos técnicos da sua construção, os pressupostos assumidos e a programação no computador.

Entrando na parte empírica, relatamos a aplicação da prova na fase de pré-teste para depois descrever a aplicação da prova, seguindo-se das conclusões e das considerações finais e recomendações.

Anexados ao trabalho temos 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 alunos, os procedimentos para aplicação do teste e a ficha de informação acadêmica do aluno.

2 ARCABOUÇO TEÓRICO

2.1 A ECCOs 4/10 E A PROVA DE RACIOCÍNIO NÃO VERBAL

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, constituindo uma versão ampliada e adaptada de sua congênere, a ECCOs 4/7. Aferida à zona Norte de Portugal continental, é o resultado de um longo trabalho iniciado no mestrado e continuado no doutorado da Dra. Maria de Lurdes Dias Brito sob a orientação do Prof. Dr. Leandro Almeida, do Instituto de Educação e Psicologia da Universidade do Minho - UMinho, de Portugal. Sua última revisão data de janeiro de 2009.

Sabe-se que a inteligência tem sido um dos construtos mais estudado pela psicologia. De fato, segundo Roazzi (2008), "o início do estudo científico da inteligência confunde-se com as próprias origens da psicologia como ciência, ambos ocorrendo entre o final do século XIX até o início do XX". Mas, apesar de tamanha relevância, Brito (2009) nos diz que 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.

Confluindo diversas posturas teóricas sobre a questão da inteligência, 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), falando sobre a ECCOs 4/10, "subjacente ao modelo de avaliação apresentado está a ideia de uma inteligência compósita, abordando processos cognitivos mais simples até processos cognitivos mais exigentes e 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 & Almeida (2009) nos contam que na ECCOs 4/10 procurou-se 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 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. Nesse sentido, 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 dois tipos de conteúdos de tarefas cada um, 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.

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

Quadro 1 – Estrutura e provas da ECCOs 4/10

Os autores da escala procuraram afixar as provas em uma sequência temporal de execução com uma exigência cognitiva cada vez mais complexa.

> Partimos 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, p.186).

Detendo-nos no raciocínio, processo cognitivo que aqui mais nos interessa, desde cedo ele esteve presente nas diversas teorias da inteligência. Segundo Brito & Almeida (2008), em diferentes estudos o raciocínio, nominado ou não, 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. Na ECCOs 4/10, busca-se avaliar o raciocínio 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). Trata-se de um processo pelo qual se avaliam e geram argumentos lógicos, podendo assumir diferentes formas:

(i) raciocínio dedutivo, que obedece às regras da lógica matemática, seguindo-se das premissas à conclusão através da aplicação de um conjunto de regras inferenciais;

(ii) raciocínio condicional, sujeito a argumentos que impõem constrangimentos às conclusões;

(iii) raciocínio silogístico, que exige o recurso a silogismos categoriais e envolve noções de todo, algum, não e algum não, que condicionam a conclusão;

(iv) raciocínio indutivo, que não segue as regras formais da lógica, estando sujeito a condições ou incertezas que impedem que a conclusão seja válida, tornando-a somente provável.

As tarefas propostas para avaliação do raciocínio devem considerar a diversidade de processos cognitivos que o mesmo pode apresentar. 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.

Segundo Brito (2009), as provas de raciocínio da ECCOs 4/10 são provas de analogias que abordam conteúdos verbais (não considerado neste estudo) e conteúdos não verbais. Tal como em outras provas, alguns itens são compostos por imagens de objetos pertencentes ao quotidiano da criança, noutras as imagens são figuras geométricas sem significado figurativo. Os itens vão crescendo seus níveis de dificuldade, 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.

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

Quadro 2 - Transformações subjacentes aos itens da prova de Imagens Incompletas

Para compreender a lógica implícita na prova de raciocínio, vamos utilizar a descrição de Brito (2009, p.65) 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". Em outras palavras, diante de três termos (figuras) exibidos para a criança ela terá que escolher uma das cinco opções (figuras) de resposta para representar o quarto termo que completará a analogia. Temos três processos aqui implicados, segundo Sternberg (1992):

(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;

No âmbito deste trabalho, nos coube construir uma versão informatizada da prova Figuras Incompletas da ECCOs 4/10 que se destina a avaliar o raciocínio abstrato utilizando-se

de analogias, tal como descrito. A escolha desta como a primeira prova da ECCOS 4/10 a ser informatizada justifica-se, primeiro, pela importância que o raciocínio costumeiramente representa nos testes de inteligência, chegando a ser considerado "o núcleo principal da maioria dos testes de inteligência" (Snijders, Tellegen, & Laros;1989). Segundo, porque sua transposição para o computador, em teoria, estaria sujeita a um número menor de obstáculos, haja vista se tratar de uma prova não verbal totalmente baseada em figuras, dispensando, por exemplo, a narração dos itens para crianças que ainda não dominam a leitura. Em terceiro lugar, a prova Figuras Incompletas dispõe de 43 itens, uma quantidade razoável frente às demais provas, quantidade esta considerada suficiente para permitir eventuais exclusões de itens sem grandes prejuízos ao teste como um todo e até mesmo a montagem de futuros testes dinâmicos.

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 produz análises entre o esperado e o encontrado, servindo para orientar intervenções. A autora destaca, ainda, que os resultados da avaliação cognitiva também constituem um bom preditor do desempenho, nomeadamente nas situações escolares de aprendizagem e rendimento (Seabra-Santos, 2000) ou até da capacidade intelectual adulta (Bracken &Walker, 1997).

A versão informatizada da prova Figuras Incompletas não modificou a proposta da prova nem descartou nenhum item a priori. Foi mantida exatamente a mesma lógica, os mesmos itens originais, com uma ligeira atualização de estilo, e a mesma sequência de apresentação. A única diferença é que como nosso objetivo maior neste trabalho não era avaliar a inteligência das crianças, mas verificar a viabilidade e os impactos de uma prova informatizada, não seguimos 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.

Devemos ressaltar que temos consciência de que algumas outras provas, como a Desenhos Absurdos, por exemplo, poderiam proporcionar o uso intenso de recursos computacionais inovadores nos testes, em particular a exibição de itens multimídia, com sons e animações, o que, certamente, poderia exercer grande poder de atração em crianças, mas que demandariam prazo e esforço para concepção e elaboração incompatível com o curso de mestrado em questão.

2.2 A MEDIDA EM PSICOLOGIA

A medição é uma atividade fundamental da ciência. Na verdade, 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. Dentro das ciências comportamentais e sociais, a psicometria, ou a medida em psicologia, tem se destacado como uma especialização em benefício da medição de fenômenos psicológicos e sociais (DeVellis, 2003). Normalmente, o procedimento de medição envolve testes (ou provas, ou questionários) e as variáveis de interesse são parte de um amplo referencial teórico, procurando-se explicar o sentido que têm as respostas dadas pelos sujeitos a uma série de tarefas, tipicamente chamadas de itens, normalmente agrupadas em um ou mais teste. 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).

A definição clássica para o termo "medição" ou "mensuração" é aquela formulada por Stevens (1946, p.677): "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". No entanto, essa é uma visão bastante limitada. Medir é 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 (cf. Bridgman, 1928) aplicando técnicas matemáticas.

De acordo com Pasquali (2003, p.14), a origem da psicometria deve ser procurada nos trabalhos do estatístico Charles Spearman (1863-1945) que, no que se refere à psicologia, seguiu os procedimentos fisicalistas da obra de Francis Galton (1822-1911). 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, p.23) ensina que a psicometria se insere dentro da teoria da medida em geral que, por sua vez, esta desenvolve uma discussão epistemológica sobre a utilização do número no estudo científico dos fenômenos naturais. Trata-se, portanto, 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. Neste ponto é importante salientar que como 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.

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 em psicologia são inteligência, extroversão, desajustamento, ciúmes, ansiedade e acuidade visual. Os sujeitos que estudamos variam ao longo dessas variáveis, apresentando escores que podem ser altos ou baixos, existir em maior ou menor quantidade ou diferenciar-se segundo um outro conjunto similar de quantificadores.

As variáveis que não podem ser medidas diretamente pela observação são chamadas variáveis latentes. Ayala (2009, p.9) diz que, ao contrário da informação colhida diretamente da observação do comportamento, o termo latente refere-se 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 ou a Teoria Clássica dos Testes, podem ser aplicadas para explicar comportamentos observáveis a partir da perspectiva de variáveis latentes contínuas. Ressaltamos que as variáveis latentes categóricas devem ser analisadas utilizando a Análise de Classe Latente (LCA), assunto que não será abordado neste estudo.

O conceito de traço latente, por sua vez, é permeado por ambiguidades e controvérsias, bastando ver a quantidade de expressões que o representam. Pasquali (2003, p.55) 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 este construto ou parâmetro.

Deixando de lado as diversas discussões, inclusive ontológicas, sobre o traço latente, mas dando-lhe algum conteúdo psicológico ao invés de puros conceitos ou parâmetros matemáticos, vamos encarar o traço latente como um processo psicológico. Pasquali (2003, p.61) esclarece que "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 são estabelecidas as correspondências entre os dados observados e a localização das pessoas na variável latente. Uma vez que tenhamos nossos sujeitos localizados na variável latente nós poderemos compará-los com outros sujeitos.

Como parte da medição, também é 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.

Segundo Pasquali (2003, p.39-45), 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 do sujeito, ou seja, o comportamento, e o critério. Na TCT critério também é entendido como comportamento (futuro) e na TRI como 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 medida. Os testes devem ser fidedignos, ou seja, devem ser capazes de medir sem erros. "Medir sem erros significa que o mesmo teste, medindo os mesmos sujeitos em ocasiões diferentes, ou testes equivalentes, medindo os mesmos sujeitos na mesma ocasião, produz resultados idênticos" (Pasquali, 2003, p.192).

O segundo ponto se refere à validade da medida. Embora existam várias discussões sobre validade de um teste, nos manuais de psicometria costuma-se definir 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, nós 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, uma vez que a ausência da propriedade de invariância faz com que as comparações entre os diferentes sujeitos tenham 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.

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

2.3 A TEORIA DA RESPOSTA AO ITEM – TRI

A Teoria de 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 na 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). Ressaltamos que o conceito de que um item tenha uma localização não é novo, podendo ser visto em Thurstone (1925, 1928), Andrich (1978), Lumsden (1978) e Yen (1986).

