Universidade de Brasília
IE - Departamento de Estatística

Estágio Supervisionado 2

Uso da Teoria de Resposta ao Item para Análise de Instrumentos de Avaliação

Leonardo Magno de Carvalho Rebelo

Relatório Final

Orientador: Prof. Luís Gustavo do Amaral Vinha

Brasília

Dezembro de 2011

Agradecimentos

À Renza Carla pelo imensurável amor e carinho, ao Euclides Rebelo por ser uma presença constante em minha vida e ao Eduardo Rebelo pelo grande amigo que é.

À Camila Milani por toda paciência, compreensão e acima de tudo pelo companheirismo de todas as horas.

Aos meus professores por contribuir de maneira tão preciosa na minha formação.

Ao professor Luís Gustavo pela atenciosa e dedicada orientação.

Ao doutor Danilo por ceder o banco de dados para o estudo.

Aos meus amigos por me ajudarem a ser a pessoa que sou.

Trabalhe com o que você tem. Esqueça a perfeição. Em tudo há uma falha, e é por ela que a luz entra.

Leonard Cohen

Resumo

O uso da Teoria de Resposta ao Item ganha importância a cada dia, pois o processo de mensuração de variáveis latentes tem sido cada vez mais requisitado e tem se alastrado para diversos ramos da ciência. Neste trabalho foi realizada uma análise de um instrumento de avaliação Escala de Depressão Geriátrica (EDG) e avaliou-se o desempenho dos itens e do teste para propor modificações no instrumento, caso necessário. Após a discussão sobre o modelo mais adequado a ser utilizado, verificou-se a existência de itens problemáticos que provavelmente perturbavam a suposição de unidimensionalidade. Foram propostas exclusões de 5 itens do teste e o instrumento modificado foi ajustado. O instrumento modificado não foi capaz de atestar a unidimensionalidade do teste, porém foram comprovadas melhoras nos índices e parâmetros dos itens restantes e, portanto, uma melhoria na aplicabilidade prática deste instrumento.

Palavras-chave: Bilog, Testfact, Depressão, TRI, análise fatorial.

Sumário

1 INTRODUÇÃO **7**

1.1 Objetivos 9

2 REFERENCIAL TEÓRICO **10**

2.1 Teoria Clássica dos Testes 10

2.1.1 Propriedade dos Itens 11

2.1.2 Fidedignidade do Instrumento 14

2.2 Teoria de Resposta ao Item 15

2.2.1 Os Modelos Logísticos para Dados Dicotômicos 16

2.2.2 Curva Característica do Item 18

2.2.3 Função de Informação 20

2.2.4 Métodos de Estimação 22

2.2.5 Teste de comparação de modelos 26

2.3 Análise Fatorial 26

3 RESULTADOS **31**

3.1 Análise do instrumento completo 32

3.1.1 Medidas da Teoria Clássica 32

3.1.2 Análise Fatorial 34

3.1.3 Os modelos de TRI 37

3.2 Análises do instrumento modificado 42

3.2.1 Medidas da Teoria Clássica 42

3.2.2 Análise Fatorial 43

3.2.3 O modelo da TRI para o instrumento modificado 45

4 CONSIDERAÇÕES FINAIS **47**

REFERÊNCIAS BIBLIOGRÁFICAS **49**

APÊNDICE I **50**

APÊNDICE II **51**

Capítulo 1

INTRODUÇÃO

Os números são poderosas ferramentas para a compreensão de fenômenos e geração de conhecimento. Quando se pode traduzir uma característica ou um fenômeno em números, consegue-se um grande ganho na capacidade de sua interpretação e compreensão. Neste contexto, o principal objetivo da teoria da medida é descrever fenômenos e manifestações por meio dos números. Entretanto, principalmente quando analisamos o comportamento humano, obtém-se características que não podem ser observadas diretamente e, portanto, requerem variáveis correlatas que viabilizem o seu estudo. Essas características são chamadas traços latentes ou constructos. A abordagem principal dessa teoria, portanto, se dá em torno das manifestações dos traços latentes e na conceitualização de sua natureza.

A Teoria Clássica dos Testes (TCT) visa estudar as variáveis latentes de modo mais simples, focando no valor do escore observado nos indivíduos, sendo este escore obtido, por exemplo, pela aplicação de um instrumento de avaliação. O modelo da teoria clássica relaciona três componentes primordiais: o escore observado, o escore verdadeiro e o erro, que é um componente aleatório. Esta teoria é baseada na distância entre o escore observado que sofre uma variação devida ao erro de

mensuração e o escore verdadeiro, que é o traço latente que se quer mensurar.

Alternativa à TCT, a Teoria de Resposta ao Item (TRI), consiste em um grupo de modelos que estabelece a correspondência entre o traço latente e sua manifestação. Sua essência está na regressão logística ou multinomial, e tem como variável resposta a probabilidade de um indivíduo escolher uma dada categoria de resposta em um item condicionado ao seu nível do traço latente.

A TRI é amplamente usada em diversos campos da ciência. No âmbito da psicologia, por exemplo, pode ser utilizada na psicometria, que tem como objetivo estabelecer uma ponte entre as ciências exatas, principalmente a estatística, e a psicologia, no que tange o estudo de um conjunto de técnicas que possibilitem a mensuração dos traços latentes (Pasquali, 2003). Na área educacional, a habilidade em matemática, por exemplo, pode ser analisada através desta técnica, por meio dos resultados do indivíduo em um instrumento de avaliação específica.

Essa técnica também pode ser utilizada na análise e modificação de instrumentos de avaliação. A TRI, aliada a outras técnicas estatísticas, permite verificar a dimensionalidade do traço latente estudado e a qualidade dos itens que compõem o instrumento. O objetivo deste trabalho é analisar um instrumento de avaliação e propor modificações para torná-lo mais rápido e prático, no que diz respeito ao traço latente em que se propõe a mensurar.

1.1 Objetivos

- **Objetivo Geral**
  - Analisar um instrumento de avaliação psicológica através da TRI e outras ferramentas estatísticas, com o intuito de identificar possíveis problemas no instrumento e propor alterações para medir apenas um traço latente de interesse.
- **Objetivos Específicos**
  - Estudar detalhadamente a TRI e exemplos práticos de aplicação;
  - Estudar outras técnicas de identificação de dimensionalidade;
  - Analisar descritivamente os dados do teste;
  - Aplicar as técnicas estudadas propondo modificações no teste para estabelecer o estudo de um único traço latente;
  - Apresentar os resultados do novo teste proposto.

Capítulo 2

REFERENCIAL TEÓRICO

Definido o traço latente a ser mensurado, deve-se criar um meio que viabilize sua medição. Normalmente, um instrumento é o meio que viabiliza a criação da escala de mensuração do traço latente de interesse, sendo esse instrumento composto por itens. São apresentados a seguir a TCT e a TRI, ferramentas normalmente utilizadas no estudo de instrumentos de mensuração de traços latentes e os resultados de índices propostos pelas técnicas para o presente estudo.

2.1 Teoria Clássica dos Testes

A TCT baseia-se na análise do instrumento como um todo e seu estudo se dá sobre o escore total observado obtido pelos indivíduos em um instrumento.

O modelo utilizado tem como equação regente

$$X_i = \mu_i + \epsilon_i, \tag{2.1}$$

onde X_i é o escore observado e $\mu_i = E(X_i)$ é o escore verdadeiro, que pode ser entendido como a média dos infinitos escores observados obtidos em aplicações independentes de instrumentos equivalentes medindo o mesmo traço latente. Desta forma, a média dos erros para um indivíduo submetido a infinitas medições é zero

$(E(\epsilon_i) = 0)$. O componente ϵ_i é o erro de medição, aleatório, que é a discrepância entre o escore observado e o escore verdadeiro. Por convenção, μ_i é chamado T_i, representando assim o escore verdadeiro. Finalmente, tem-se então a representação do modelo dado por

$$X_i = T_i + \epsilon_i. \tag{2.2}$$

Para construirmos uma escala de mensuração adequada para variáveis latentes, dependemos de variáveis correlacionadas ao traço latente de interesse. Essas variáveis podem ser medidas através dos itens componentes do instrumento e suas caracterizações. Os itens, por sua vez, possuem propriedades que norteiam a criação dessas escalas, permitindo comparações entre indivíduos e itens. Essas propriedades atuam como âncoras que compõem a criação da escala de mensuração e possibilitam a alocação de indivíduos nesta escala.

2.1.1 Propriedade dos Itens

Dificuldade

O parâmetro de dificuldade, no âmbito educacional, é um conceito intuitivo, pois se pensarmos no traço latente "habilidade em matemática", por exemplo, temos que um item dito difícil requer que o respondente tenha habilidade apurada em matemática para acertá-lo.

No entanto, o âmbito comportamental exige uma interpretação diferenciada deste conceito. A dificuldade do item retrata que um indivíduo precisa ter um nível mais elevado do traço latente para endossar o item. Neste contexto, endossar o item significa responder de maneira concordante ao traço latente que está sendo

mensurado.

Um indicador deste parâmetro pode ser dado pela simples proporção de respondentes que endossaram ou acertaram o item. Este índice é chamado **Índice de Dificuldade (ID)** e compõe um dos indicadores usados pela TCT para estudar o item. Assim, a razão entre o número de respondentes que endossaram ou acertaram o item e o número total de respondentes do item formam o ID.

Discriminação

O parâmetro de discriminação representa o poder que o item possui de discriminar respondentes em diferentes níveis de traço latente. Quanto mais discriminativo for o item, mais precisa se torna a diferenciação de respondentes com traços latentes diferentes, seja no âmbito educacional ou comportamental.

A TCT propõe alguns índices capazes de diagnosticar a discriminação, o Índice de Discriminação, ou Índice D, e os coeficientes baseados na correlação item-teste.

- **Índice D**

  O Índice D determina a construção de dois grupos-critério, um chamado superior e outro inferior, primariamente. O banco de dados é ordenado pelos escores observados, e assim os respondentes que possuem os 27% maiores escores compõem o grupo superior, e de forma análoga, os 27% menores escores compõem o grupo inferior. A diferença entre as proporções de endosso do grupo superior e inferior é o Índice D. A literatura aborda intervalos de valores para classificar os itens de acordo com o valores observados por este índice (Algina e Crocker, 1987):

-Se Índice D $\geq$ 0,40 o item é funcionalmente satisfatório;

-Se 0,30 $\leq$ Índice D $\leq$ 0,39, é necessária ou não uma pequena revisão do item;

-Se 0,20 $\leq$ Índice D $\leq$ 0,29, o item precisa ser revisado;

-Se Índice D $\leq$ 0,19 o item deve ser eliminado ou completamente revisado.

- **Correlação item-total** ou **Coeficiente Bisserial**

  O coeficiente bisserial é obtido correlacionando a resposta dada ao item com o escore total observado, sob a suposição de que o traço latente em estudo seja normalmente distribuído. Este coeficiente é dado por

$$\rho_b = \frac{\mu_+ - \mu_x}{\sigma} \times \frac{p}{h(p)}, \tag{2.3}$$

  em que

  - μ_+ é a média dos escores observados no teste para os respondentes que endossaram ou acertaram o item em questão;
  - μ_x é a média global dos escores observados de todos os respondentes do teste;
  - σ é o desvio padrão dos escores observados para todos os respondentes;
  - p é a proporção de endosso ou acerto do item;
  - $h(p)$ é o valor da densidade da distribuição Normal (0,1) acumulada até o ponto p descrito acima.

- **Coeficiente Ponto-Bisserial** Esse coeficiente é uma medida paralela à bisserial e consiste na correlação de Pearson entre o desempenho no item e no

teste. Sua equação é semelhante à bisserial

$$\rho_{pb} = \frac{\mu_+ - \mu_x}{\sigma} \times \frac{\sqrt{p}}{\sqrt{1-p}}, \tag{2.4}$$

onde os termos são os mesmos estabelecidos para o cálculo da bisserial.

2.1.2 Fidedignidade do Instrumento

Como medida de fidedignidade de um instrumento ou sua consistência interna, é proposto o coeficiente α de Cronbach. Desta forma, este coeficiente determina a confiabilidade de um teste ao medir um traço latente, mensurando sua consistência. Este coeficiente é dado por

$$\alpha = \frac{I}{I-1}\left[1 - \frac{\sum_{i=1}^{I} p_i(1-p_i)}{s^2}\right], \tag{2.5}$$

em que I é a quantidade de itens que o teste possui, p_i é a proporção de endosso do item i e s é o desvio padrão dos escores observados. A literatura aborda o valor 0,70 como critério de análise desse coeficiente. Valores de α superiores a 0,70 são considerados satisfatórios e portanto uma alta consistência interna.

