



Universidade de Brasília
Departamento de Estatística

Mensuração psicofísica do olfato: uma aplicação de técnicas de Psicometria

Luísa de Moura Ferreira Dourado

Brasília
2022

Luísa de Moura Ferreira Dourado

Mensuração psicofísica do olfato: uma aplicação de técnicas de Psicometria

Orientador(a): Prof. Luís Gustavo do Amaral Vinha

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2022**

Resumo

Este trabalho teve o intuito de analisar as propriedades psicométricas do Teste Olfatório da Universidade da Pensilvânia (UPSIT), buscando fundamentar sua utilização como referência para um novo procedimento digital de mensuração do olfato e explorar os resultados de um estudo preliminar desse novo método. Na análise do UPSIT foram utilizados os dados coletados na última etapa da validação do método para a população brasileira e os resultados apontaram, por meio da análise de unidimensionalidade, da Teoria Clássica dos Testes e da Teoria da Resposta ao Item, evidências de validade e fidedignidade. Utilizando o UPSIT como referência e a partir de uma amostra coletada em um estudo preliminar, foi observada uma grande potencialidade do teste digital para mensuração do olfato. Entretanto, foram observados, em geral, itens centrados na mensuração de baixas capacidades olfativas, o que diminui o intervalo de maior qualidade de mensuração. Além disso, alguns itens destacaram-se por não apresentarem resultados tão satisfatórios quanto os demais.

Palavras-chave: Testes Psicofísicos de Avaliação do Olfato, Propriedades Psicométricas, Teoria Clássica dos Testes, Teoria da Resposta ao Item

Lista de Tabelas

1	Análise dos itens do UPSIT utilizando a TCT	26
2	Valores de AIC e BIC dos modelos de TRI	27
3	Estimativas dos parâmetros dos itens - Modelo com dois parâmetros e probabilidade de acerto ao acaso igual a 0,25	28
4	Análise dos itens dos testes utilizando a TCT	34

Lista de Quadros

1	Aromas correspondentes a cada um dos itens do UPSIT	24
2	Opções de resposta correspondentes a cada um dos itens do UPSIT	40

Lista de Figuras

1	Exemplo de curva característica do item (CCI)	21
2	Correlações tetracóricas entre os itens do UPSIT	25
3	Screeplot resultante da Análise Paralela dos itens	26
4	Curvas características dos itens por ordem de dificuldade	29
5	Curvas de informação dos itens do teste	30
6	Curva de informação do teste	30
7	Análise do perfil dos indivíduos - Sexo e idade	32
8	Análise do perfil dos indivíduos - Classificação quanto à capacidade olfativa	32
9	Comparação dos escores totais dos indivíduos em cada um dos testes . . .	33
10	Comparação das durações das realizações de cada um dos testes	33
11	Comparação das proporções de acertos dos indivíduos por itens em cada um dos testes	35
12	Curvas características dos itens 1 a 20	43
13	Curvas características dos itens 21 a 40	44
14	Curvas de informação dos itens 1 a 20	45
15	Curvas de informação dos itens 21 a 40	46

Sumário

1	Introdução	13
2	Referencial Teórico	15
2.1	Traço latente	15
2.2	Propriedades psicométricas	15
2.2.1	Fidedignidade	15
2.2.2	Validade	16
2.3	Unidimensionalidade	17
2.4	Teoria Clássica dos Testes (TCT)	17
2.4.1	Coeficiente de correlação ponto-bisserial (ρ_{pb})	18
2.4.2	Coeficiente alfa de Cronbach (α)	18
2.5	Teoria da Resposta ao Item (TRI)	19
2.5.1	Modelo logístico unidimensional de três parâmetros	20
2.5.2	Curva Característica do Item (CCI)	21
2.5.3	Função de informação	21
2.5.4	Estimação	22
2.6	Comparação e seleção de modelos	22
3	Resultados	24
3.1	Avaliação das propriedades psicométricas do UPSIT	24
3.1.1	Análise de unidimensionalidade	25
3.1.2	Análise utilizando a TCT	26
3.1.3	Modelos de TRI	27
3.2	Estudo preliminar do teste digital	31
3.2.1	Caracterização da amostra	31
3.2.2	Correlação entre os resultados dos testes	33
3.2.3	Análise dos itens	34
4	Conclusão	36
5	Referências	38
6	Apêndice	40

1 Introdução

O olfato é um dos cinco sentidos do corpo humano e está relacionado à capacidade de sentir odores. Ele é um fenômeno que não pode ser estudado facilmente em animais inferiores, sendo assim o sentido menos explorado e, conseqüentemente, o menos conhecido (GUYTON; HALL et al., 1997). Ao analisar a relação entre a disfunção olfativa e a qualidade de vida de um indivíduo, é possível perceber que essa alteração tem um grande impacto em relação ao comportamento alimentar, de segurança, higiene, social e afetivo do sujeito (PEREIRA, 2020).