Até o surgimento da TRI, nos anos 1950, e sua popularização nos anos 1980, predominava na construção de testes a Teoria Clássica dos Testes (TCT), vinculada à psicometria tradicional. Mas todos os indícios são de que a TRI veio para ficar e substituir grande parte da teoria clássica da psicometria. Segundo Hambleton, Swaminathan e Rogers

(1991), o enorme impacto da TRI se deve ao fato dela superar certas limitações teóricas graves que a psicometria tradicional contém, em especial a dependência dos sujeitos da amostra (group-dependent ou sample-dependent), a incapacidade de produzir os mesmos escores com os mesmos sujeitos 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. Por sua vez, a TRI avalia as características do item, não dependendo da amostra de sujeitos, fornece escores semelhantes independentemente do teste utilizado, orienta a análise de cada item individualmente, permite acrescer 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, a TRI adota dois axiomas fundamentais:

(i) o desempenho do sujeito em uma tarefa se explica em função de um conjunto de traços latentes. 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 sujeitos com aptidão maior terão maior probabilidade de responder corretamente ao item e vice-versa (Θi é a aptidão e Pi(Θ) a probabilidade de responder corretamente ao item).

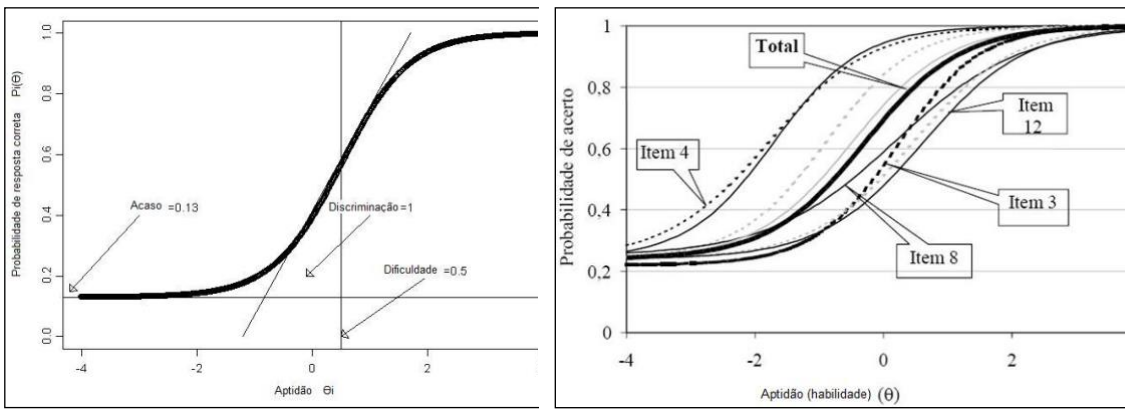


Figura 1 – Curvas Características do Item – CCI

Dessa forma, analisando as respostas dadas por uma pessoa a uma série de itens podemos inferir sobre seu traço latente, hipotetizando relações desse sujeito com seu nível de

traço latente através de uma equação matemática que descreve a forma de função que estas relações assumem.

Quando a TRI é apropriada e quando o modelo adere aos dados, então a TRI oferece vantagens 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 a 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 não sustentabilidade do modelo em relação aos dados.

Ayala (2009, p.4) 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 predizer as respostas observadas. Os processos cognitivos usados por um indivíduo para responder a um item não são modelados na maioria dos modelos de TRI.

A TRI vem se tornando o modelo psicométrico básico no campo dos testes, sendo a técnica predominante na construção dos testes adaptativos (ou sob medida) (Pasquali, 2003). Ela reúne modelos estatísticos usados para fazer predições, estimativas ou inferências sobre as habilidades (ou competências) medidas em um teste. 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 escore 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 escore 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.

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, pp.95-60), 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.

2.3.1 Unidimensionalidade e independência local

Entre as suposições para a aplicação de modelos baseados na TRI, duas são de especial relevância: a unidimensionalidade e a independência local, que discutiremos a seguir.

Sobre a independência local, o postulado afirma que se forem mantidas constantes as aptidões que afetam o teste, menos o teta dominante, as respostas dos sujeitos a quaisquer dois itens são estatisticamente independentes. Ou seja, o desempenho do sujeito num item não afeta seu desempenho em outro item: cada item é 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 postulam que deve existir apenas uma aptidão (teta) responsável pela realização de um conjunto de tarefas (itens de um teste) para que a medição possa ser realizada – na verdade um postulado intrínseco a qualquer tipo de medida. Por outro lado, como afirmam Pasquali & 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 vem dos 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.

Na verdade, 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. Como podemos ver, 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 falar da unidimensionalidade no tópico do modelo de Rasch.

2.3.2 Modelos

Para que possamos extrair as características de itens e pessoas e predizer as chances que as pessoas terão de responder acertadamente determinado item de um teste, como proposto pela TRI, precisamos vincular um modelo que possa suportar essas capacidades e represente 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). 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 quantificada. Nos modelos ML2, quantifica-se a dificuldade e a capacidade de discriminação dos itens. 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 das questões (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 muito popular atualmente é o modelo de Rasch (1960) de 1 parâmetro. Esse foi o modelo utilizado na pesquisa aqui relatada e será detalhado a seguir.

2.4 O MODELO DE RASCH

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 utilizamos, visto que a maior parte da análise desta pesquisa depende dessa compreensão.

Segundo Bond & Fox (2007), para um modelo ser útil à investigação de aspectos da condição humana representada nas teorias do desenvolvimento, assim como em outras teorias, 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 P1 é mais desenvolvida, mais capaz, ou mais reabilitada do que a pessoa P2);

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. Georg Rasch, um matemático dinamarquês, formulou sua abordagem em 1953 para analisar as respostas a uma série de testes de leitura (Rasch G, Probabilistic Models for Some Intelligence

and Attainment Tests, Chicago: MESA Press, 1992, com prefácio e posfácio instrutivo de B.D. Wright).

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 (Rasch, 1960/1980) é um modelo de probabilidade binomial para uma resposta dicotômica, 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:

a person having a greater ability than another person should have the greater probability of solving any item of the type in question, and similary, once item being more difficult than another means that for any person the probability of solving the second item is the greater one. (Rasch, 1960, p.117)

Ou seja, uma pessoa com habilidade maior do que outra deverá ter uma probabilidade maior de resolver qualquer item do mesmo tipo e, do mesmo modo, se temos um item mais difícil do que outro significa que para qualquer pessoa a probabilidade de resolver o segundo item será maior do que a probabilidade de resolver o primeiro.

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 & Fox (2007), a característica central do modelo de Rasch é uma tabela de probabilidades esperadas desenhada para responder à seguinte questão central: quando uma pessoa com habilidade x encontra um item com dificuldade y, qual é a probabilidade dessa pessoa responder corretamente ao item? 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, 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 & Fox (2007), as estimativas das habilidades das pessoas e das dificuldades dos itens são apresentadas ao longo de uma escala logit (log adds 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%.

2.4.1 A análise dos itens – validade e 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 (multidimensionalidade) ou por conter vieses a favor ou contra gênero, grupos étnicos e outros.

Fidedignidade (reliability) e validade são assuntos fundamentais para as medidas em psicologia. Enquanto fidedignidade se refere a quanto uma variável influencia um conjunto de itens, 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, stress 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 o que ele se propõe a medir.

Segundo DeVellis (2003), 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 (1951), correspondendo fortemente à clássica definição de fidedignidade como proporção da variância em uma escala que pode ser atribuída ao escore verdadeiro da variável latente. O Winsteps implementa uma versão especial do alfa para itens que são dicotômicos chamado Kuder-Richardson fórmula 20 (KR-20). Através do software SPSS é possível extrair um interesse indicador da fidedignidade do teste baseado no alfa: 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 sujeito" 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 um pouco.

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 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 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.

2.4.2 Separação

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

A fidedignidade para os sujeitos 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) dos sujeitos sobre esse construto, ou seja, permita identificar a separação dos sujeitos (Fox & Jones, 1998). Portanto, alta fidedignidade dos sujeitos significa que temos uma escala em que algumas pessoas terão maior pontuação e outras pessoas pontuações menores e que nós podemos esperar coerência destas 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). O índice de confiabilidade é equivalente ao alfa de Cronbach e o KR20. É a estimativa de replicabilidade de posicionamento dos itens dentro da hierarquia de dificuldade de alunos 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 & Fox (2007), no modelo de Rasch a separação é estimada para pessoas e itens. O índice de separação dos sujeitos (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 do sujeito 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.

2.4.3 As estatísticas de aderência ao modelo

As estatísticas de aderência (fit) servem para estimar a magnitude do desajuste dos dados das pessoas e itens ao modelo. Segundo Linacre (2002, p.878), no contexto de Rasch as estatísticas que indicam precisão e permitem previsibilidade de aderência dos dados ao modelo são:

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 confidência 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.

2.4.4 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 (*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. Consequentemente, 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 do modelo de Rasch é que as informações dos dados sejam explicadas pelas variáveis latentes. Então, a parte 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 Cattel 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. No caso da TRI, o modelo precisa ser unidimensional. Linacre (1998) sugere três passos para investigar a dimensionalidade dos dados em Rasch e, consequentemente, definir a aderência dos dados ao modelo:

(1) Identificar e examinar correlações point-biseriais 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.

2.5 RELAÇÃO ENTRE O COMPUTADOR E A TESTAGEM

O computador tem 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.

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. Os processos 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.

Na testagem dinâmica (também chamada de adaptativa ou sob medida) "o computador não só apresenta os itens, como também seleciona o próximo item a ser apresentado com base nas respostas anteriores do participante" (Hogan, 2006, p.27), produzindo testes únicos, otimizados para níveis de habilidades diferenciadas e, portanto, mais adequados para expressar nossa diversidade de processos cognitivos. Os testes dinâmicos estão revolucionando o campo da testagem, representando uma das áreas que mais cresce atualmente, ainda segundo Hogan (2006).

A utilização do computador para aplicação de testes também vem propiciando a utilização de uma nova categoria de itens: os itens multimídia. Esse tipo de item pode englobar estímulos visuais, como figuras, vídeos e animações, estímulos sonoros e até mesmo táteis empregando joysticks com vibração. Santos (2007, p.43) relata que "a utilização de computadores com crianças normalmente resulta em aumento da motivação, diminuição do

comportamento inadequado, 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.

Mas 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 proporcionar. Não há 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, 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.

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 aos sujeitos que os realizam. De fato, nos dias de hoje, as crianças estão mais interessadas em manipular equipamentos do que escrever com lápis e papel.

Dessa forma, nos interessou investigar se a utilização do computador favoreceu a aplicação do teste, seja pelo fato de substituir o papel, seja pelo aumento do caráter lúdico do teste permitido pela ferramenta computacional. Será o teste mais prazeroso ou mais estimulante para o aluno quando aplicado através do computador?