A medida que indica a precisão com qual o teste está mensurando o traço latente é o Erro Padrão da Medida (EPM). O EPM é determinado apartir do α e do desvio padrão dos escores obeservados. O EPM é dado por

$$EPM = s\sqrt{1-\alpha}. \tag{2.6}$$

Essas medidas compõe o estudo da Teoria Clássica feitas para este trabalho. Para mais detalhes ver Algina e Crocker (1987).

2.2 Teoria de Resposta ao Item

A TRI é um sistema de modelos que definem uma maneira de estabelecer a correspondência entre o traço latente e sua manifestação. A TRI, porém, não responde questões como porquê o respondente deu uma determinada resposta a um item ou como ele procedeu para escolher essa resposta (Ayala, 2009). Os modelos da TRI baseiam-se no aumento da probabilidade de um indivíduo endossar um item a medida que o nível do traço latente do indivíduo aumenta.

Os modelos de TRI apresentados pela literatura dependem fundamentalmente de alguns fatores (Andrade et al., 2000). O primeiro deles trata da natureza dos itens usados, que podem ser dicotômicos ou dicotomizados e não-dicotômicos. Outro fator considerado é o número de populações envolvidas no estudo, além da quantidade de traços latentes sendo mensurados (dimensionalidade).

Em geral, é de interesse o uso de modelos aplicados apenas a uma população e que sejam unidimensionais. Mas é possível ainda ajustar modelos para dados multidimensionais, e/ou que tenham respostas do tipo politômicas. Como citado anteriormente, para este trabalho serão utilizados modelos unidimensionais com itens dicotômicos e será considerado que os indivíduos são provenientes de uma única população. Além disso, é considerada uma outra suposição importante que procura a melhor forma funcional de um modelo para que este tenha o melhor ajuste possível nos dados. Os modelos da TRI requerem algumas suposições para sua aplicação. Em particular, para os modelos utilizados para a mensuração de apenas um traço latente, as suposições são relacionadas à unidimensionalidade do instrumento e à

independência local, além da escolha do modelo mais adequado.

A independência local prevê que seja satisfeita a condição de que dado um nível de traço latente (θ), todos as respostas dadas aos itens são independentes entre si. Dessa forma, a resposta dada a um item não deve influenciar, nem ter sido influenciada por uma resposta dada a qualquer outro item.

Considerando testes unidimensionais com itens dicotômicos, os modelos da TRI podem ser ajustados com uso de 1, 2 ou 3 parâmetros. Estes modelos são apresentados a seguir.

2.2.1 Os Modelos Logísticos para Dados Dicotômicos

Os parâmetros que podem ser usados na composição dos modelos são discriminação (a), dificuldade (b) e probabilidade de endosso ou acerto do item por um indivíduo com nível latente muito baixo (c). O parâmetros a e b, comumente, têm espaço paramétrico compreendido entre -3 e 3, sendo este, um intervalo empírico. A ocorrência de parâmetros com magnitudes fora deste intervalo são indícios de problemas, principalmente no processo de estimação, e devem ser verificados. Outro fator que pode influenciar na ocorrência de valores incomuns para estes parâmetros é a não observação de respondentes em algumas regiões da escala de latência. Se, por exemplo, um item não foi endossado por nenhum respondente, a estimação do parâmetro b deste item pode ser comprometida e apresentar valores maiores do que o normal. O mesmo pode ocorrer para itens acertados por todos os respondentes.

O parâmetro c, por sua vez, é sempre compreendido entre 0 e 1 por se tratar de uma probabilidade, no caso, de endosso de um item respondido ao acaso.

No **modelo logístico de 1 parâmetro (ML1)** a probabilidade de endosso do item depende apenas da diferença entre o traço latente do indivíduo avaliado e a dificuldade do item. Este modelo é dado por

$$P(U_{i,j}|\theta_j, b_i) = \frac{1}{1 + e^{-a(\theta_j - b_i)}}, \tag{2.7}$$

com os itens $i = 1, 2, \cdots, I$, respondentes $j = 1, 2, \cdots, n$, em que $P(U_{i,j}|\theta_j, b_i)$ é a probabilidade do respondente j endossar ou acertar o item i dado seu nível de traço latente θ_j e a dificuldade do item b_i e a é uma constante de discriminação comum a todos os itens.

Um modelo amplamente utilizado é o modelo de *Rasch*, que tem como particularidade o parâmetro a ser igual a um. Uma vez que o parâmetro de discriminação dos itens é comum, a estimativa do traço latente depende apenas do resultado geral do teste, independentemente do padrão de resposta dado. Isto significa alocar na escala de latência todos os respondentes que obtiveram um determinado resultado, não levando em conta quais os itens foram endossados, e por isso estes respondentes tem a mesma intensidade de traço latente. Sendo assim, os modelos podem ter o ajuste melhorado ao terem inseridos outros parâmetros em sua composição. Por outro lado, o modelo de *Rasch* tem uma interpretação simplificada, dado que todos seus itens possuem a mesma discriminação.

Em muitas situações, a suposição de que os itens têm a mesma discriminação é muito restritiva. Com a inserção do parâmetro de discriminação de cada item, além do parâmetro da dificuldade, obtém-se o **modelo logístico de 2 parâmetros (ML2)**, o qual também será proposto para análise dos dados e verificação de sua

funcionalidade. Esse modelo é dado pela equação

$$P(U_{i,j}|\theta_j, a_i, b_i) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}}, \tag{2.8}$$

com as mesmas especificações do ML1, exceto pela inserção do parâmetro a_i de discriminação de cada item i.

Os modelos anteriores não consideram a possibilidade de um indivíduo com baixíssimo traço latente endossar o item, essa situação é comum no âmbito educacional, onde são utilizados itens com múltipla escolha. Neste caso, é inserido o parâmetro c_i que estima a probabilidade de um respondente acertar um item i respondendo o mesmo aleatoriamente. O **modelo logístico de 3 parâmetros (ML3)** é dado por

$$P(U_{i,j}|\theta_j, a_i, b_i) = c_i + (1 - c_i)\frac{1}{1 + e^{-a_i(\theta_j - b_i)}}. \tag{2.9}$$

Estes modelos são escolhidos de acordo com o escopo do estudo. Neste trabalho, acredita-se que seja desnecessária sua inserção do parâmetro c por não existir resposta ao acaso. Consequentemente, os prováveis modelos a serem utilizados são o ML1 e ML2.

Comumente, os parâmetros a e b

2.2.2 Curva Característica do Item

A análise gráfica dos modelos de TRI pode ser feita através da curva característica do item (CCI) que evidencia a inclinação e localização da curva, mostrando cada um dos parâmetros utilizados de acordo com o modelo ajustado. Esse gráfico demonstra uma relação não-linear entre a probabilidade de endosso do item e o nível latente

do respondente e apresenta formato sigmoidal. Na Figura 2.1 é apresentado um exemplo deste gráfico.

O parâmetro de discriminação é proporcional à derivada da tangente da curva no ponto em que esta muda de concavidade e, por isso, é chamado também de parâmetro de inclinação. Ele representa a discriminação que o item é capaz de fazer entre os respondentes que possuem diferentes intensidades do traço latente, principalmente para valores próximos ao parâmetro de dificuldade. O parâmetro de dificuldade, por sua vez, é também chamado de parâmetro de localização por ser responsável pelo posicionamento da CCI no eixo dos níveis do traço latente (Andrade et al., 2000). O parâmetro c é a assíntota da curva, dado que se o respondente tem grau baixíssimo de latência, esta curva tende ao valor da probabilidade de endosso ao acaso.

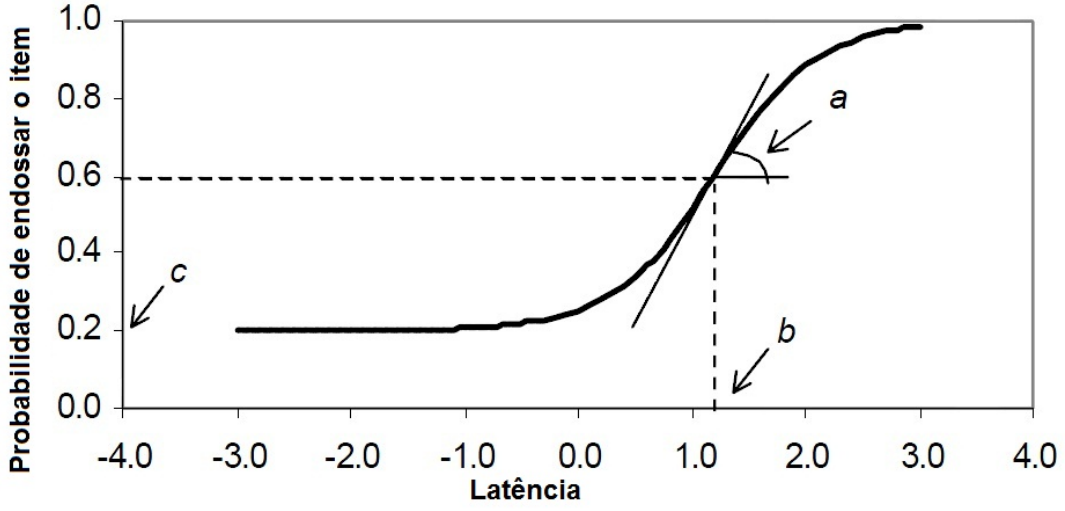

Figura 2.1: Curva Característica do Item

A Figura 2.1 ilustra um item discriminativo, dada sua íngreme inclinação. Além disso o item tem dificuldade igual a 1,2, e o parâmetro c vale 0,20. Dado seu formato em "S", percebe-se a relação não-linear entre o traço latente do indivíduo e

a probabilidade deste endossar o item. O eixo X é a escala de latência onde deve-se destacar a ordem existente entre seus pontos e não necessariamente sua magnitude (Andrade et al., 2000).

A intepretação do parâmetro b, neste contexto, é do grau de latência necessário para que o indivíduo tenha probabilidade $(1+c)/2$ de endossar este item, neste caso, 0,6.

2.2.3 Função de Informação

A Função de Informação (FI) também possibilita a construção de um recurso gráfico, que conjuntamente usado com a CCI, formam um importante mecanismo de análise de itens. A FI quantifica a informação proveniente do item em estudo para estimação traço latente. O item contém maior informação quanto maior for sua discriminação. A função de informação do item é dada por

$$I_i(\theta) = \frac{\left[\frac{d}{d\theta}P_i(\theta)\right]^2}{P_i(\theta)Q_i(\theta)}, \tag{2.10}$$

em que $I_i(\theta)$ é a informação fornecida para um nível θ de traço latente, $P_i(\theta)$ é a probabilidade do indivíduo endossar ou acertar o item, de acordo com nível deste traço e além disso, $Q_i(\theta) = 1 - P_i(\theta)$. Na Figura 2.2 são apresentadas a CCI e a FI de um item. Observa-se que a FI é unimodal e atinge seu pico no ponto em que o traço latente assume um valor igual ao parâmetro de localização.