Segundo Doty (1997), a importância da avaliação quantitativa do olfato na medicina está relacionada com: estabelecer a validade da queixa de um paciente, caracterizar a natureza específica do problema, acessar com confiabilidade a efetividade dos tratamentos e estabelecer modos de compensação para distúrbios olfatórios encontrados. Além disso, a avaliação olfatória é um recurso importante para auxiliar no diagnóstico precoce de doenças como o Alzheimer (VIEIRA, 2020), o Mal de Parkinson (BASTOS et al., 2019) e a COVID-19 (SEDAGHAT; GENGLER; SPETH, 2020).

Os testes de avaliação do olfato podem ser objetivos (eletrofisiológicos) ou subjetivos (psicofísicos) (FORNAZIERI, 2013). Para a prática clínica são utilizados principalmente os métodos psicofísicos, pois os eletrofisiológicos se restringem aos ambientes acadêmicos em poucos países em razão do custo e da complexidade de execução (FORNAZIERI, 2013). Apesar dos testes subjetivos serem comercialmente mais acessíveis e mais práticos, não quantificam diretamente a capacidade olfativa, mas, sim, mensuram indiretamente esse sentido utilizando instrumentos de medida constituídos por questões relacionadas a ele. Nesse contexto, o olfato pode ser interpretado como um traço latente e é de suma importância avaliar as propriedades psicométricas dos instrumentos utilizados na sua mensuração.

O Teste Olfatório da Universidade da Pensilvânia (UPSIT - *University of Pennsylvania Smell Identification Test*) é um dos testes psicofísicos mundialmente mais utilizados na atualidade. O método foi validado para a população brasileira em 2013 e é considerado por muitos como o padrão-ouro da avaliação olfatória (FORNAZIERI, 2013). Em resumo, ele consiste em um questionário de 40 itens, cada um correspondendo a um aroma que o indivíduo deve identificar entre quatro opções apresentadas.

Entretanto, estamos na era digital e é de grande relevância buscar um método semelhante ao padrão-ouro atual que não seja coletado de maneira manual, que diminua a poluição gerada e que facilite o armazenamento dos dados. Com isso, surgiu uma proposta de um novo procedimento psicofísico de mensuração do olfato baseado no UPSIT, mas utilizando um dispositivo digital que emite odores. Desse modo, foi realizado, por um

pesquisador da área, um estudo preliminar em Brasília (DF) com 179 participantes para iniciar a validação do novo procedimento. Nesse estudo, os participantes realizaram o Teste Olfatório da Universidade da Pensilvânia da maneira convencional e, em outro momento, responderam às mesmas questões utilizando o dispositivo digital (NAKANISHI et al., 2022).

Embora o UPSIT já tenha sido validado para a população brasileira por meio de diversas fases, o teste considerado como padrão-ouro ainda não foi avaliado em sua versão final considerando suas propriedades psicométricas. Sendo assim, o presente estudo buscou analisar as propriedades psicométricas do Teste Olfatório da Universidade da Pensilvânia, com o objetivo de validar sua utilização como referência e explorar os resultados do estudo preliminar do novo procedimento digital. Os resultados podem ser utilizados para propor melhorias no novo instrumento e, conseqüentemente, auxiliar a construção do método que pode melhorar e facilitar ainda mais a forma de mensuração do olfato.

2 Referencial Teórico

2.1 Traço latente

Algumas características de indivíduos não podem ser medidas diretamente, como, por exemplo, o nível de depressão, o grau de inteligência, o nível de conhecimento, etc. Essas características são chamadas de traços latentes e, apesar de não poderem ser quantificadas diretamente, em geral, podem ser mensuradas indiretamente utilizando instrumentos de medida compostos por itens a elas associados.