Desejou-se compreender se a aplicação da prova no computador aumentará a atenção e concentração na tarefa por parte do aluno, haja vista que a criança passa a ser um elemento ativo no teste, inclusive com controle da passagem dos itens, ao contrário da situação passiva normalmente verificada nas provas tradicionais quando o psicólogo mostra os itens às crianças

e elas simplesmente indicam a resposta. Teremos menos abandonos nos testes, com mais crianças fazendo o teste motivadas até o fim? A prova informatizada poderá ser aplicada em grupos, visto não carecer de tanta assistência do psicólogo quanto a prova tradicional?

3 CONSTRUÇÃO DO INSTRUMENTO DE COLETA DE DADOS

3.1 ELABORAÇÃO DO INSTRUMENTO

Para construir a versão informatizada da prova Figuras Incompletas da ECCOs 4/10 consideramos a existência de duas tarefas bem distintas no processo: coleta e interpretação de dados. Enquanto a coleta dos dados 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, no âmbito deste trabalho, quando falamos em informatizar a prova estamos falando em construir um instrumento informatizado para coleta de dados, deixando a análise e interpretação dos dados para ser feita com a ajuda de programas estatísticos consagrados para essa função, como o SPSS ou o Winsteps, a partir dos dados exportados pelo instrumento de coleta de dados criado.

3.1.1 Linguagem de programação

O programa de coleta de dados da versão informatizada da prova Figuras Incompletas foi construído na linguagem ActionScript. A ActionScript é uma linguagem de programação orientada a objetos, baseada em scripts e padronizada pela *European Computer Manufacturers Association*. É ela que está por trás do Adobe Flash, um consagrado recurso para exibição de animações na internet, sendo utilizada principalmente para construção de aplicações ricas na internet rica (do inglês "Rich Internet Applications").

O Adobe Flash, ou simplesmente Flash, é um software primariamente de gráfico vetorial utilizado geralmente para a criação de animações interativas que funcionam embutidas em um navegador web. Os programas criados em Flash propiciam grande portabilidade, podendo funcionar em plataformas e dispositivos diferentes, tais como nos sistemas operacionais Windows, Linux, Macintosh, Android ou Symbian, através de computadores, smartphones, tablets e outros dispositivos eletrônicos.

Apesar de o Flash apresentar algumas limitações para armazenamento local de dados, a portabilidade, a leveza dos gráficos, a capacidade de incorporar imagens, músicas e vídeos, a possibilidade de funcionar em páginas na internet e os recursos de interatividade pesaram a

favor da escolha pelo Flash / ActionScript como linguagem e ambiente para desenvolvimento da prova informatizada da ECCOs.

3.1.2 Modelagem da base de dados

Conforme Brito (2009, p.17), "a informação recolhida através dos testes deve ser complementada por informações veiculadas através dos pais ou dos educadores/professores". Diz, ainda, que a própria avaliação psicológica não pode se limitar à aplicação e cotação dos testes, devendo-se integrar à informação recolhida pela observação do próprio psicólogo em momentos diversos do processo avaliativo.

Dessa forma, conforme pode ser observado na estrutura de dados apresentada no Apêndice B, foi idealizada uma base de dados que além de registrar as informações pertinentes à realização e desempenho no próprio teste, contempla campos para registro de informações adicionais que sejam relevantes para a interpretação dos resultados no tocante aos aspectos socioeconômicos, acadêmicos e de eventuais comprometimentos físicos, sensórios, psicológicos ou mentais da criança. Esses campos serão preenchidos através das respostas obtidas com professores e pais dos alunos que participaram da pesquisa.

3.1.3 Interface

Não podemos esquecer que o programa terá sempre dois perfis 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.

Havemos de considerar, também, 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. Dessa forma, concebeu-se 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 Apêndice A, nas figuras numeradas de 8 a 10.

Adicionalmente à simplicidade da interface, utilizamos monitores com telas sensíveis ao toque (*touch screen*) para a coleta de dados, possibilitando às crianças interagirem com o teste simplesmente apontando com o dedo na tela a opção desejada.

4 A PESQUISA DE CAMPO

4.1 OBJETIVOS

De uma forma geral, este trabalho teve por objetivo construir uma versão informatizada da prova de avaliação do raciocínio não verbal da ECCOs 4/10, denominada Figuras Incompletas, à luz da TRI. Importou validar seus resultados em comparação com a prova original, detectando eventuais problemas e limites processuais ou tecnológicos, além de procurar identificar se as características psicométricas do teste foram mantidas e se a versão informatizada ampliou os benefícios para crianças e psicólogos.

Antes de tudo, é necessário destacar a incorporação de dois novos e importantes eixos de trabalho em relação à ECCOs 4/10 que se traduzem em objetivos específicos complementares, mas que têm origens conceituais diferentes: (1) a própria informatização da prova e (2) a introdução da TRI como conceito e ferramenta de suporte e análise para a definição do teste. Tratam-se de dois assuntos ainda pouco explorados no campo dos testes psicológicos no Brasil e, portanto, esperamos que os ganhos decorrentes do domínio da TRI e da informatização de testes transcendam este trabalho e possam ser replicados para outros estudos.

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 aos sujeitos que os realizam.

Dessa forma, nos interessava 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. De fato, nos dias de hoje, as crianças estão mais interessadas em manipular equipamentos do que escrever com lápis e papel. Seria o teste mais prazeroso ou mais estimulante para o aluno quando aplicado através do computador?

Queríamos compreender 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, ao contrário da situação passiva normalmente verificada nas provas tradicionais, quando o psicólogo mostra os itens às crianças e elas simplesmente indicam a resposta. Teremos menos abandonos nos testes, com mais crianças fazendo o teste motivadas até o fim? Também pretendíamos verificar se a prova

informatizada poderia ser aplicada em grupos, visto não carecer de tanta assistência do psicólogo quanto a prova tradicional.

A TRI é outra novidade fundamental que incorporamos ao trabalho e que nos conduz a outros objetivos específicos. Recorrer à TRI para a construção e validação de testes psicológicos é uma tendência mundial. A TRI retira o foco no teste como um todo, voltando-se ao conhecimento específico sobre cada item que forma o teste, expressando, através de modelos matemáticos, a relação entre a probabilidade de acerto no item e a aptidão medida pelo teste, identificando o posicionamento do item dentro do *continuum* da habilidade que se deseja avaliar, mapeando e tratando eventuais lacunas ou sobreposições dos itens.

Dessa forma, nosso interesse era ser capaz de compreender como montar provas com melhor distribuição dos itens e de predizer a chance de a criança ter sucesso diante de cada novo item. Através da abordagem da TRI teremos novas possibilidades de interpretação sobre os dados coletados, culminando com a possibilidade de no futuro construirmos testes computadorizados sob medida (também chamados de dinâmicos ou adaptativos), onde o computador seleciona cada novo item que será apresentado em função do nível de habilidade que o participante vai demonstrando ter. Também nos interessava identificar se a prova concebida no conceito da Teoria Clássica dos Testes estaria bem estruturada quando olhada pela ótica da Teoria da Resposta ao Item.

4.2 PROCEDIMENTOS E INSTRUMENTOS UTILIZADOS

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

O software desenvolvido ao longo deste projeto serviu como instrumento de coleta de dados, sendo utilizado em um computador com tela de 21 polegadas sensível ao toque, onde os alunos podiam indicar sua escolha apontando na tela a opção desejada para cada item. Quando o teste foi aplicado para grupos de alunos utilizou-se outros três computadores com mouse ao invés de tela sensível ao toque. A orientação para utilização do software está descrita Apêndice G. A aplicação do teste ocorreu dentro do ambiente escolar.

Para fins de análise do teste foram utilizados os programas SPSS versão 18 e Winsteps Rasch Measurement versão 3.70.1.1, este destinado exclusivamente ao auxílio do pesquisador

para muitas aplicações baseadas na TRI utilizando o modelo de Rasch, particularmente nas áreas de testes educacionais, pesquisas de comportamento e análise de escalas.

Como a base de dados modelada para o programa de coleta de dados prevê a inserção de alguns dados referentes ao desempenho acadêmico e a percepção dos professores sobre os alunos, bem como indicadores da situação socioeconômica dos alunos, através do nível de escolaridade e profissão dos pais e mães desses alunos, solicitamos previamente às escolas o preenchimento da ficha de informação do aluno (vide Apêndice H), de forma que no momento do registro do aluno na base de dados o cadastro já fosse completamente preenchido.

Os dados coletados foram interpretados segundo a ótica da Teoria da Resposta ao Item.

4.3 PRÉ-TESTE

No pré-teste nosso objetivo prioritário foi identificar itens problemáticos, eventuais ajustes no programa de coleta de dados ou na forma de sua aplicação e a obtenção de uma amostra de dados que pudesse ser posteriormente confrontada com outras amostras.

4.3.1 Participantes

Participaram do pré-teste 73 crianças de uma escola particular para classe média (notadamente classes C e B) do município de Jaboatão dos Guararapes, em Pernambuco, sendo 16 crianças matriculadas na Educação Infantil (pré-escolar) e 57 crianças no Ensino Fundamental I (1º ciclo do ensino básico). Elas foram distribuídas por idade e sexo no quadro a seguir.

		Classificação por Idade						
Classificação por sexo	Total de Alunos	**4 anos**	**5 anos**	**6 anos**	**7 anos**	**8 anos**	**9 anos**	**10 anos**
Masculino	42	4	3	7	6	9	6	7
Feminino	31	2	4	3	8	4	7	3
Total:	**73**	**6**	**7**	**10**	**14**	**13**	**13**	**10**

Quadro 3 – Distribuição da população por idade e sexo – pré-teste

Uma aluna desistiu de realizar o teste antes de começar a pontuar e não foi considerada nesse total e dois outros alunos fizeram o teste de forma visivelmente descompromissada e também não foram considerados neste total.