A função de informação pode ser também usada para avaliação da informação total fornecida pelo teste. Essa função é obtida através da soma das informações de cada item que compõe o teste. A curva é usada para auxiliar na análise da qualidade do instrumento, e além disso, é comumente traçado em conjunto com esta curva,

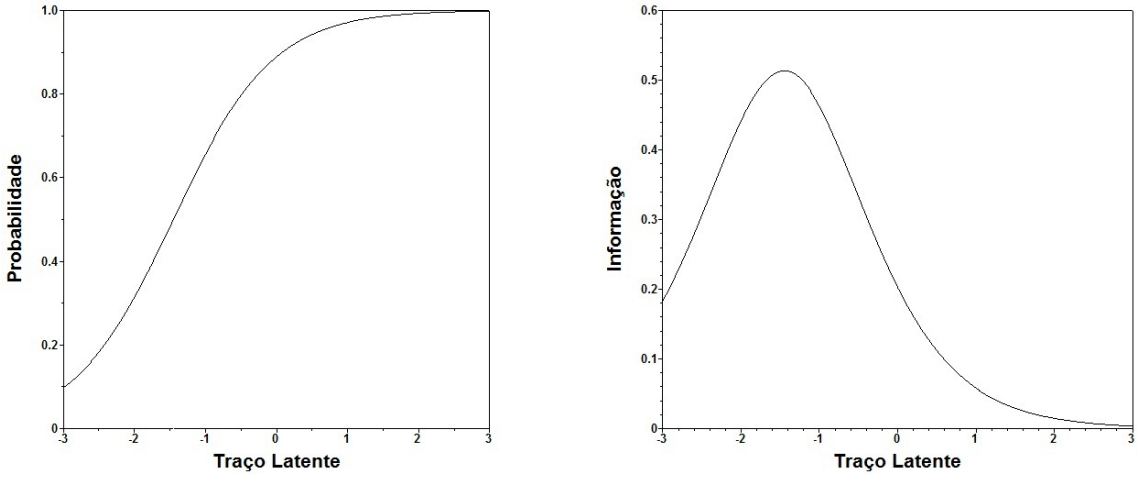

Figura 2.2: Exemplo das curvas CCI e FI

o erro padrão da medida, que por sua vez, é inversamente proporcional à curva de informação. A Figura 2.3 ilustra uma Função de Informação Total (FIT).

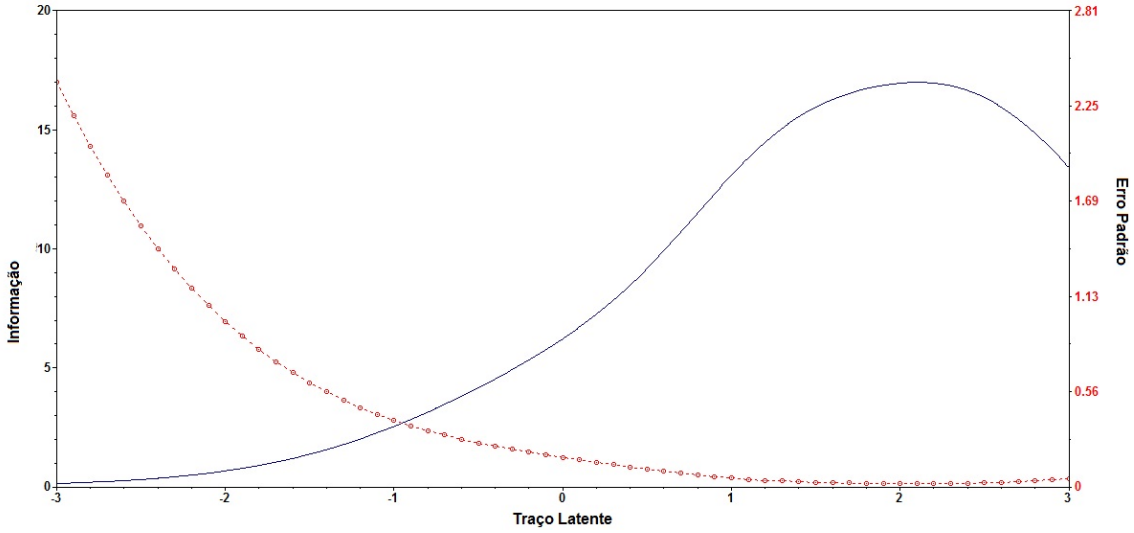

Figura 2.3: Exemplo de FIT e o erro padrão

A FIT nos fornece o comportamento do instrumento em função da escala de latência. O seu ponto máximo corresponde ao nível de latência em que o teste tem minimizado o erro padrão da medição. Desta forma, na extremidade superior do gráfico a informação fornecida é maior e assim, para respondentes com intensidades muito altas do traço latente o instrumento é mais preciso na mensuração da latência.

Por outro lado, na extremidade inferior da curva, observa-se a baixa informação fornecida pelo teste e, consequentemente um alto erro padrão da medida.

Os recursos gráficos apresentados, aliados às medidas de adequação, são utilizados para avaliação de um determinado modelo proposto. Para mais detalhes, Ayala (2009).

2.2.4 Métodos de Estimação

Apresentados os modelos logísticos usados pela TRI, faz-se necessária uma apresentação da metodologia utilizada na estimação dos parâmetros envolvidos no processo. Esses parâmetros podem ser estimados por duas formas principais: estimação conjunta e marginal.

Método de estimação conjunta

A estimação conjunta dos parâmetros consiste em maximizar a função de verossimilhança tanto para os parâmetros dos respondentes quanto dos itens. Assumindo independência condicional e respostas dicotômicas, tem-se a seguinte função que indica a probabilidade de endosso de todo o instrumento, por um indivíduo é dada por

$$P(\mathbf{U}|\theta, \mathbf{a}, \mathbf{b}) = \prod_{j=1}^{I} p_j^{x_j}(1-p_j)^{1-x_j} \tag{2.11}$$

em que $P(\mathbf{x}|\theta, \mathbf{a}, \mathbf{b})$ é a probabilidade do vetor $\mathbf{x}$ de respostas condicionado à latência θ do respondente e a discriminação e dificuldade dos itens $(\mathbf{a}, \mathbf{b})$. Os parâmetros dos itens são calculados de acordo com um modelo estabelecido. Para obter a função conjunta para itens e respondentes multiplica-se a equação (2.11) pelos n respon-

dentes e pelos I itens

$$L = \prod_{i=1}^{N} \prod_{j=1}^{I} p_j(\theta_i)^{x_{ij}} (1 - p_j(\theta_i))^{1-x_{ij}}, \tag{2.12}$$

e aplicando o logaritmo natural (ln) obtém-se

$$lnL = \sum_{i=1}^{N} \sum_{j=1}^{I} [x_{ij} ln(p_j(\theta_i)) + (1 - x_{ij}) ln(1 - p_j(\theta_i))]. \tag{2.13}$$

Assim, os valores de θ, **a** e **b** que maximizam a equação (2.13) são as estimativas dos parâmetros dos respondentes e itens respectivamente. As estimativas são determinadas igualando a derivada dessa função a zero.

Num primeiro momento, os parâmetros dos itens são estimados usando estimativas provisórias das latências dos respondentes (θ's), e então essas estimativas preliminares são tratadas como "conhecidas" para estimar os parâmetros dos itens. Desta forma, usa-se essa estimativa dos itens para refazer a estimação dos θs, que no próximo passo são usadas para refazer a estimação dos parâmetros dos itens. Esse procedimento é feito pelo algoritmo de *Newton-Raphson*. Esse *loop* é repetido até que as sucessivas melhorias nas estimativas se tornem mínimas, se tornando menores do que um critério de parada pré-estabelecido.

Todavia, a determinação simultânea das estimativas dos parâmetros acarretam algumas implicações. Aprimorar a estimativa implica, entre outros quesitos, aumentar a precisão dessa estimação. Sendo assim, um aumento no número de respondentes amostrado seria capaz de aumentar a precisão, mas em contrapartida aumenta também o número de parâmetros a serem estimados.

Outro empecilho desse método é o da eficiência. Como o método usa a estimação das latências na estimação dos parâmetros dos itens, e portanto os parâmetros dos

itens e dos respondentes são atrelados. Sendo assim, se é encontrado algum item que precisa ser revisado, e por isso não se ajusta no modelo, então todo o instrumento deve ser recalibrado. Essa recalibração é necessária para remover os efeitos adversos causados pelos itens "ruins" nas estimativas das latências dos respondentes.

Sendo assim, é proposto outro método de estimação, chamado de estimação marginal de máxima verossimilhança que minimiza os problemas do método conjunto e por isso é o mais utilizado. Método esse que também foi utilizado para estimar os parâmetros dos itens e respondentes deste trabalho. Para mais detalhes Ayala (2009).

Método de estimação marginal

Um outro método amplamente usado é o de máxima verossimilhança marginal, método esse que estima inicialmente apenas os parâmetros dos itens. Com os parâmetros dos itens estimados, uma abordagem Bayesiana é adotada para estimar os parâmetros dos respondentes. Esse método é conhecido como expectância *a posteriori*.

Uma vantagem apresentada por esse método, que deve ser ressaltada, é o aumento da precisão na estimação, que ocorre em alguns instrumentos, já que são separados os parâmetros de respondentes e itens no processo de estimação. O método marginal é proposto assumindo que os respondentes são aleatoriamente amostrados da população, caracterizando assim um "efeito aleatório". Por outro lado, os itens são considerados fixos. Nesse método, o fator aleatório é o mecanismo utilizado para introduzir a informação sobre a população no processo de estimação dos itens, sem

ter a estimativa dos parâmetros dos respondentes (Ayala, 2009). O método tem a equação de probabilidade de um vetor de respostas $\mathbf{x}$ dada por

$$P(\mathbf{U}|\theta, \mathbf{a}, \mathbf{b}) = \prod_{j=1}^{I} p_j^{x_j}(1 - p_j)^{1-x_j}, \tag{2.14}$$

em que $P(\mathbf{U}|\theta, \mathbf{a}, \mathbf{b})$ é condicionado a latência θ e à matriz de parâmetros dos itens. Matematicamente, a aplicação da idéia de aleatoriedade da amostragem dos respondentes de uma população e portanto, é dada a integral que em relação à distribuição da população

$$P(\mathbf{x}) = \int_{-\infty}^{\infty} p(\mathbf{U}|\theta, \mathbf{a}, \mathbf{b}) g(\theta|\upsilon) d\theta, \tag{2.15}$$

em que $P(\mathbf{U}|\theta, \mathbf{a}, \mathbf{b})$ é dado pela equação (2.14) e $g(\theta|\upsilon)$ é a distribuição contínua, como por exemplo a Normal, que representa a população de respondentes. Sendo assim, $p(\mathbf{x})$ é não condicionada ao θ e por isso é a probabilidade marginal de resposta do vetor $\mathbf{x}$, dada por respondentes aleatoriamente amostrados de uma população com distribuição contínua $g(\theta|\upsilon)$ assumida *a priori*. A suposição da distribuição da população impede que o número de parâmetros a serem estimados aumente com o aumento do tamanho da amostra, pois não diferencia os respondentes considerando que todos pertencem a mesma distribuição.

Para proceder o cálculo da probabilidade marginal são usados métodos computacionais de integração númerica. O valor de $P(\mathbf{x})$ obtido é usado na equação do cálculo maximização da função de log-verossimilhança marginal dada por

$$lnL = \sum_{i}^{n} lnP(\mathbf{x}). \tag{2.16}$$

Para mais detalhes ver Andrade et al. (2000).

2.2.5 Teste de comparação de modelos

Um teste estatístico utilizado para decidir qual o modelo mais adequado para os dados é baseado na razão de verossimilhanças. Esse teste compara o ajuste de modelos hierarquicamente relacionados. Essa relação se dá pelo fato do modelo que usa 3 parâmetros, ao ter o parâmetro c constante e zero, se torna o modelo de 2 parâmetros, e analogamente, o modelo de 2 parâmetros se torna o de um parâmetro se tiver o parâmetro a constante, logo, os modelos envolvidos no teste podem ser considerados como completo e restrito. O teste tem estatística dada por

$$\Delta G^2 = -2ln(L_R) - (-2ln(L_C)), \tag{2.17}$$

em que L_C é a máxima verossimilhança para o modelo completo e L_R para o modelo reduzido. Essa estatística tem distribuição χ^2 com graus de liberdade dado pela diferença do número de parâmetros dos modelos envolvidos. Esse teste foi utilizado para auxiliar na escolha do modelo a ser ajustado no presente estudo.

2.3 Análise Fatorial

Uma das suposições iniciais citadas neste trabalho, importante na construção e interpretação de um modelo da TRI, é a unidimensionalidade do instrumento. Um teste unidimensional requer itens altamente correlacionados com o traço latente a ser mensurado, formando um teste fidedigno. A análise fatorial é uma forma de verificar a dimensionalidade de um banco de dados averiguando se essa suposição é satisfeita para que, em um próximo passo, sejam discutidos e ajustados os modelos da TRI formando uma análise completa do instrumento.