4.3.2 Observações e recomendações oriundas do pré-teste

Como consequência da TRI, as habilidades das pessoas e as dificuldades dos itens são convertidas para a mesma escala (logit), podendo, então, serem comparadas. O mapa de distribuição de pessoas e itens, extraído através do Winsteps e exibido a seguir, nos deu as primeiras impressões sobre o teste:

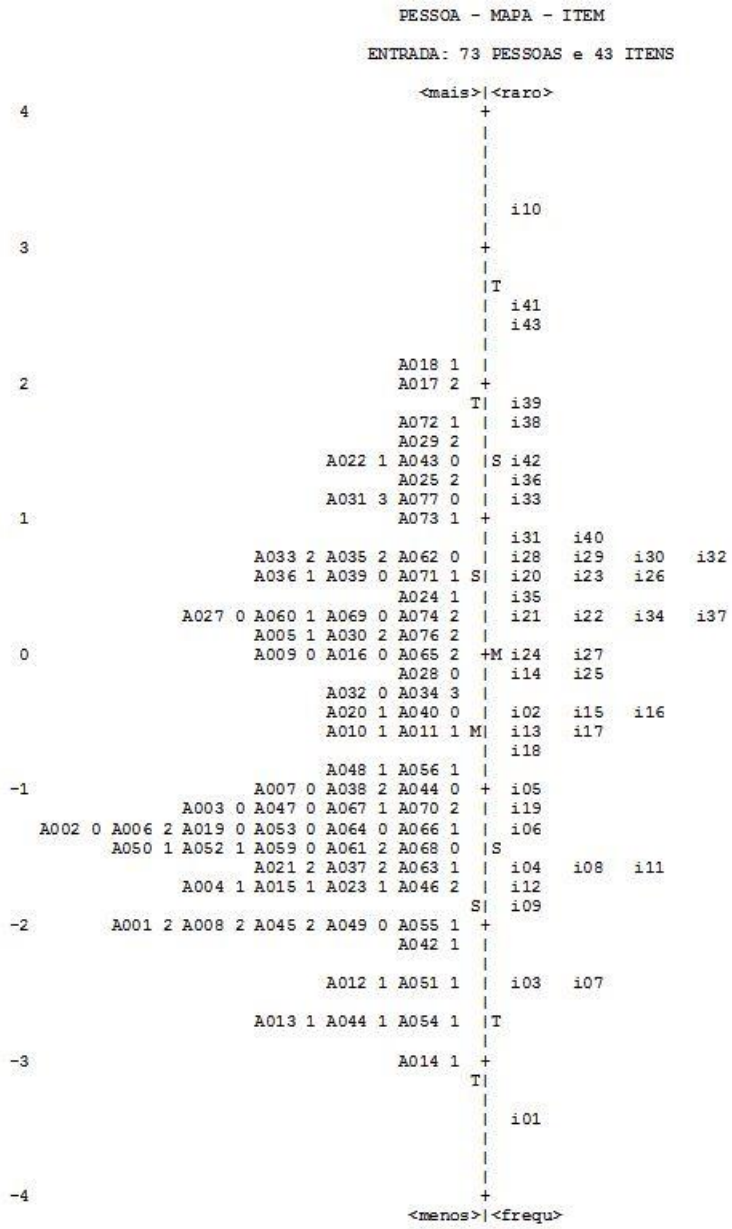


Figura 2 – Mapa comparativo entre pessoas e itens no *continuum* da variável latente – pré-teste

Em uma análise preliminar, ao inspecionar o mapa diversos pensamentos ficam evidentes. Vemos que se trata de um teste equilibrado, mas razoavelmente difícil (também visualizamos isso em outros indicadores, como a Curva Característica do Teste), havendo uma maior concentração de pessoas abaixo de 0 logits e uma concentração maior de itens acima disso. Os itens dão cobertura a, praticamente, toda faixa de habilidade das pessoas, mas o nível de dificuldade do item i10 está estimado muito alto para a amostra e existe uma pequena deficiência de itens no início e no centro do diagrama, o que não chega a comprometer a escala. Enquanto os itens descreveram uma discriminação interessante, variando de -3,39 a +3,22

logits, não tivemos participantes que dessem conta de todo esse espectro. Mesmo um pouco desnivelada, a distribuição é suficiente para exibir a média e dois desvios padrão, tanto para itens quanto para pessoas. Como, supunha-se, que os itens estivessem numerados pelo seu nível de dificuldade, percebemos diversas trocas de posicionamento no *continuum*, notadamente os itens i10, i02 e i12.

O aparente desajuste de alguns itens se confirmou através das estatísticas do item. Certos itens apresentaram valores altos para outfit e infit e correlações point-biserial negativas. O item i10, em especial, pareceu ser um item muito desajustado: teve grau de dificuldade muito acima do esperado e apresentou outfit muito elevado de 9,90 (valores admitidos são entre -2 e 2) e correlação point-biserial negativa de -0,21. Os itens i41, i39 e i01 também apresentaram indicadores de desajustes em graus variados.

ESTATISTICAS DOS ITENS: ORDEM DE DESAJUSTE (FIDEDIGNIDADE = .94)

NUMERO ENTRADA	ESCORE TOTAL	CONT. TOTAL	MEDIDA	ERRO PADRÃO	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBISERIAL CORR.	PTBISERIAL EXP.	PREVISIBIL. OBS%	PREVISIBIL. EXP%	ITEM
10	3	73	3.22	.61	1.22	.6	9.08	3.5	A-.21	.21	95.9	95.9	i10
41	5	73	2.63	.49	1.24	.7	2.48	1.6	B .02	.27	93.2	93.1	i41
39	9	73	1.88	.39	1.26	1.1	2.28	1.9	C .12	.34	84.9	87.8	i39
1	66	73	-3.39	.42	1.00	.1	2.02	1.4	D .14	.23	90.4	90.4	i01
2	34	73	-.47	.27	1.65	4.5	1.95	4.5	E-.03	.46	47.9	73.0	i02

Quadro 4 – Estatísticas dos itens

De posse desses indicadores, procedemos com a averiguação da qualidade dos itens e constatamos que a analogia do item 10 não estava muito clara, optando por remover este item do teste. Quanto aos demais itens não constatamos nenhum desvio de qualidade. A versão revisada do programa deverá considerar as questões observadas e as levantadas pelos alunos, conforme explicitado mais adiante.

Quanto às estatísticas dos sujeitos, já sem o item i10, o Winsteps indicou índices elevados de desajuste ao modelo para os alunos A008, A021, A031 e A044 o que nos levou a retirar esses alunos da amostra.

Após os ajustes na base de dados, voltamos a verificar as estatísticas produzidas pelo Winsteps e constatamos algumas diferenças em favor das capacidades psicométricas da escala, como melhor relação entre os índices de dificuldade e os de habilidades das pessoas e melhora considerável dos valores de outfits, infits e da correlação point-biserial. O valor de 0,91 para o alfa de Cronbach e índice de fidedignidade de 0,89 para pessoas e 0,94 para itens nos levou a

acreditar na consistência da medida. A análise de residuais indicou a possibilidade de existir uma segunda dimensão com força de 3,2 eigenvalues, equivalente à 3 itens, aproximadamente. No entanto, simulando dados através do próprio Winsteps, verificamos os residuais são reduzidos à medida que o banco de dados cresce, de forma que resolvemos não separar os itens em dois testes e prosseguir com a coleta de dados.

Em decorrência da aplicação do pré-teste também surgiram algumas observações e recomendações que contribuiriam para nortear alguns ajustes de procedimentos e do próprio programa de coleta de dados na versão final. Essas foram as nossas principais impressões durante o pré-teste:

a) A grande maioria dos alunos se mostrou interessada em fazer o teste, inclusive com algumas crianças questionando se não poderiam ter o "jogo" em casa.

b) Todos os alunos conseguiram compreender o objetivo do teste e foram capazes de operá-lo, já demonstrando que a interface está condizente com as idades dos alunos.

c) A escolha pela utilização de computador com tela sensível ao toque (touch screen) foi considerada extremamente acertada, haja vista que algumas crianças não eram familiarizadas com o mouse e, fatalmente, não conseguiriam fazer o teste ou o fariam com dificuldades.

d) Constatou-se a necessidade de aumentar a área do botão de avanço de itens para facilitar seu acionamento.

e) Constatou-se um erro no programa no cálculo da idade da criança.

f) Percebeu-se que a maioria das crianças demonstra cansaço entre os itens 15 e 25, dependendo da idade. Começam a perguntar se ainda falta muito, se o coleguinha acertou aquele item, etc.

g) Apenas 3 crianças não chegaram até o item 43, final do teste. Uma se levantou e disse que não queria mais fazer e, no caso das outras duas, o pesquisador preferiu interromper o teste por entender que a criança não estava mais concentrada ou interessada no teste.

h) Embora não fosse comum, mesmo crianças com 5 anos foram capazes de ir até o final do teste de forma compenetrada.

i) Usando a "*reflexão falada*", onde a criança explicita seu pensamento, foi possível identificar que a questão 11 sugere para algumas crianças a figura de uma banheira ao invés de uma cama para o gato, provavelmente por causa da cor azul (cor de água) do colchão

dentro do cesto. Aliás, parece comum as crianças fazerem o raciocínio "gato deitado -> gato em pé ou gato dormindo -> gato acordado" ao invés de "gato dormindo fora do cesto -> gato dormindo dentro do cesto" como seria de se esperar.

j) Algumas crianças parecem perder a referência do raciocínio que vinham fazendo quando o item deixa de representar imagens comuns ao cotidiano das crianças para apresentar figuras geométricas. Muitas querem saber "o que é isso?".

k) As crianças muitas vezes tocam simultaneamente em mais de um ponto da tela. O programa resistiu a esses comandos equivocados, mas a ação da criança não se traduziu em comando nesses casos.

l) Algumas crianças de pouca idade parecem gostar do movimento que ocorre na tela ao selecionar uma opção e ficam, às vezes, fazendo diversas escolhas sem confirmar a escolha, "brincando" com o teste.

m) O conjunto de mesa e cadeira que foi utilizado para o teste não estava adequado para o tamanho das crianças menores e precisou ser alterado. A criança conseguia operar o computador, mas como ela olha de baixo para cima, percebe cores ou tonalidades diferentes daquelas que deveria perceber. No exercício que aparecem rostos (smiles) amarelos, algumas citavam "rostinhos verdes".

n) A prova dispõe de 2 itens de exercício antes do começo do teste. Ampliar esses itens de exercício para 3 ou 4 no futuro poderá ser uma alternativa interessante, pois muitas crianças se mostram inseguras nos primeiros itens de teste quando diferem dos exercícios.

4.4 A APLICAÇÃO DO TESTE FORMAL

4.4.1 Participantes

Responderam ao teste 304 alunos de duas escolas particulares para classe média (notadamente classes C e B) dos municípios de Recife e Jaboatão dos Guararapes, em Pernambuco, sendo 56 pré-escolares e 250 estudantes do Ensino Fundamental I (1º ciclo do ensino básico), compreendendo às idades de 4 a 10 anos, que são aquelas para as quais a ECCOs 4/10 foi destinada. A distribuição por idade e sexo está representada no quadro a seguir.