A análise fatorial tem como objetivo principal descrever a variabilidade original de um conjunto de variáveis aleatórias X, em termos de um número menor de variáveis aleatórias, chamadas de fatores comuns e que estão relacionados com X através de um modelo linear (Mingoti, 2007). O propósito dessa técnica é agrupar em subconjuntos de novas variáveis (fatores) mutuamente não correlacionadas que sumarizem significativamente a variável original, isto é, os fatores devem explicar significativamente a variância das variáveis originais X. A proporção não explicada é atribuída ao erro aleatório. A formulação matemática do modelo linear é dado por

$$X_i = \beta_{1i}F_1 + \ldots + \beta_{ki}F_k + u_i, \tag{2.18}$$

em que X_i é o valor observado da i-ésima variável, a_{ki} é a carga fatorial da variável observada X_i no fator k e u_i é a proporção não explicada da variância original. Essa componente é chamada de unicidade ou variância não compartilhada.

Para a operacionalização do modelo, algumas suposições são necessárias:

1. $E[F_k] = 0$, para todo k e $var[\mathbf{F}] = \mathbf{I_k}$;
2. $E[u_i] = 0$, para todo i e $var[\mathbf{u}] = D = \begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \sigma_p^2 \end{pmatrix}$;
3. $Cov[F_k; u_i] = 0$, para quaisquer k e i.

Um modelo fatorial que atende às suposições é chamado ortogonal, em que a ortogonalidade se refere ao fato de que os fatores são ortogonais, ou seja, independentes entre si.

O objetivo da análise fatorial é, portanto, encontrar uma decomposição da matriz de correlação das p variáveis ortogonais formada por duas parcelas de variação, sendo

uma advinda dos k fatores e a outra advinda do componente erro. As matrizes que surgem dessa decomposição são extraídas com base na matriz original de correlações. A covariância entre os fatores e os dados originais são chamados de cargas fatoriais.

A matriz de covariâncias Σ dos dados originais pode ser decomposta da seguinte maneira:

$$\mathbf{\Sigma} = \mathbf{BB'} + \mathbf{D}, \tag{2.19}$$

de forma que o objetivo da técnica é encontrar as matrizes B e D que possam representar a matriz σ para um dado valor k de fatores menor que o número de variáveis originais. Assim esses fatores devem representar a composição das variâncias dos dados originais e da covariância entre eles. A variância dos dados originais pode ser expressa por

$$\sigma^2_{Xi} = var[X_i] = \sum_{j=1}^{K} \beta^2_{ij} + u^2_i. \tag{2.20}$$

Por esta relação pode-se verificar que $h_i = \sum_{j=1}^{K} \beta^2_{ij}$ representa a parcela da variância de X_i explicada pelos fatores e u^2_i é a parcela não explicada. Essas porções da variância são conhecidas por comunalidade e especificidade, respectivamente.

Para identificar o número de fatores a serem retidos, a literatura trata de alguns critérios. O primeiro é o critério da análise da proporção da variância relacionada com cada autovalor obtido. Permanecem os fatores cujos autovalores mais explicam a variância total dos dados originais. Esse critério determina que pelo menos 70% da variância dos dados originais sejam explicados por esses fatores. Outro critério amplamente utilizado é o de Kaiser que determina que autovalores maiores que um devem ser retidos, fundamentado na idéia de que se um fator é comum deve conter

ao menos a variância equivalente a uma variável (Mingoti, 2007). O terceiro critério é baseado na análise do gráfico *Scree Plot*. Esse gráfico é feito com os autovalores dos fatores ordenados em ordem decrescente. O critério consiste na detecção de um "ponto de salto" simbolizado por um importante decréscimo nesses autovalores. O número de autovalores anteriores ao "ponto de salto" é o número de fatores que devem ser retidos. A Figura 2.4 mostra o *Scree Plot*.

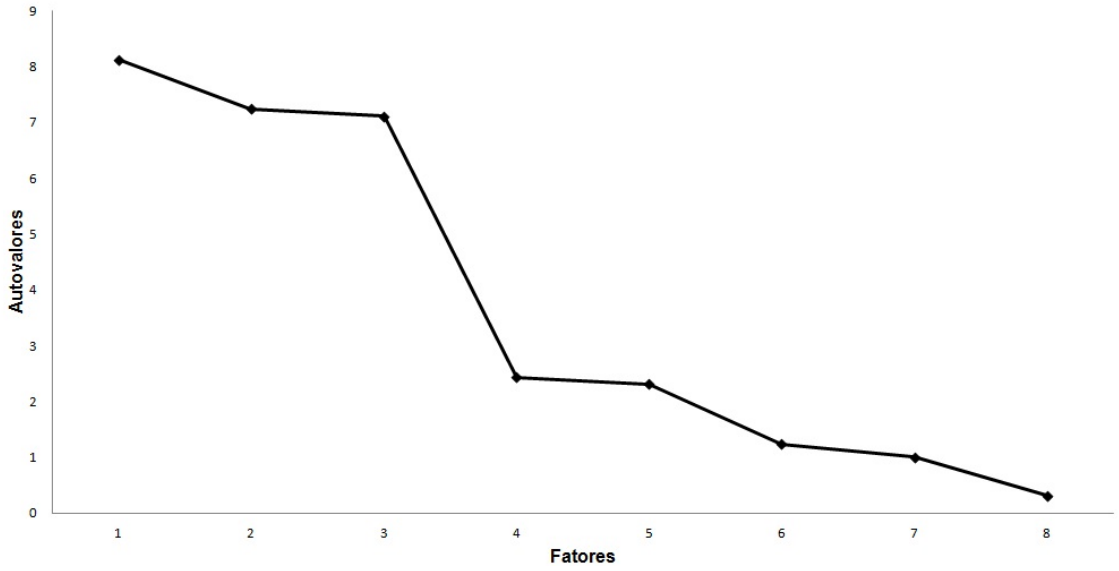

Figura 2.4: Exemplo de Scree Plot

Analisando a Figura 2.4, fica evidente, pelo critério do *Scree Plot*, que 3 fatores devem ser retidos, tendo em vista o importante decréscimo detectado entre o terceiro e quarto fatores. Estipulado um valor k de fatores, as matrizes $\mathbf{B}$ e $\mathbf{D}$ podem ser estimadas pelo método, por exemplo, de máxima verossimilhança. Esse método é indicado apenas quando a suposição de normalidade dos dados originais é satisfeita (Mingoti, 2007).

Para este estudo os itens analisados são dicotômicos. Por isso, faz-se necessário o uso da correlação tetracórica. Essa correlação é a mais apropriada para cons-

trução da matriz de correlação para variáveis dicotômicas. A matriz de correlação tetracórica utilizada na análise fatorial pode ser obtida pelo *software* TESTFACT.

Capítulo 3

RESULTADOS

O instrumento estudado procura mensurar o grau de depressão que afeta indivíduos da terceira idade. O instrumento possui 15 itens e os indivíduos foram classificados em uma escala de depressão de acordo com o grau de latência estimado. As respostas dadas aos itens assumem os valores 1 ou 0, caso sejam concordantes com o traço latente ou não. Foram entrevistados 102 pacientes do Hospital Regional da Asa Norte (HRAN) e de outros hospitais.

Foram utilizados pacotes estatísticos para a análise computacional dos dados, dentre eles o BILOG-MG, que é direcionado para estudo dos dados em TRI e o TESTFACT para a análise fatorial. O programa BILOG-MG é amplamente usado, principalmente na análise de itens dicotômicos. O programa TESTFACT também é mais indicado para uso em itens com essa característica. Além disso, como dito anteriormente, o TESTFACT usa a matriz de correlações tetracóricas em suas estimativas, fato esse que justifica ainda mais seu uso, dado que uma vez mais, essa matriz também é a mais adequada para análise de itens dicotômicos.

A princípio, foram analisados todos os dados disponíveis englobando todos os itens e respondentes extraindo a maior quantidade possível de informações sobre os

dados originais. Como o objetivo do estudo é propor uma modificação do instrumento, são propostas modificações, com a exclusão de alguns itens, posteriormente fomentando a comparação e discussão dos resultados. Para isso, a análise será apresentada de maneira dividida, e todas as análises serão mostradas para ambas as modalidades: instrumento completo e modificado.

3.1 Análise do instrumento completo

3.1.1 Medidas da Teoria Clássica

No primeiro momento, foram realizados os cálculos da TCT para o banco de dados estudado, iniciando o processo de análise. Esta abordagem clássica dos testes tem um caráter preliminar neste trabalho, e mostra indícios do comportamento dos itens, além de proporcionar uma avaliação preliminar do teste como um todo. Desta forma pode-se ter uma idéia da dificuldade do teste e do grau de correlação dos itens com o teste, verificando sua consistência. Foram calculadas as medidas mencionadas no referencial teórico deste trabalho e são apresentadas pela Tabela 3.1. Verifica-se que os itens que possuem maior coeficiente bisserial, ponto bisserial e índice D são os itens 1, 7 e 3, em contrapartida os itens 13, 15 e 9 possuem os menores valores destes coeficientes.

Tabela 3.1: Medidas da Teoria Clássica

Item	Bisserial	P-Bisserial	Índice D	Dificuldade
1	0,877	0,667	0,740	0,696
2	0,602	0,474	0,555	0,607
3	0,823	0,651	0,851	0,588
4	0,619	0,491	0,629	0,568
5	0,634	0,464	0,481	0,754
6	0,587	0,468	0,555	0,470
7	0,948	0,745	0,925	0,617
8	0,661	0,512	0,555	0,343
9	0,528	0,421	0,555	0,470
10	0,567	0,452	0,555	0,470
11	0,802	0,581	0,629	0,764
12	0,743	0,589	0,740	0,431
13	0,499	0,388	0,518	0,647
14	0,724	0,554	0,703	0,313
15	0,505	0,391	0,481	0,343

Esses coeficientes tratam de uma correlação denominada item-teste, têm-se indícios que os itens 1, 2 e 7 são os que mais incorporam o constructo em estudo e consequentemente são os itens que mais contribuem para a fidedignidade do teste. O coeficiente bisserial pode também nos dar indícios sobre o estudo da dimensionalidade do teste em questão. Com a suposição de unidimensionalidade, um teste composto por itens que possuem alto coeficiente bisserial têm um bom indicativo de fidedignidade, tendo em vista que todos os itens que o compõe possuem alta correlação com o teste como um todo. Sendo assim, os itens 13, 15 e 9, em princípio, podem perturbar a suposição inicial de unidimensionalidade. No apêndice A do presente trabalho é apresentado o questionário que foi feito na entrevista dos pacientes.

Baseado no intervalo anteriormente citado, todos os itens estão acima de 0,40 e portanto indicam bons níveis de discriminação pela TCT, verificando os Índices D

dos itens. Observa-se também que os itens 8, 14 e 15 são os mais "difíceis" pela ótica da TCT, e em contrapartida, os itens 5 e 11 seriam aqueles em que o respondente tem a capacidade de endossar mesmo que o grau de depressão seja baixo.

O coeficiente α para o modelo completo é de 0,8138. É um valor satisfatório, o que indica alta consistência interna do teste.

3.1.2 Análise Fatorial

A análise fatorial depende dos critérios usados para retenção de fatores. Além dos critérios abordados no trabalho, usados nesta análise, o programa TESTFACT possui um critério padrão que retém inicialmente um número $\frac{I}{2}$ de fatores, sendo I o número de itens componentes do instrumento. Sendo assim, serão abordados os valores dos 7 fatores e suas cargas, e então analisá-los de acordo com os critérios descritos anteriormente. A Tabela 3.2 mostra os autovalores, os autovalores gerados pela análise paralela, a proporção da variância original que cada um dos fatores explica e a proporção acumulada. Observa-se que o primeiro fator é predominante sobre os demais pois possui um autovalor muito maior que os outros e por consequência é o fator que mais explica a variabilidade total, 41,4%. A técnica de análise fatorial paralela é uma técnica que se baseia na geração de autovalores aleatórios propiciado a comparação entre os autovalores gerados e os empíricos. Esses autovalores são gerados por um *software* que cria um conjunto de dados aleatórios com o mesmo número de observações e variáveis que os dados originais. A matriz de correlação é calculada a partir desse conjunto gerado e seus autovalores são então calculados.