		Classificação por Idade						
Classificação por sexo	Total de Alunos	**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**

Quadro 5 – Classificação dos participantes por idade e sexo

4.4.2 Análise dos dados

Investigando a existência de outliers, decidimos pela exclusão de 32 participantes, haja vista apresentarem, principalmente, valores altos de outfit, um sinal de inconsistência no padrão das respostas, pois poderiam perturbar a análise do teste como um todo caso fossem mantidos. Dessa forma, passamos a considerar 272 alunos e 42 itens (o item i10 foi excluído durante o pré-teste).

Da mesma forma que no pré-teste, as habilidades das pessoas e as dificuldades dos itens foram convertidas para a mesma escala (logit). O mapa de distribuição de pessoas e itens ficou assim representado:

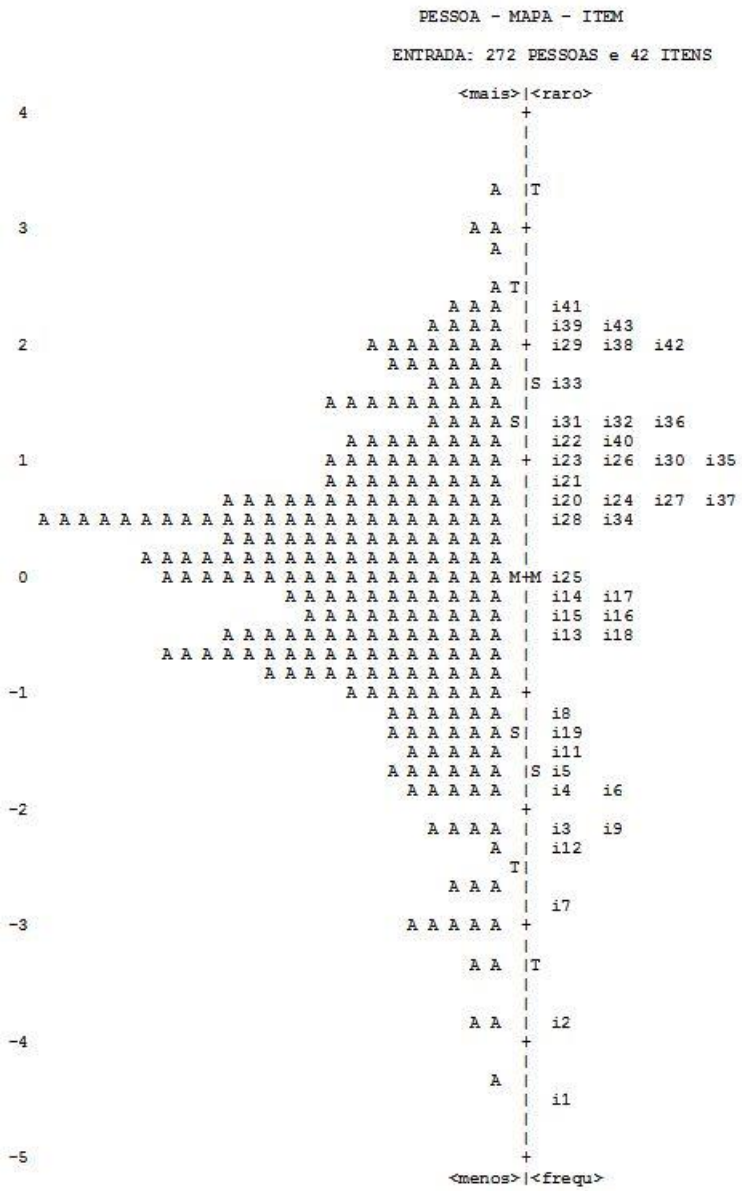


Figura 3 – Mapa comparativo entre pessoas e itens no continuum da variável latente

Visualmente percebemos um teste satisfatório, com boas características psicométricas representadas pelo equilíbrio entre as habilidades das pessoas, exibidas no lado esquerdo do mapa, e as dificuldade dos itens, exibidas no lado direito do mapa. A distribuição é suficiente para exibir a média e dois desvios padrão, tanto para itens quanto para pessoas, embora, idealmente, constata-se a necessidade de mais alguns itens com dificuldade elevada para melhor discriminar a faixa de altas habilidades. De qualquer forma, os itens dão cobertura a, praticamente, toda faixa de habilidade das pessoas. As medidas de dificuldade dos itens foram estimadas entre 2,41 e -4,48 logits, enquanto a estimativa de habilidade das pessoas variou de 3,39 a -4,34 logits. Em relação à ECCOs 4/10 original, pode-se observar 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 1 e 2, 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 os posicionar ao longo da escala é boa.

SUMÁRIO DA MEDIDA DE 272 PESSOAS

	ESCORE TOTAL	CONTAGEM TOTAL	MEDIDA	ERRO MODELO	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MÉDIA	20.5	42.0	.01	.41	1.00	.0	.95	.0
DESV. 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	DESV VERD	1.21	SEPARAÇÃO	2.77	CONFIABILIDADE .88
MODELO RMSE	.42	DESV 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 1 – Sumário da medida de 272 pessoas

SUMÁRIO DA MEDIDA DE 42 ITENS

	ESCORE TOTAL	CONTAGEM TOTAL	MEDIDA	ERRO MODELO	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MÉDIA	132.8	272.0	.00	.17	1.00	-.1	.95	-.2
DESV. 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	DESV VERD	1.67	SEPARAÇÃO	9.30	CONFIABILIDADE .99
MODELO RMSE	.18	DESV 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

Tabela 2 – Sumário da medida de 42 itens

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 nossos dados não têm uma segunda dimensão e podemos prosseguir com a análise.

Tabela de of variância recidual 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% 100.0%	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%	

Tabela 3 – Saída da Análise dos Principais Componentes (em eigenvalue)

Corrobora 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 (pt-biserial=0,12) e i41 (pt-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 ambos 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, conforme pode ser visto no Apêndice C.

A função de informação do teste junto com as curvas características dos itens e do teste (Apêndice E) 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.

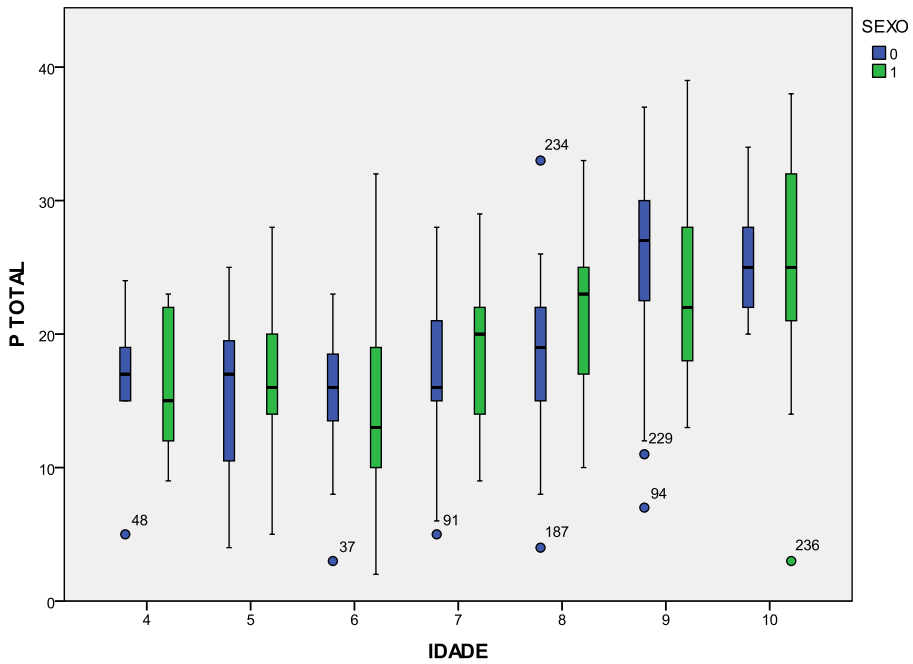


Figura 4 - Boxplot idade x pontuação total

4.5 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 dúvidas, uma grande vantagem de aplicar uma prova automatizada é que a base de dados já está 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 os submeter 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 compromissados e produzam respostas com a provisoriedade 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 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 servem 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.

4.6 CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES

Frente às possibilidades que os testes computadorizados, a TRI e a ECCOs 4/10 proporcionam, podemos dizer que ainda temos um longo caminho pela frente no sentido da modernização dos testes, em especial trabalhando sobre a ECCOs 4/10.

Na análise inicial que fizemos para definir qual das provas da ECCOs 4/10 iríamos informatizar vimos que para algumas provas o recurso computacional poderia ser muito limitante e para outras extremamente libertador. As provas não verbais, como Figuras Incompletas, Composição de Padrões ou Desenhos Absurdos são, notadamente, bastante adequadas para utilização do computador, inclusive com a possibilidade de utilizar itens multimídia e interativos. Já em relação às provas verbais, onde se requer maior interação com

o pesquisador, não se pode dizer o mesmo. Para esses casos parece não ser possível garantir que as condições ideais de realização do teste 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. O atual estado da arte das tecnologias já permitiria 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.

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 poderá melhorar a discriminação da escala em algumas localizações. Esse mesmo conceito pode ser estendido para outras provas.

Outra possibilidade interessante e viável para expandir este estudo, seria a produção de uma versão sob medida (ou adaptativa) do teste de raciocínio, onde os itens que vão sendo apresentados pelo computador ao aluno são determinados pelas suas respostas anteriores em função do nível de habilidade do aluno e do nível de dificuldade dos itens, ambos estabelecidos pela análise dos dados com a TRI.

REFERÊNCIAS

Albuquerque, C. P., & Simões, M. R. (2000). Escala de Inteligência de Wechsler para Crianças – Terceira edição (WISC-III): Validade dos resultados para a intervenção educativa. Psychologica, 25,123-141.

Andrich, D. (1978) Relationships between the Thurstone and Rasch approaches to item scaling. Applied Psychological Measurement, 2, 449-460

Ayala, R. J. (2009). The Theory and practice of item response theory. Nova York, NY : The Guilford Press.

Bond, Trevor G., & Fox, Christine M. (2007). Applying The Rasch Model: Fundamental Measurement in the Human Sciences. New Jersey – EUA: Lawrence Erlbaum Associates, Inc..

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).

Candeias, A...[et.al] (2008) - São Paulo: Casa do Psicólogo. Vários autores

Cattell, R. B. (1966). The scree test for the number of factors. Multivariate Behavioral Research, 1, 629-637.

DeVellis, R. F. (2003). Scale development: theory and applications (second edition). Estados Unidos: Sage Publications, Inc.