Além disso, pelo critério da proporção explicada, 4 fatores são retidos dado que

Tabela 3.2: Análise dos fatores

Fator	Autovalor	Paralela	Proporção da variância	Proporção Acumulada
1	6,256	1,717	0,414	0,414
2	1,605	1,538	0,106	0,520
3	1,531	1,411	0,101	0,621
4	1,255	1,308	0,083	0,704
5	0,997	1,210	0,066	0,770
6	0,874	1,121	0,058	0,828
7	0,699	1,038	0,046	0,874

explicam pelo menos 70% da variância das variáveis originais. Concordante com o critério da proporção explicada, o critério de Kaiser também determina que 4 fatores sejam retidos pois atingem valores superiores a um. A análise fatorial paralela apontou a existência de 3 fatores. Outro critério abordado é baseado na análise do gráfico *Scree Plot*. Na Figura 3.1, o *Scree Plot* da análise fatorial para o instrumento completo.

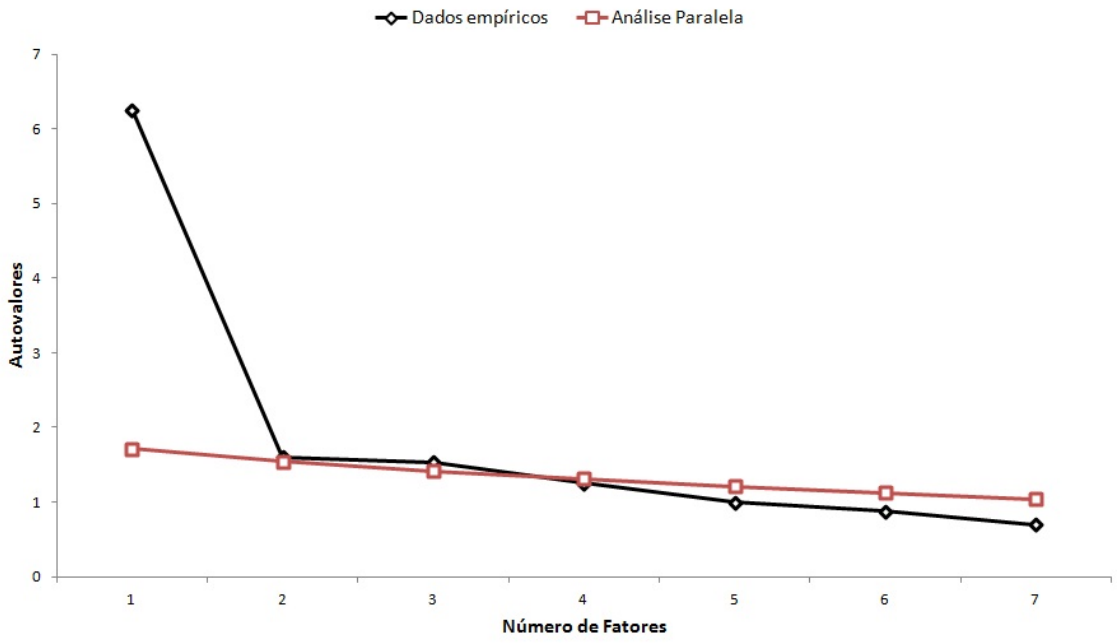

Figura 3.1: Scree Plot

Pelo critério do *Scree Plot* um fator deve ser retido dado a clara redução da magnitude dos autovalores. Além disso, a reta da análise paralela evidencia a existência

de 3 fatores, que é o número de pontos acima da interseção entre as retas. Sendo assim, temos critérios diferentes de análise promovendo diferentes interpretações, já que pelo critério de Kaiser e pelo critério da variância total explicada, 4 fatores devem ser usados no modelo da análise fatorial. Por outro lado, a Figura 3.1 evidencia a presença de apenas um fator, pelo critério gráfico e 3 pela análise paralela. Essa discordância na interpretação dos critérios não possibilita de forma clara a verificação da suposição inicial de unidimensionalidade, por não haver uma clareza na definição do número de fatores presentes no modelo.

A próxima parte da análise é feita a partir dos valores das comunalidades e cargas fatorias obtidas para o fator principal, supondo que esse fator predominante represente depressão. Estes valores representam a correlação do fator com cada um dos itens. A Tabela 3.3 mostra as comunalidades e cargas fatoriais do fator principal para cada item componente do teste. Por essa tabela, observa-se que existem alguns itens que são pouco explicados pelo principal fator, sendo estes provavelmente os responsáveis pela indicação de 4 fatores pelos critérios de Kaiser e de proporção de variabilidade explicada.

Verifica-se também a presença de itens com alta comunalidade e, são aqueles que mais contém correlação com o fator e portanto são os mais capazes de extrair dos respondentes o grau de depressão que aflige cada um deles (itens 1 e 7). Os itens 13 e 15 apresentaram as menores comunalidades. Além disso, os itens 6, 9 e 10 também apresentaram cargas insatisfatórias, todas elas inferiores a 0,2.

A interpretação da análise fatorial executada é sintetizada de forma que, é claro

Tabela 3.3: Comunalidades dos fatores

Item	Comunalidade	Carga Fatorial
1	0,788	0,888
2	0,224	0,473
3	0,572	0,756
4	0,300	0,548
5	0,282	0,531
6	0,195	0,442
7	0,822	0,907
8	0,332	0,576
9	0,144	0,379
10	0,176	0,420
11	0,538	0,733
12	0,423	0,650
13	0,141	0,375
14	0,428	0,654
15	0,136	0,369

que existe um fator que se destaca dos outros, o qual norteia o conteúdo dos itens, fator esse que supostamente trata do grau de depressão. Porém, existem itens no teste que possibilitam que mais fatores sejam indicados pelos critérios, fato que fomenta discussões sobre a composição deste teste. Aliado a esse fato, temos uma baixa proporção sobre a variância total sendo explicada por esse único fator em destaque. Sendo assim, tem-se fortes indícios de que alguns itens devem ser revisados ou reformulados, para que atrelado aos itens com alta comunalidade já presentes no instrumento, componham um teste mais preciso e que seja unidimensional.

3.1.3 Os modelos de TRI

O teste da razão de verossimilhança foi usado para determinar o modelo mais adequado a ser utilizado. Ao descartarmos o ML3, foram comparadas as verossimilhanças para o ML2 (modelo completo) e o ML1 (modelo restrito). Sendo assim:

$$\Delta G^2 = 3547,76 - 3465,60 = 82,16. \tag{3.1}$$

Tabela 3.4: Parâmetros dos itens e os erros padrão da estimativa

Item	a_i	ep(a_i)	b_i	ep(b_i)
1	2,490	0,512	0,644	0,123
2	0,994	0,229	-0,531	0,240
3	1,917	0,374	-0,297	0,136
4	1,154	0,251	-0,302	0,200
5	1,167	0,279	1,219	0,286
6	0,928	0,217	0,153	0,234
7	2,786	0,555	0,372	0,106
8	1,193	0,266	0,703	0,219
9	0,813	0,200	0,168	0,263
10	0,871	0,209	0,160	0,248
11	1,711	0,379	1,033	0,194
12	1,421	0,293	0,275	0,168
13	0,801	0,203	0,865	0,323
14	1,436	0,309	0,759	0,193
15	0,807	0,205	0,921	0,329

Como são 15 itens, o número de graus de liberdade é 14, já que para o ML1 são estimados 15 parâmetros b_i somados ao parâmetro a comum a todos os itens, e para o ML2 são estimados os a_i e b_i para os itens, totalizando 30. O p-valor obtido é menor que 0,001 logo verifica-se uma significante melhora no ajustamento do modelo, se comparado com o ML1. Sendo assim, o ML2 é o modelo mais adequado para proceder a análise da TRI. Ajustado o modelo ML2, a Tabela 3.4 mostra as estimativas dos parâmetros dos itens e os erros padrão dessas estimativas.

Analisando a Tabela 3.4 pode-se verificar que os valores obtidos dos erros padrão são razoáveis e que os parâmetros de discriminação apresentam certa variabilidade (de 0,801 a 2,786), o que confirma que o modelo ML2 é o mais adequado. Observando também os valores obtidos tem-se que o item 7 é o mais discriminativo, seguido do item 1. Além disso, os itens 9, 10, 13 e 15 são os menos discriminativos. Se tratando do parâmetro de dificuldade, os itens 5 e 11 têm maior parâmetro de posição, ou

seja, são itens que requerem maior latência para serem endossados.

Os parâmetros de dificuldade e discriminação de cada item são então analisados para tentarmos identificar os itens que menos contribuem na medição do grau de depressão. Os itens 13 e 15 são os itens com menor valor no parâmetro de discriminação e por isso têm baixa capacidade de discernir respondentes que possuem diferentes intensidades de latência. A Figura 3.2 mostra as CCI's e FII's desses itens e no apêndice B deste trabalho encontram-se as CII's e FII's dos itens.

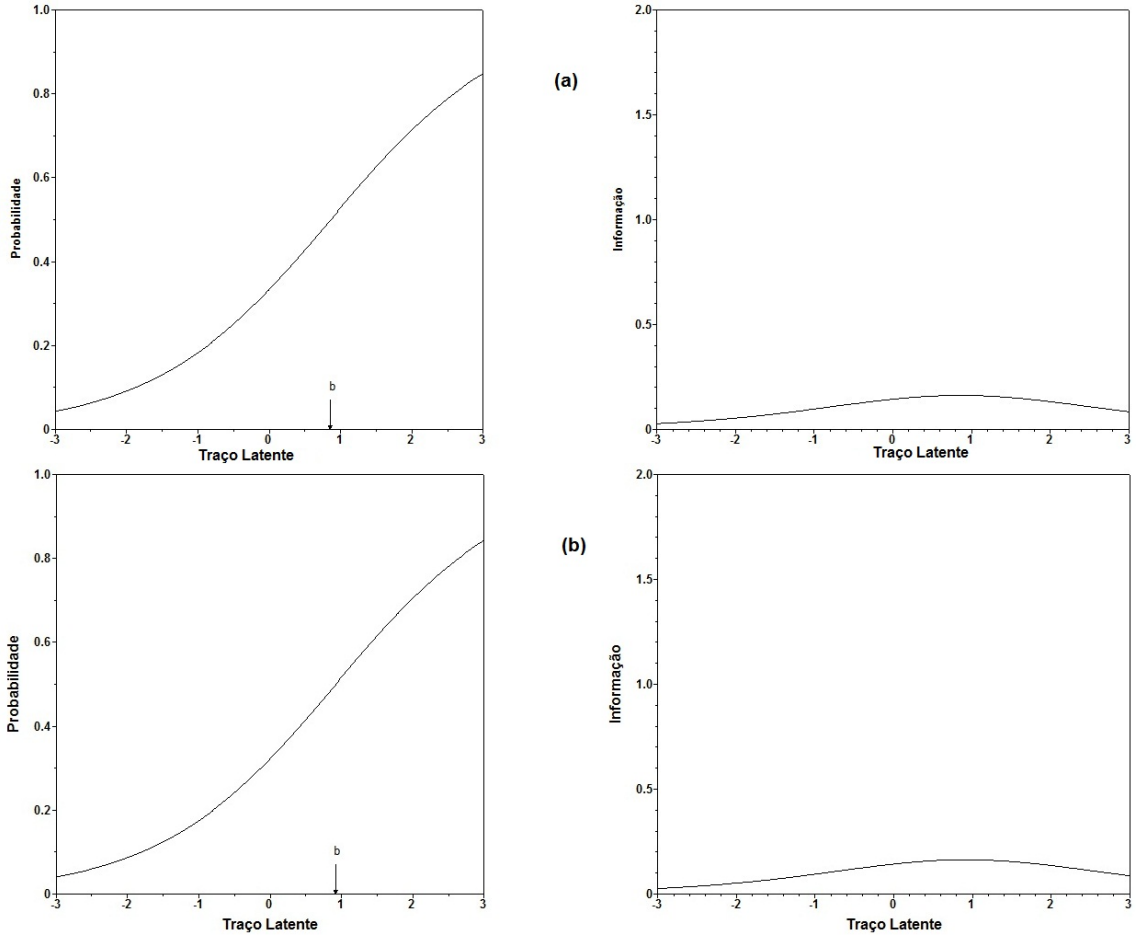

Figura 3.2: (a) CCI e FII do item 13 e (b) CCI e FII do item 15

Com os mais baixos valores de discriminação, esses itens também possuem os piores resultados quanto à informação fornecida, se assemelhando a uma reta. Sendo

assim, tem-se um indício que, principalmente esses dois itens, são os que mais contribuem para existência de outros fatores e baixa proporção em relação à variância total explicada verificada na análise fatorial. Além desses, os itens 6, 9 e 10, como dito anteriormente, também são pouco discriminativos .

Por outro lado, devemos ressaltar os melhores itens e fazer uma comparação gráfica também. Os itens 1 e 7 possuem os maiores valores discriminativos e acentuadas curvas de informação (vide Tabela 3.3), ressaltando a grande capacidade desses itens, em especial, a de diferenciar com precisão os respondentes que são alocados em diferentes pontos da escala de latência (depressão).

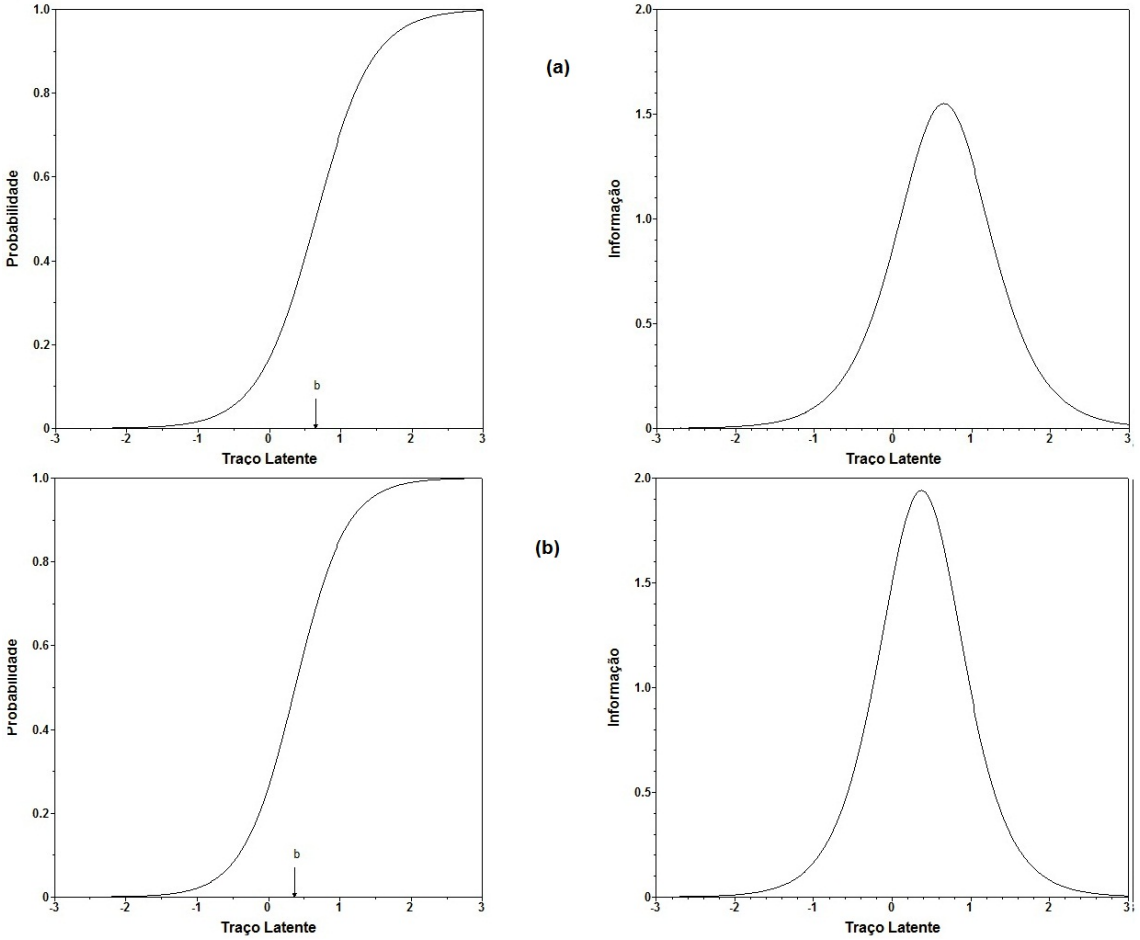

Figura 3.3: (a) CCI e FII do item 1 e (b) CCI e FII do item 7

Tendo em vista os gráficos das CII's e FII's apresentadas, verifica-se claramente a diferença na capacidade de discernimento desses itens. Além disso, as curvas de informação dos itens 1 e 7 atingem um ápice elevado, diferentemente dos itens 13 e 15.

A Figura 3.4 mostra a FIT para o estudo deste instrumento. A curva em preto mostra a função de informação do teste, enquanto a curva em vermelho mostra o comportamento do erro padrão. A FIT desse teste, mostra que em torno da intensidade 0,5 do traço latente depressão o teste é mais informativo, e portanto mais preciso em suas estimativas. Isso significa que o teste tem mais capacidade de diferenciar os respondentes que estão nesta região.

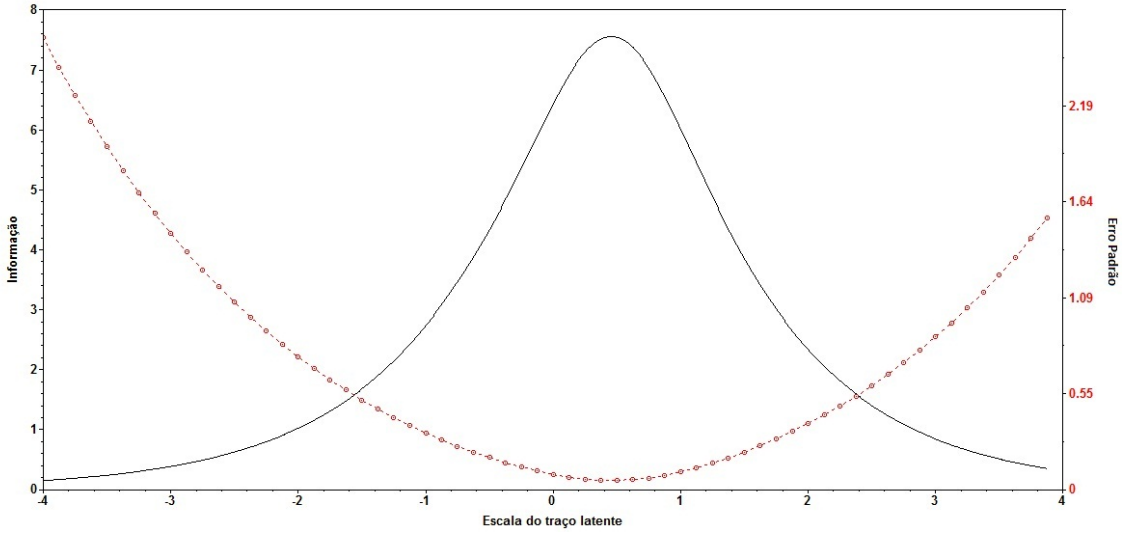

Figura 3.4: FIT do teste

A Figura 3.4 mostra uma FIT com imprecisão de estimação nas extremidades da escala de latência. Isso pode se dever ao fato de não se ter itens suficientes localizados nessas áreas da escala, medindo com imprecisão o grau de pessoas muito depressivas ou de pessoas sem nenhum indício de depressão.

Tabela 3.5: Medidas da TCT

Item	Bisserial	P-Bisserial	Índice D
1	0,944	0,718	0,778
2	0,624	0,499	0,556
3	0,887	0,702	0,814
4	0,698	0,554	0,703
5	0,717	0,524	0,481
7	0,955	0,751	0,889
8	0,722	0,559	0,629
11	0,789	0,572	0,630
12	0,752	0,597	0,741
14	0,788	0,602	0,667

3.2 Análises do instrumento modificado

Portanto, as análises sobre o instrumento completo apontaram alguns itens problemáticos, que entre outros problemas, perturbam a suposição de unidimensionalidade do instrumento. Sendo assim, é proposto um modelo com o instrumento reduzido afim de verificar as implicações provenientes da exclusão dos cinco itens mais problemáticos. Os itens 6, 9, 10, 13 e 15, além de apresentarem baixa correlação com o fator principal são os que possuem menor capacidade de discriminação e portanto foram excluídos. Com a indicação, tanto pela análise fatorial, quanto pela TRI, de exclusão desses itens, é proposto um modelo para o instrumento modificado.

3.2.1 Medidas da Teoria Clássica

A exemplo do que foi feito para o instrumento completo, a Tabela 3.5 mostra algumas medidas da Teoria Clássica para estabelecermos uma comparação. Nessa tabela, não foi mostrado o índice de dificuldade dos itens, pois a retirada de itens não afeta essa medida.

Todos os resultados demonstram uma melhora no teste, mesmo que não sendo

substancial. Analisando os índices, percebe-se uma pequena alteração nos valores das medidas, que por sua vez são ligeiramente maiores para esse modelo do que no modelo para o instrumento completo. Como os coeficientes bisserial e ponto-bisserial são mais altos, temos indícios que os itens excluídos tornam o teste mais consistente. Os itens que continuaram compondo o teste se tornam mais discriminativos, contribuindo também para o aumento da qualidade do teste. Além disso, os mesmos itens 1 e 7 permanecem sendo os mais correlacionados com o teste e dessa vez, os itens que apresentaram menores valores desses coeficientes foram os itens 2 e 4.

O coeficiente α de Cronbach também sofreu um pequeno acréscimo, e passou para $0,8152$. Apesar do fato do valor do α para o instrumento completo ser muito próximo ao obtido pelo instrumento modificado. O pequeno decréscimo do EPM também segue a tendência das outras medidas e aponta uma melhora no teste. O EPM para esse modelo é de $1,470$, enquanto no instrumento completo foi de $1,628$.

3.2.2 Análise Fatorial

Uma análise semelhante à executada para o instrumento completo foi feita para esse segundo modelo também na análise fatorial. As comparações devem ser feitas em torno do número de fatores retidos pelo programa, da análise destes fatores pelos critérios descritos, da análise paralela, da proporção da variância explicada pelos fatores retidos e pelos valores das novas comunalidades e cargas fatoriais obtidas (Tabela 3.6).

Percebe-se um acréscimo de aproximadamente 5% na explicação da variância promovida apenas pelo fator 1, supostamente grau de depressão, enquanto nos 4

Tabela 3.6: Análise dos fatores para o instrumento modificado

Fator	Autovalor	Paralela	Proporção da variância	Proporção acumulada
1	5,380	1,527	0,482	0,482
2	1,448	1,367	0,115	0,597
3	1,026	1,231	0,070	0,667
4	0,731	1,118	0,061	0,728
5	0,534	1,019	0,051	0,779
6	0,380	0,927	0,044	0,823
7	0,221	0,841	0,032	0,855

Tabela 3.7: Comunalidades dos fatores para o instrumento modificado

Item	Comunalidade	Carga Fatorial
1	0,793	0,628
2	0,581	0,337
3	0,929	0,863
4	0,839	0,703
5	0,744	0,553
7	0,849	0,720
8	0,677	0,458
11	0,789	0,622
12	0,905	0,819
14	0,749	0,561

fatores que permanecem retidos pelo critério de Kaiser, constatou-se um aumento de 6%. Esse acréscimo evidencia que os itens retirados perturbavam a suposição de unidimensionalidade, porém o acréscimo ainda é insuficiente para constatá-la de fato, tendo em vista que a variância explicada permanece pequena.

Pela Tabela 3.7, analisando as comunalidades e cargas percebe-se que com a ausência dos itens apontados como problemáticos inicialmente foi verificado um aumento pequeno na proporção da variância explicada dos itens pelo fator principal. Este resultado sustenta a proposição de que um instrumento modificado é mais correlacionado com o traço latente depressão.

Tabela 3.8: Parâmetros dos itens e os erros padrão

Item	a_i	ep(a_i)	b_i	ep(b_i)
1	1,832	0,373	0,622	0,104
2	0,654	0,138	-0,486	0,216
3	1,316	0,246	-0,264	0,123
4	0,783	0,156	-0,267	0,178
5	0,840	0,184	1,086	0,224
7	1,711	0,332	0,383	0,102
8	0,821	0,169	0,634	0,131
11	1,063	0,226	0,728	0,155
12	0,874	0,172	0,658	0,130
14	0,997	0,201	0,709	0,164

3.2.3 O modelo da TRI para o instrumento modificado

Procedendo de maneira análoga, o ML2 também foi ajustado para o instrumento modificado.

Comparando os valores obtidos após feitas as modificações (Tabela 3.8), constata-se uma oscilação no valor dos parâmetros e seus respectivos erros. Em geral, o valor dos parâmetros, principalmente o de discriminação, diminuiu e consequentemente, o erro-padrão da medida também.