Field, A. (2009). Discovering statistics using SPSS (third edition). Dubai: Oriental Press.

Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. Journal of Counseling Psychology, 45(1), 30-45.

Hambleton, R.K., Swaminathan, H. & Rogers, J. (1991). Fundamentals of item response theory. Beverly Hills, CA:SAGE.

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.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Linacre, J. M. (1998). Detecting multidimensionality: which residual data-type works best? Journal of Outcome Measurement, 2, 3, 266-283.

Linacre, J. M. (2002). Rasch Measurement Transactions, 1 Volume 16:2, Autumn 2002 6:2

Lumsden, J. (1978). Tests are perfectly reliable. British Journal of Mathematical and Statistical Psychology, 31, 19-26.

Pasquali, L. (2003). Psicometria : teoria dos testes na psicologia e na educação. Petrópolis, RJ : Vozes.

Pasquali, L., & Primi, R. (2003). Fundamentos da Teoria da Resposta ao Item –TRI. Avaliação Psicológica, p.99-110.

Pasquali, L. (1997). O investimento em Testes Psicológicos. Anais do I Congresso Ibero-Americano de Avaliação Psicológica (pp. 59-60). Porto Alegre: PUCRS.

Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests (Expanded ed.). Chicago: University of Chicago Press. (Trabalho original publicado em 1960.)

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.

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.

Seabra-Santos, M. J. (2000). Avaliação psicológica em idade pré-escolar: O caso da avaliação da inteligência. Psychologica, 25, 143-162.

Stevens, S.S. (1946). On the theory of scales of measurement. Science, 103, 677-680

Snijders-Oomen, N., Tellegen, P. J., & Laros, J. A. (1989). S. O. N.-R 5 ½ -17. Wolters-Noordhoff.

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.

Torrecilla, F. J. M. (2006). Estudios sobre eficacia escolar en Iberoamérica: 15 buenas investigaciones. 1. ed. Bogotá: Convenio Andrés Bello.

Tusing, M. E., & Ford, L. (2004). Examining preschool cognitive abilities using a CHC framework. International Journal of Testing, 4 (2), 91-114.

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

Winsteps Help for Rasch Analysis. (2010). Disponível em http://www.winsteps.com/winman/multidimensionality.htm?zoom_highlightsub=unidimen

Woodcock, R. W. (2002). New looks in the assessment of cognitive ability. Peabody Journal of Education, 77 (2), 6-22.

Wright, B. D. (1967). Sample-free Test Calibration and Person Measurement - ETS Invitational Conference on Testing Problems (Mesa Research Memorandum Number 1 – Mesa Psychometric Laboratory). Disponível em http://www.rasch.org/memo1.htm.

Wright, B. D., & Mead, R. J., & Bell, S. R. (1980). Bical: Calibrating items with the Rasch Model. Research Memorandum Number 23C. Statistical Laboratory Department . of Education The University of Chicago.

Wright, B. D., & Stone, M. H. (1999). Measuremente Essentials (2nd Edition). Wide Range, Inc. Wilmington, Delaware

Wright, B.D. (1996). Local dependency, correlations and principal components. Rasch Meas Trans, 10, 3, 509-511.

Yen, W.M. (1986). The choice of scale for educational measurement: An IRT perspective. Journal of Educational Measure, 23, 299-326.