Analisando a informação total do teste tem-se a Figura 3.5 que mostra a FIT para o instrumento modificado além da função do erro padrão.

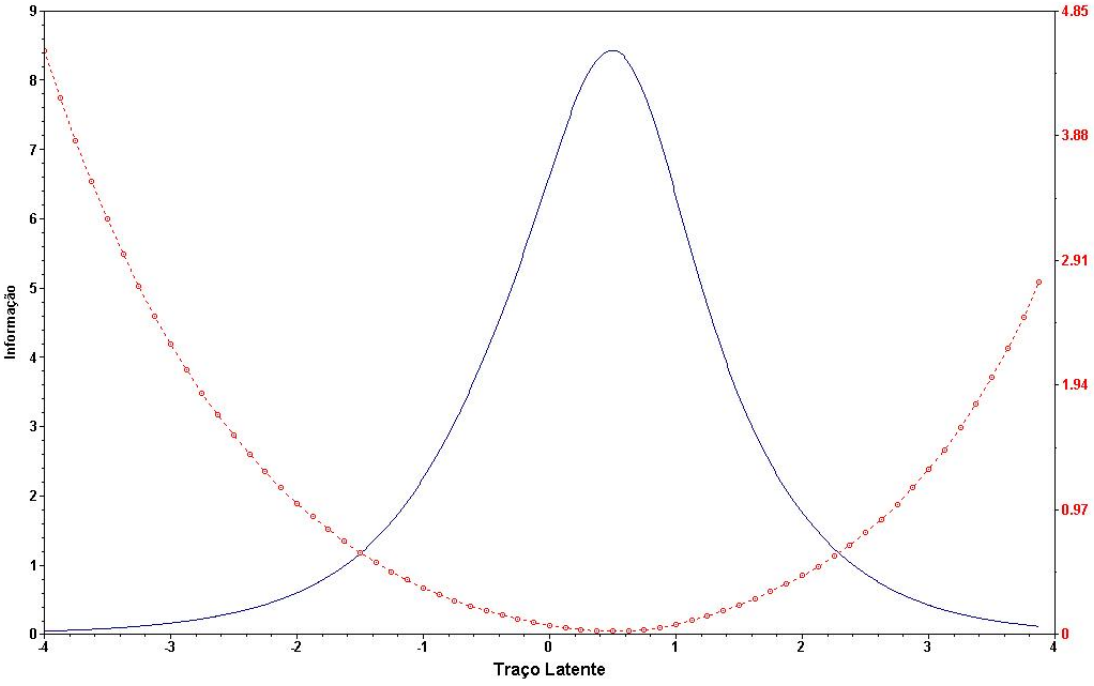

Figura 3.5: FIT do teste modificado

Analisando o gráfico, verifica-se um decréscimo no valor máximo atingido pela função de informação do teste, em relação ao modelo para o instrumento completo e pequena variação no erro padrão do teste. Porém, no geral, o comportamento do teste permanece inalterado obtendo em torno do valor 0,5 de latência a maior capacidade de dicernimento dos respondentes.

Com a retirada dos 5 itens estipulados, foram observadas mudanças radicais no ajuste do modelo da TRI.

Capítulo 4

CONSIDERAÇÕES FINAIS

De acordo com as análises feitas para o instrumento completo, não se teve plenamente satisfeita a suposição de unidimensionalidade e portanto foi proposta uma modificação no instrumento estudado visando melhorá-lo. Essa modificação aconteceu em detrimento dos resultados dessas análises que apontaram itens que perturbavam a suposição de unidimensionalidade, apontaram que esses itens eram também os menos discriminativos além de serem muito pouco explicados pelo fator dominante.

Os resultados do instrumento modificado mostraram uma pequena melhora geral no teste mas ainda insuficiente para atestar sua unidimensionalidade. Houveram aumentos nos coeficientes bisseriais e do α de Cronbach na TCT e o fator dominante explicou uma parte maior da variância dos itens restantes. A TRI apontou uma melhoria nas magnitudes de informação do teste e diminuiu os erros padrão das medidas. Dessa forma, conclui-se que o teste modificado não tem perda significativa de informação e não tem prejuízos na estimação dos parâmetros dos itens. Sendo assim, o teste modificado é menor e consequentemente mais rápido. Este fator é importante, tendo em vista que o instrumento é executado via entrevista com pacientes em idade avançada.

Seria interessante, em trabalhos futuros a utilização de uma amostra maior, visando a obtenção de resultados mais confiáveis. Uma análise profunda da composição do teste. Avaliar os questionamentos feitos e suas implicações psicológicas nos pacientes ajudam a compor um teste mais proveitoso, que seja unidimensional e forneça informações melhores e mais precisas sobre o grau de depressão geriátrica.

Referências Bibliográficas

Algina, J. e Crocker, L. (1987). *Introduction to Classical and Modern Test Theory.* Wadsworth Publishing.

Andrade, D.; Tavares, H. e Valle, R. (2000). *Teoria de Resposta ao Item: Conceitos e Aplicações.* SINAPE.

Andriola, W. B. (2009). Psicometria moderna: Características e tendências. *Est. Avaliação Educacional*, volume 20:319–340.

Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory.* Gilford.

Franchignoni, F. (2010). Using psychometric techniques to improve the balance evaluation systems test: The mini-bestest. *Journal of Rehabilitation Medicine*, pág:323–331.

Johnson, R. A. e Wichern, D. W. (2006). *Applied Multivariate Statistical Analysis.* Prentice Hall.

Mingoti, S. A. (2007). *Análise de Dados Através de Métodos de Estatística Multivariada.* Editora UFMG.

Pasquali, L. (2003). *Psicometria, Teoria dos Testes na Psicologia e na Educação.* Editora Vozes.

APÊNDICE I

1	Você se considera globalmente satisfeito com sua vida?	☒ sim	☐ não
2	Você tem abandonado muitas de suas atividades e interesses?	☒ não	☐ sim
3	Você tem a sensação de que sua vida está vazia?	☒ não	☐ sim
4	Você se aborrece com freqüência?	☐ não	☒ sim
5	Você habitualmente está de bom humor?	☒ sim	☐ não
6	Você tem medo de que algo ruim possa lhe acontecer?	☒ não	☐ sim
7	Você se sente feliz na maior parte do tempo?	☒ sim	☐ não
8	Você se sente freqüentemente sem ajuda, desamparado?	☒ não	☐ sim
9	Você prefere ficar em casa ao invés de sair e fazer alguma coisa?	☐ não	☒ sim
10	Você acha que sua memória é pior do que a das outras pessoas?	☐ não	☒ sim
11	Você acha maravilhoso viver nos dias de hoje?	☒ sim	☐ não
12	Você atualmente se sente sem valor?	☒ não	☐ sim
13	Você se sente cheio de energia?	☒ sim	☐ não
14	Você se julga sem esperança em relação a sua situação atual?	☒ não	☐ sim
15	Você acha que a maioria das pessoas vive melhor do que você?	☒ não	☐ sim