APÊNDICE A – TELAS DE EXEMPLO DA PROVA FIGURAS INCOMPLETAS DA ECCOs INFORMATIZADA

Figura 5 – Menu Principal

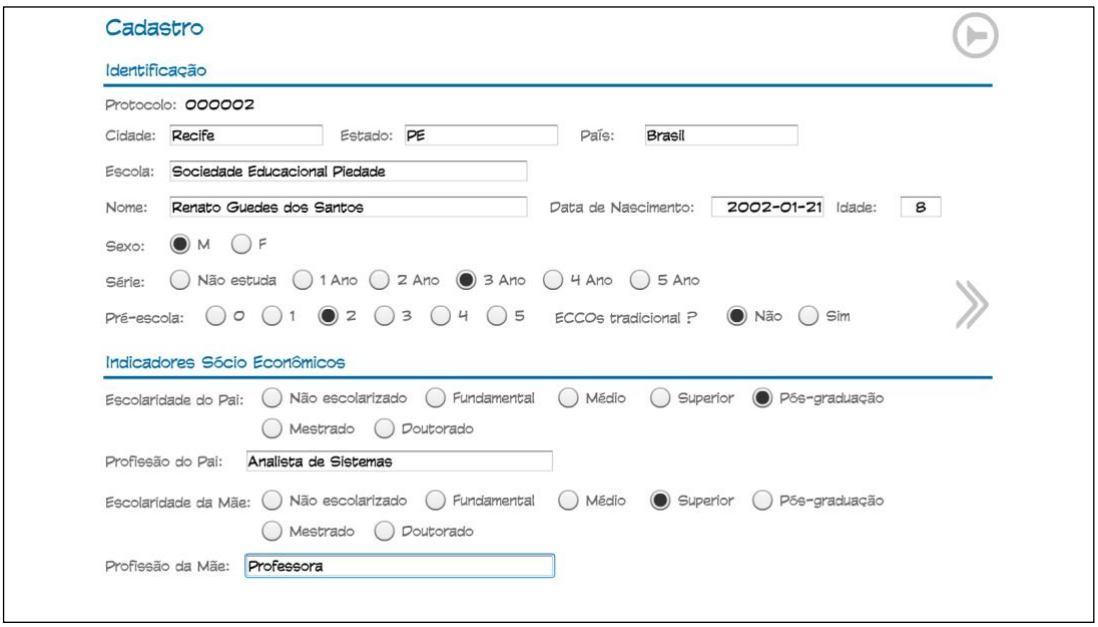


Figura 6 – Cadastro: identificação e indicadores socioeconômicos

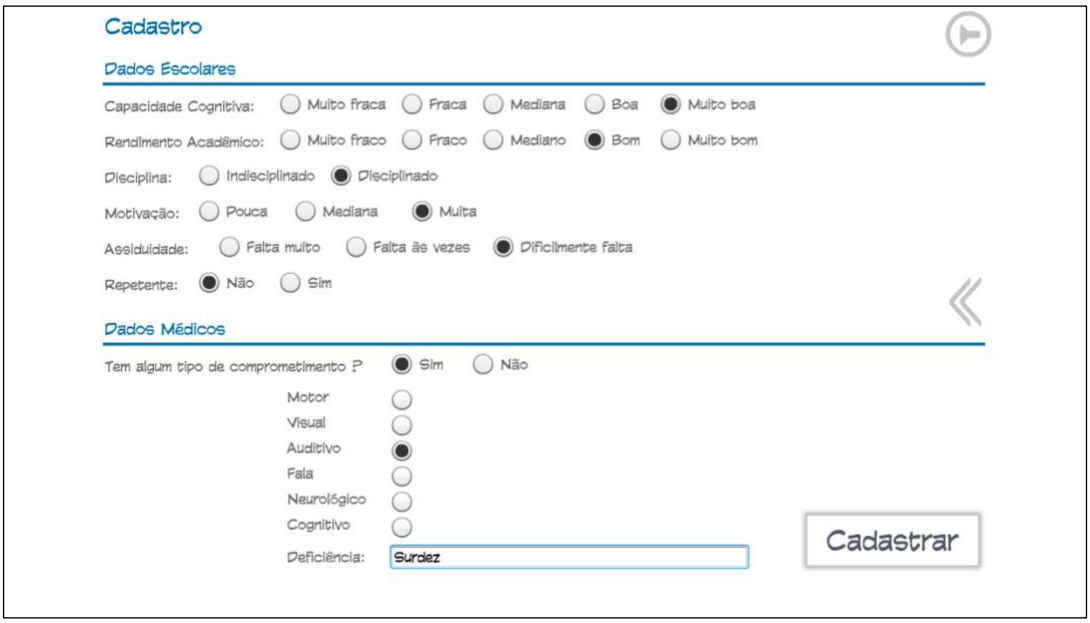


Figura 7 – Cadastro: dados escolares e médicos

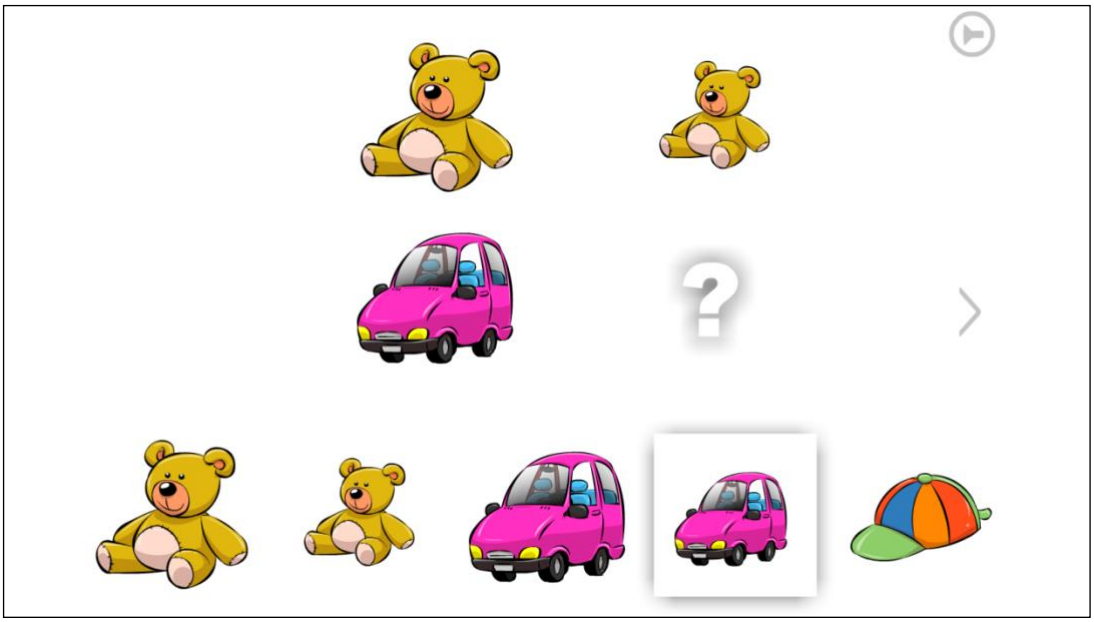


Figura 8 – Exemplo de exercício

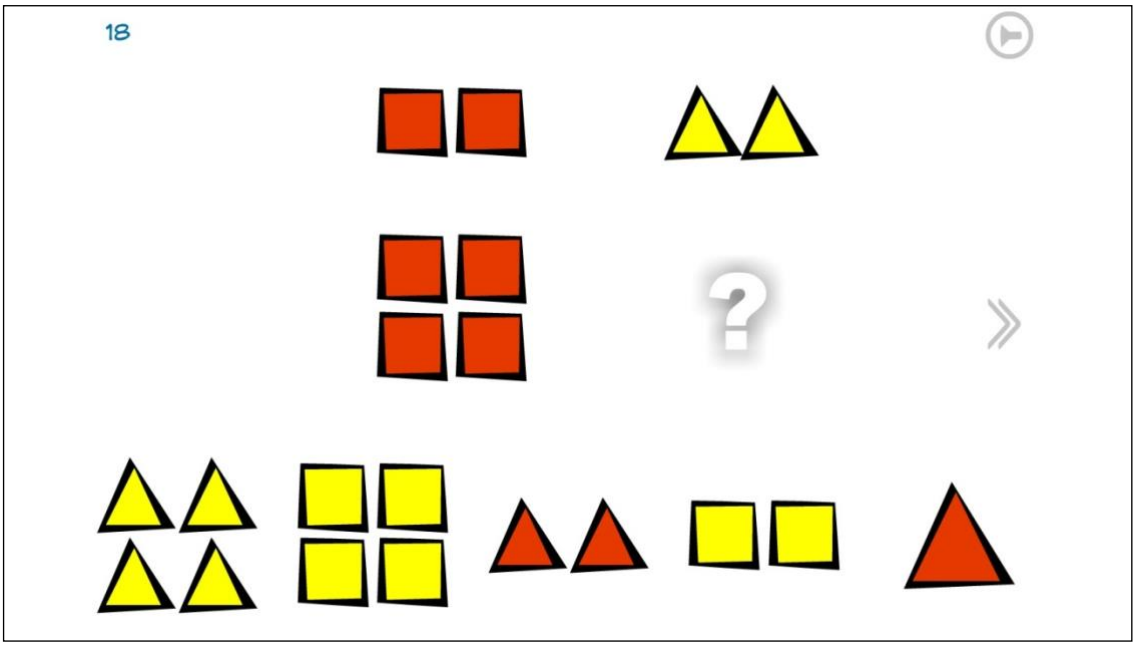


Figura 9 – Exemplo de exercício

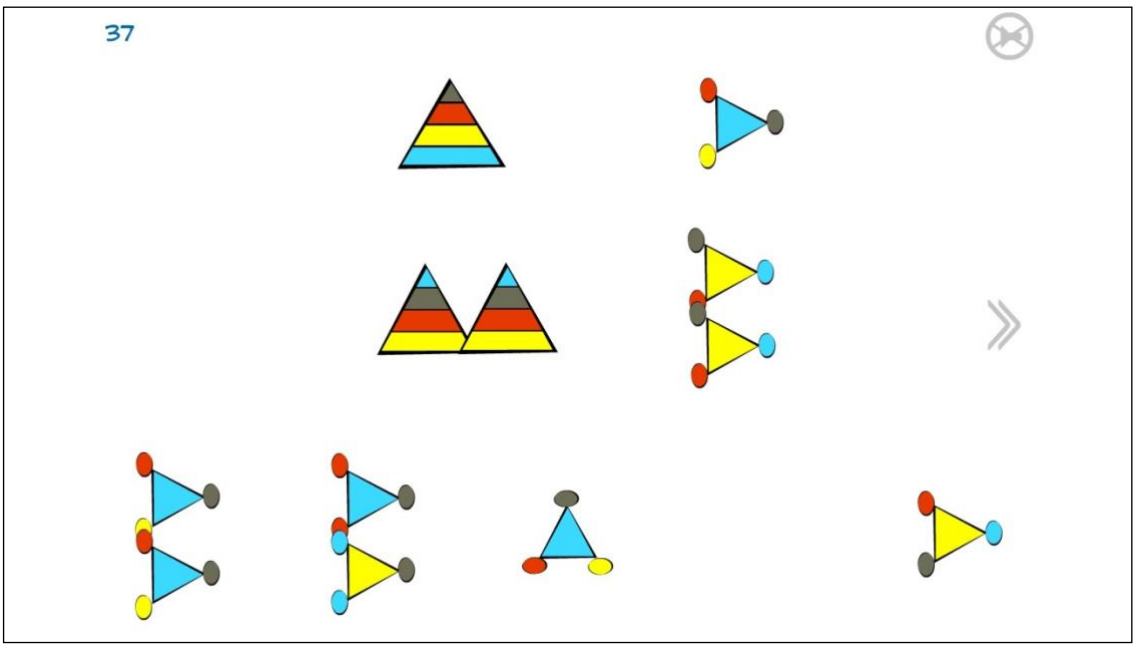


Figura 10 – Exemplo de exercício

APÊNDICE B – ESTRUTURA DE DADOS DA PROVA DE FIGURAS INCOMPLETAS DA ECCOs

Variável	**Tipo**	**Tam**	**Formato**	**Descrição**	**Origem**
protocolo	número	6	999.999	número único de identificação do registro	auto
dt_teste	data	8	dd/mm/aaaa	data de aplicação do teste	auto
hr_teste	hora	4	hh:mm	hora de aplicação do teste	auto
cidade	texto	50		cidade de aplicação do teste	Psic.
estado	texto	30		estado /região de aplicação do teste	Psic.
país	texto	30		país de aplicação do teste	Psic.
escola	texto	50		nome da escola onde o participante estuda	Psic.
nome	texto	50		nome do participante	Psic.
dt_nascimento	data	8	aaaa/mm/dd	data de nascimento do participante	escola
idade	número	3	99,9	idade do participante ao realizar o teste em anos	auto
sexo	número	1	9	sexo do participante	Psic.
escolaridade_pai	número	1	9	escolaridade do pai do participante	família
profissao_pai	texto	30		profissão do pai do participante	família
escolaridade_mae	número	1	9	escolaridade da mãe do participante	família
profissao_mae	texto	30		profissão da mãe do participante	família
ano_escolar	texto	1		série / ano escolar na data de aplicação do teste	escola
comprometimento	número	1	9	o aluno tem algum tipo de comprometimento?	escola
comp_motor	número	1	9	comprometimento motor?	escola
comp_visual	número	1	9	comprometimento visual?	escola
comp_auditivo	número	1	9	comprometimento auditivo?	escola
comp_fala	número	1	9	comprometimento da fala?	escola
comp_neuro	número	1	9	comprometimento neurológico?	escola
comp_cognitivo	número	1	9	comprometimento cognitivo?	escola
deficiencia	texto	25		nome da deficiência do participante, quando houver	escola
cap_cognitiva	número	1	9	capacidade cognitiva informada pelo professor	escola
rendimento	número	1	9	rendimento acadêmico (capacidade de aprendizagem) do participante informada pelo professor	escola
disciplina	número	1	9	o participante apresenta disciplina condizente com o ambiente escolar?	escola
motivacao	número	1	9	indicador da motivação do aluno frente às atividades escolares	escola
assiduidade	número	1	9	indicador da assiduidade do aluno	escola
repetente	número	1	9	o participante já foi reprovado em algum ano?	escola
pre-escola	número	1	9	quantos anos de pré-escola o participante frequentou?	escola /família
acertos	número	2	99	total de acertos no teste	auto
eccos_tradicional	número	1	9	participante já foi submetido à ECCOs tradicional?	Psic.
qtd_item	número	2	99	quantidade de itens respondidos	auto
tempo_total	número	3	999	tempo de resposta ao total de itens, em minutos	auto
Dos itens do teste:					
i01_resp	texto	1	9	indica a resposta que foi assinalada no item 01	auto
i01	número	1	9	indica se a resposta dada ao item 01 está correta (1) ou errada (0)	auto
i01_tempo	número	4	9999	tempo para responder ao item 01, em segundos	auto
[...] [...]				[...]	
i43_tempo				tempo para responder ao item 43, em segundos	auto

APÊNDICE C – ESTATÍSTICAS DOS ITENS DO SPSS

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[b]
	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

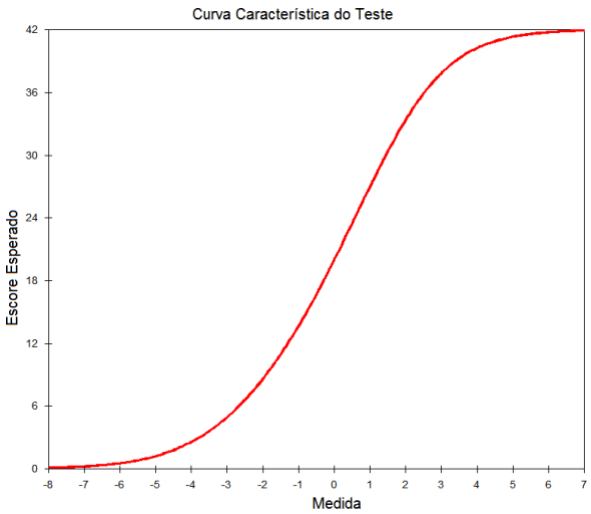
Alfa de Cronbach	Alfa de Cronbach baseado nos itens padronizados	Nr. de Itens
,888	,887	42

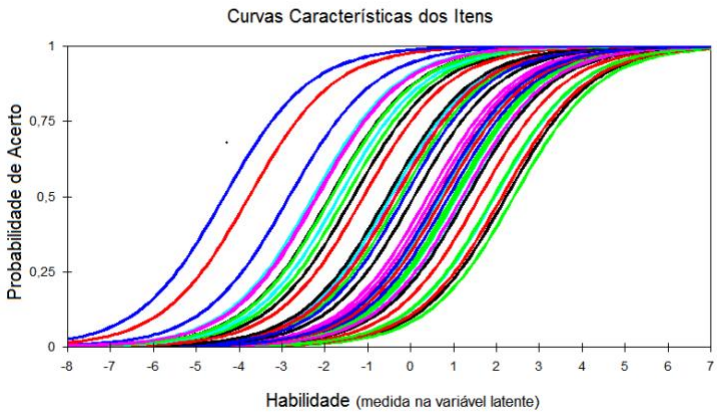
APÊNDICE D – ESTATÍSTICA DOS ITENS DO WINSTEPS

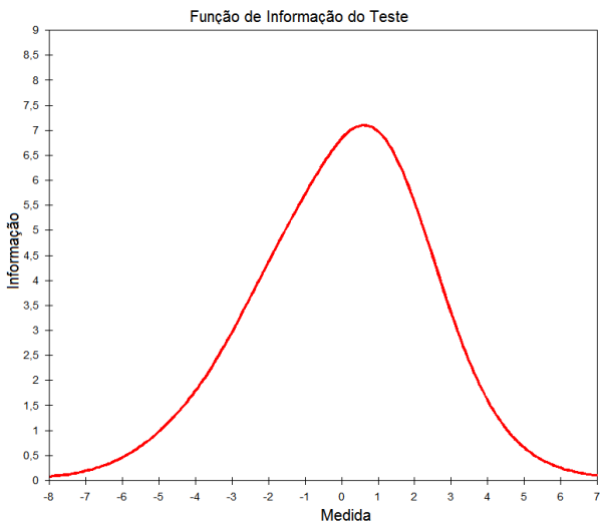
ESTATÍSTICAS DOS ITENS: ORDENADO PELA MEDIDA

NUMERO ENTRADA	ESCORE TOTAL	CONT. TOTAL	MEDIDA	ERRO PADRÃO	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBISERIAL CORR.	PTBISERIAL 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	i41
42	40	272	2.23	.19	.89	-1.0	.90	-.3	.37	.31	87.5	86.0	i43
38	41	272	2.20	.18	1.11	1.0	1.00	.1	.23	.31	84.2	85.6	i39
28	48	272	1.97	.17	.99	-.1	1.08	.4	.32	.33	85.3	83.5	i29
37	48	272	1.97	.17	1.16	1.5	1.09	.4	.22	.33	82.4	83.5	i38
41	48	272	1.97	.17	1.05	.6	.94	-.2	.29	.33	82.4	83.5	i42
32	60	272	1.64	.16	1.08	.9	1.20	1.0	.29	.35	77.6	80.2	i33
31	72	272	1.34	.15	.90	-1.4	.82	-1.0	.46	.37	80.1	77.3	i32
30	73	272	1.32	.15	.88	-1.7	.80	-1.2	.47	.37	82.7	77.1	i31
35	73	272	1.32	.15	.84	-2.2	.75	-1.5	.50	.37	82.7	77.1	i36
21	76	272	1.25	.15	1.23	3.0	1.37	2.0	.19	.38	71.3	76.5	i22
39	78	272	1.20	.15	.77	-3.5	.64	-2.4	.58	.38	80.9	76.0	i40
29	84	272	1.07	.15	1.03	.4	1.11	.7	.35	.39	75.0	74.8	i30
22	85	272	1.05	.15	.95	-.7	.85	-1.0	.44	.39	74.6	74.6	i23
34	88	272	.98	.15	1.15	2.3	1.15	1.0	.28	.39	68.8	74.0	i35
25	90	272	.94	.14	.97	-.4	.89	-.7	.41	.39	74.6	73.7	i26
20	98	272	.78	.14	1.20	3.2	1.24	1.7	.24	.40	65.4	72.4	i21
36	101	272	.72	.14	1.17	2.8	1.29	2.1	.26	.40	65.1	71.9	i37
23	102	272	.70	.14	.91	-1.5	.87	-1.0	.47	.40	77.2	71.8	i24
19	105	272	.64	.14	1.00	.0	1.05	.5	.40	.41	71.0	71.4	i20
26	106	272	.62	.14	.99	-.2	.92	-.7	.42	.41	72.8	71.3	i27
33	110	272	.54	.14	.92	-1.5	.84	-1.4	.48	.41	74.3	70.9	i34
27	116	272	.42	.14	.99	-.2	.96	-.3	.42	.41	71.7	70.4	i28
24	134	272	.08	.14	.90	-2.0	.84	-1.6	.51	.42	75.4	69.7	i25
13	145	272	-.13	.14	1.03	.5	.99	-.1	.40	.42	68.0	70.1	i14
16	151	272	-.24	.14	.92	-1.4	.89	-1.1	.48	.42	76.5	70.6	i17
14	156	272	-.34	.14	.92	-1.5	.87	-1.3	.49	.42	75.4	71.0	i15
15	156	272	-.34	.14	.84	-2.9	.75	-2.7	.55	.42	77.6	71.0	i16
17	163	272	-.48	.14	.92	-1.4	.92	-.8	.48	.42	74.3	71.7	i18
12	167	272	-.56	.14	1.03	.5	1.07	.7	.38	.42	72.1	72.2	i13
8	193	272	-1.12	.15	.87	-1.7	.82	-1.3	.50	.41	80.5	77.3	i8
18	203	272	-1.36	.16	1.42	4.5	1.61	3.3	.12	.40	70.2	79.7	i19
10	208	272	-1.49	.16	1.22	2.3	1.45	2.4	.26	.40	75.0	80.9	i11
5	217	272	-1.74	.17	.81	-2.0	.71	-1.6	.50	.38	87.5	83.2	i5
6	221	272	-1.86	.18	1.06	.6	.92	-.3	.36	.38	83.8	84.2	i6
4	222	272	-1.89	.18	.94	-.5	.76	-1.2	.44	.38	84.9	84.5	i4
3	231	272	-2.19	.19	1.29	2.2	1.35	1.4	.19	.36	83.5	87.0	i3
9	231	272	-2.19	.19	.78	-1.8	.63	-1.6	.49	.36	90.1	87.0	i9
11	233	272	-2.27	.20	.97	-.2	.84	-.5	.38	.36	87.9	87.6	i12
7	246	272	-2.85	.23	.83	-1.0	.61	-1.2	.41	.32	91.9	91.3	i7
2	260	272	-3.85	.32	.93	-.2	.57	-.8	.29	.24	96.0	95.7	i2
1	265	272	-4.48	.40	.87	-.3	.35	-1.2	.27	.19	97.4	97.4	i1
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	

APÊNDICE E – CURVAS DO TESTE







APÊNDICE F – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Projeto de pesquisa: Construção e validação da prova informatizada Figuras Incompletas 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 que poderá ter até 43 questões, 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

APÊNDICE G - PROCEDIMENTOS PARA APLICAÇÃO DO TESTE (1/2)

Material	Prova constante composta por: - dois exercícios de treino - quarenta e três itens constituintes da prova

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 que a criança indique a alternativa de resposta correta. Caso tal se mostre difícil ou a indicação que a criança der for errada, analisar cada alternativa de resposta com ela, explicando a razão pela qual não se adéqua.

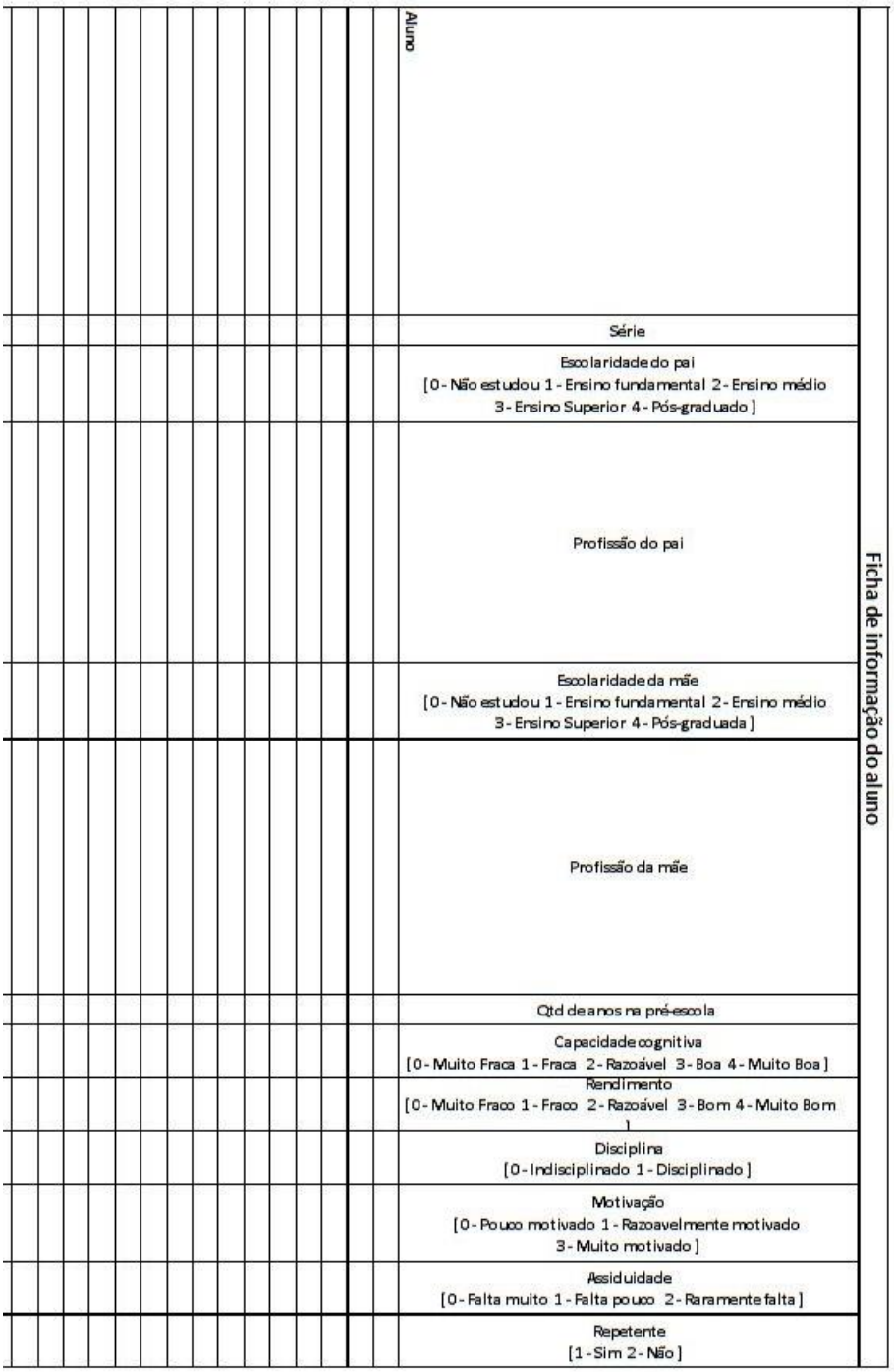
Passar ao segundo exemplo, procurando verificar se a tarefa foi compreendida pela criança. Caso surjam dúvidas, deve proceder-se a uma análise mais pormenorizada tal como atrás foi referido.

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.

Classificação	As respostas serão cotadas com um ou zero pontos consoante a alternativa de resposta selecionada pela criança for certa ou errada, respectivamente.
Leque de cotações	De zero a 42 pontos.

APÊNDICE H – PROCEDIMENTOS PARA APLICAÇÃO DO TESTE (2/2)

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		

APÊNDICE I – FICHA DE INFORMAÇÃO DO ALUNO