Figura 4.1: Instrumento estudado

APÊNDICE II

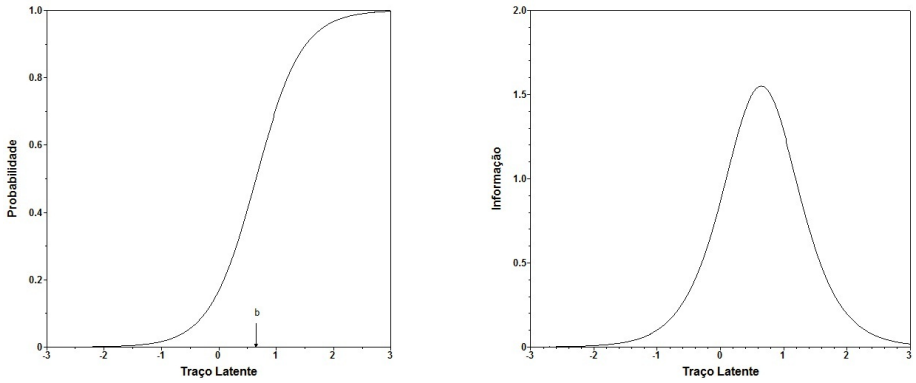

Figura 4.2: CII do item 1

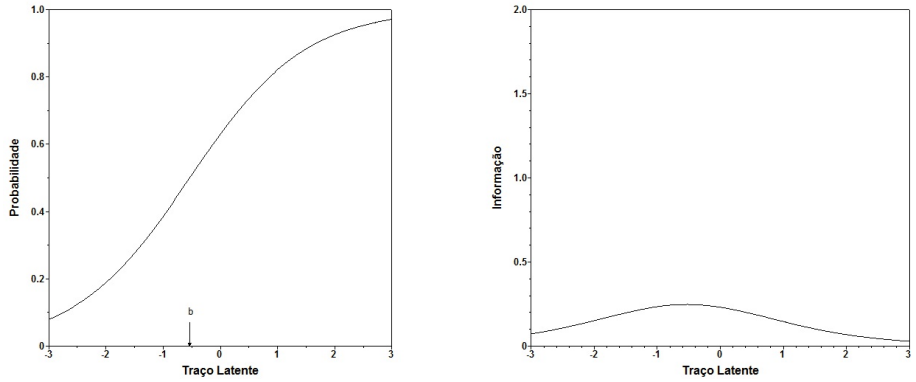

Figura 4.3: CII do item 2

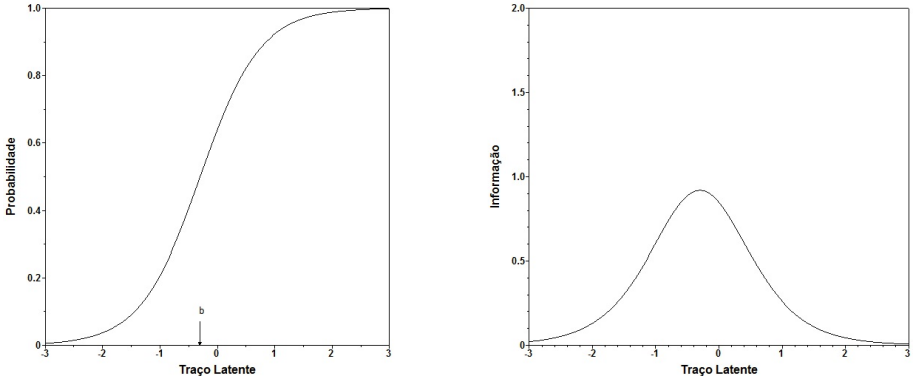

Figura 4.4: CII do item 3

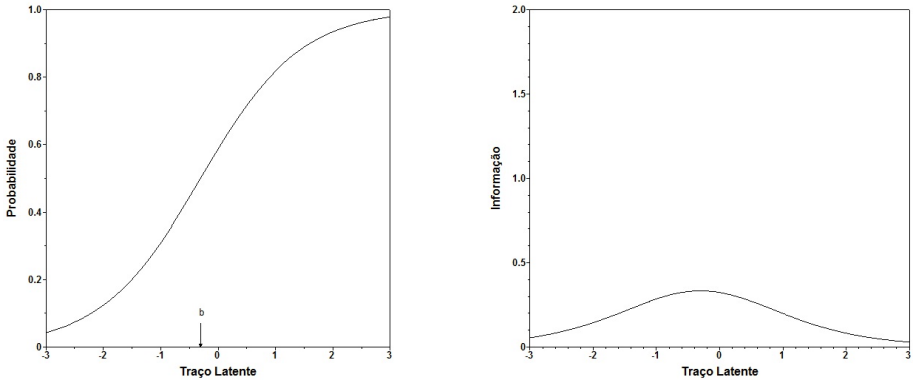

Figura 4.5: CII do item 4

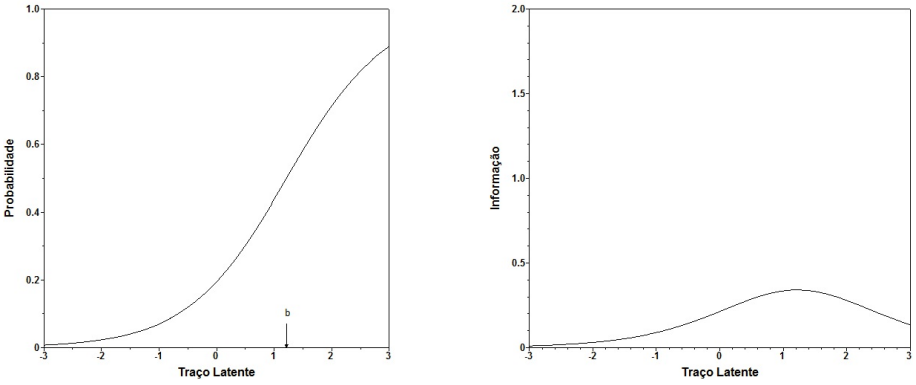

Figura 4.6: CII do item 5

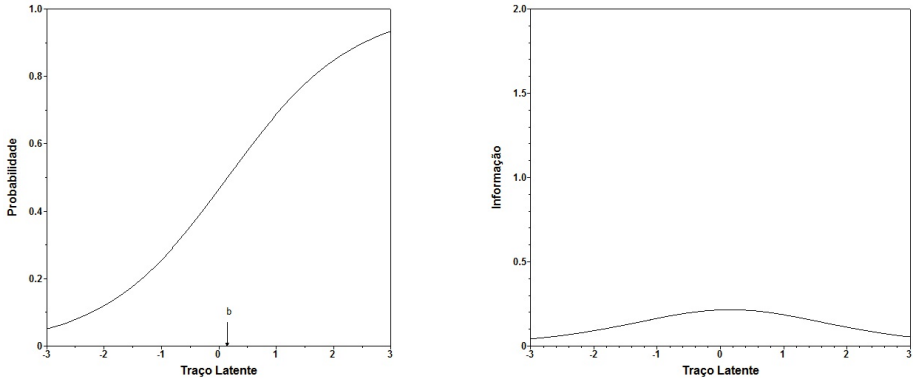

Figura 4.7: CII do item 6

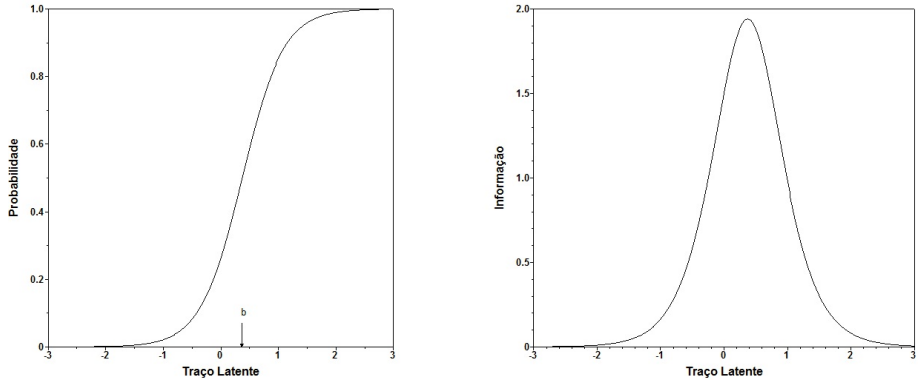

Figura 4.8: CII do item 7

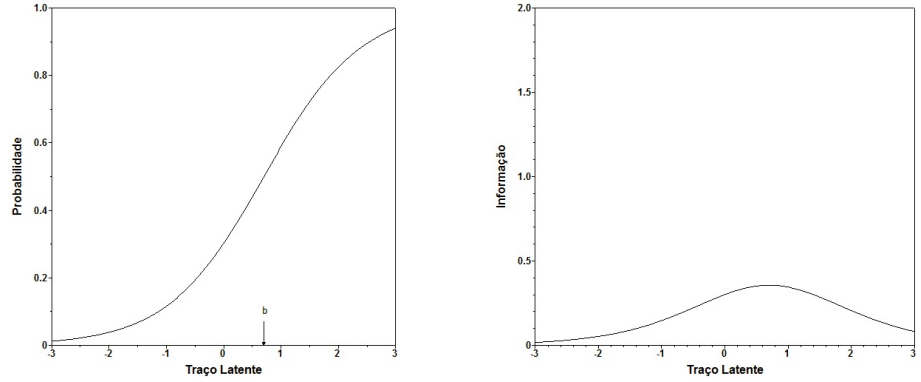

Figura 4.9: CII do item 8

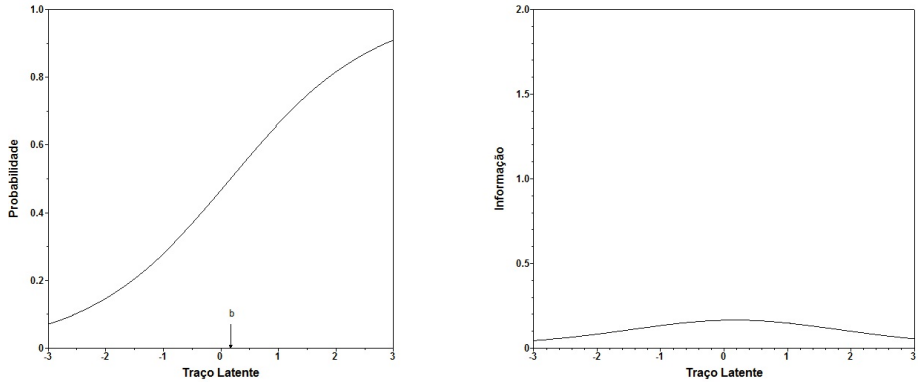

Figura 4.10: CII do item 9

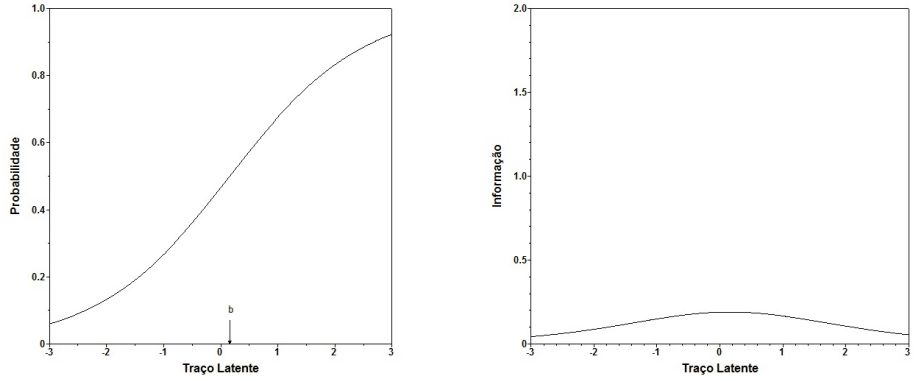

Figura 4.11: CII do item 10

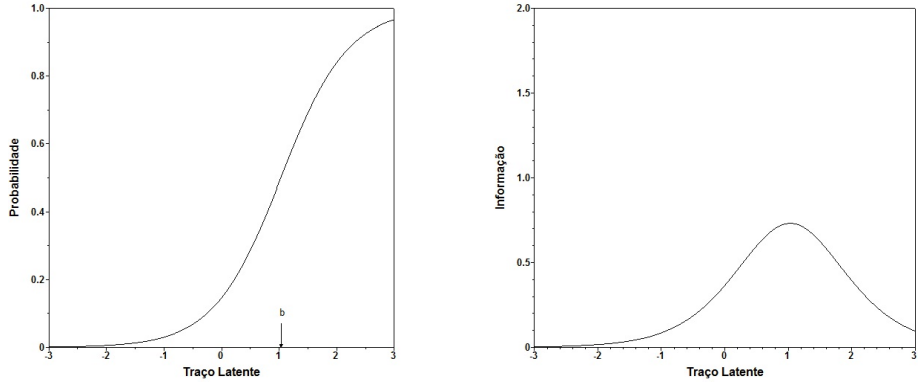

Figura 4.12: CII do item 11

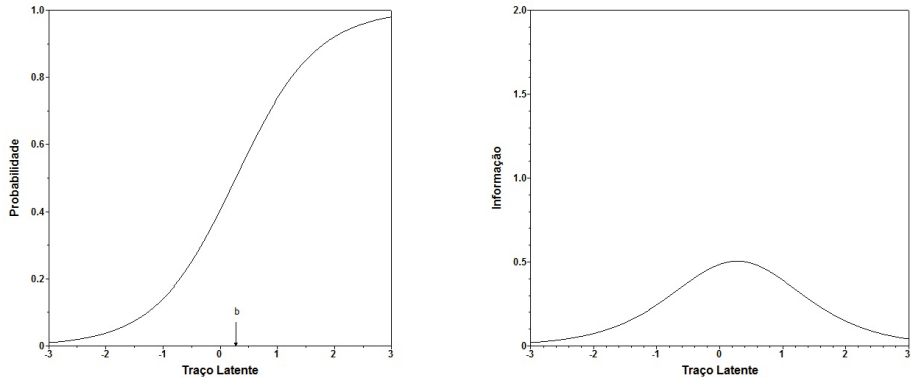

Figura 4.13: CII do item 12

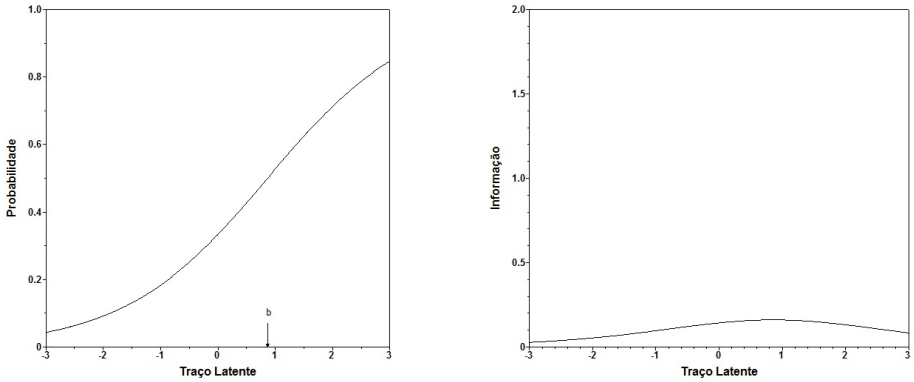

Figura 4.14: CII do item 13

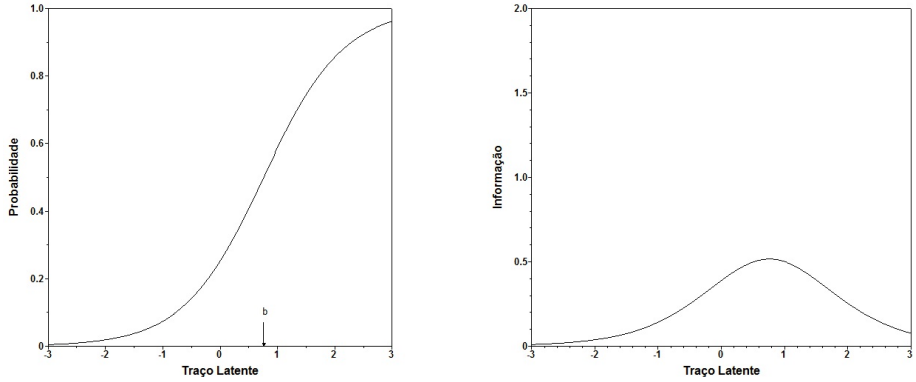

Figura 4.15: CII do item 14

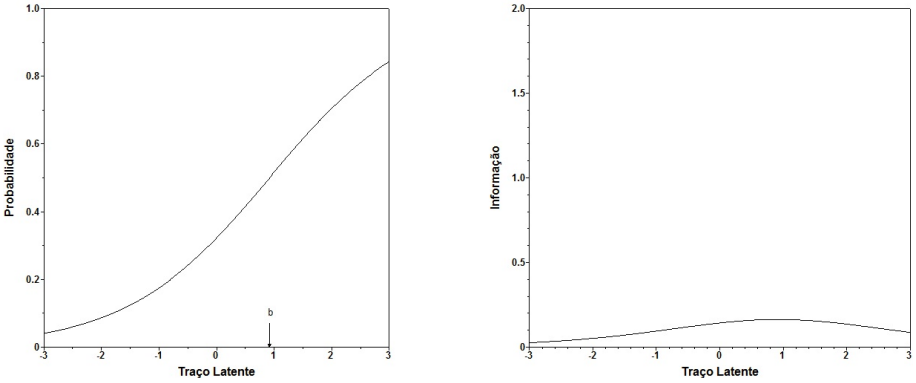

Figura 4.16: CII do item 15