Esses instrumentos podem ser compostos por itens dicotômicos, em que as opções de resposta são binárias, por exemplo, certo ou errado, sim ou não. Podem ser formados também por itens politômicos, quando mais de duas opções de resposta são apresentadas, mas não há uma considerada correta. Em geral, esses itens politômicos representam grau de concordância ou frequência de acontecimentos de eventos.

Neste estudo, o olfato foi tratado como o traço latente, já que foi avaliado por métodos psicofísicos em que o sentido não é medido diretamente, mas sim por meio de um questionário. Nos instrumentos avaliados, os itens foram considerados como dicotômicos, dado que possuem quatro opções de resposta, porém apenas uma delas é considerada correta. Por isso, os procedimentos foram descritos considerando itens dicotômicos de certo e errado.

2.2 Propriedades psicométricas

A comunidade científica desenvolveu uma série de parâmetros mínimos que a medida psicométrica deve apresentar para se constituir em instrumento legítimo e válido. Os parâmetros mais básicos se referem, além da análise dos itens, à fidedignidade e à validade do instrumento, que constituem temas centrais da Psicometria (PASQUALI, 2017).

2.2.1 Fidedignidade

A fidedignidade determina quanto da variabilidade nos escores observados ao longo das inúmeras replicações de um teste é devida ao escore verdadeiro, e não ao erro aleatório (HUTZ; BANDEIRA; TRENTINI, 2015). Em termos práticos, essa propriedade se refere à estabilidade com que os escores dos respondentes conservam-se em aplicações alternativas de um mesmo teste ou em formas equivalentes de testes distintos (HUTZ; BANDEIRA; TRENTINI, 2015). Ou seja, essa propriedade pode ser vista como uma

forma de avaliar a intensidade do erro cometido na medida.

Existem diferentes formas de obter uma medida que expresse o nível de fidedignidade de um teste. A correlação entre os escores é utilizada nos métodos Teste-Reteste, Formas Paralelas e Inter-Avaliadores. No primeiro método, um mesmo teste é aplicado para o mesmo grupo de respondentes em momentos diferentes, enquanto no procedimento Formas Paralelas dois testes diferentes e equivalentes são aplicados aos mesmos indivíduos em um mesmo momento. Já no Inter-Avaliadores, o procedimento consiste na aplicação de um só teste e na correção feita por dois ou mais avaliadores diferentes. Além disso, também pode ser utilizada a correlação entre os itens de um único questionário buscando avaliar a consistência interna do teste.

2.2.2 Validade

Considera-se um teste válido aquele em que o instrumento mede os comportamentos que são a expressão do traço latente que se deseja mensurar (HUTZ; BANDEIRA; TRENTINI, 2015). Trata-se, portanto, do grau em que um teste realmente mede aquilo que se propõe a medir. Os conceitos de validade e fidedignidade estão bem relacionados, mas considera-se a fidedignidade um pré-requisito para a validade. Um teste pode ser fidedigno e não ser válido, mas não pode ser válido sem ser fidedigno.

Idealmente, em um teste válido os itens devem refletir o traço latente como se fossem um espelho, no que se refere ao conceito, ao conteúdo e às relações com outras variáveis (HUTZ; BANDEIRA; TRENTINI, 2015). Considerando isso, existem vários tipos de classificação da validação, mas as principais são a validade de conteúdo, de critério e de construto.

A validade de conteúdo trata do quanto o teste aborda comportamentos que são expressão do traço latente em questão, ou seja, se os itens do teste constituem uma amostra representativa do universo de itens relacionados construto. A avaliação dessa propriedade é realizada pela análise dos itens de forma mais teórica (ERTHAL, 2010).

Já a validade de critério indica a relação entre o desempenho no teste e algum outro critério tido como um importante indicador do traço latente estudado. Nesse caso, o critério pode ser outro instrumento que mensure a mesma característica do que aquele que precisa ser validado, mas é importante que seja relevante, válido e não contaminado (HUTZ; BANDEIRA; TRENTINI, 2015).

A validade de construto é considerada a maneira mais fundamental de validade dos instrumentos psicológicos, visto que estabelece uma forma direta de averiguar a hipótese da legitimidade da representação comportamental dos traços latentes (PASQUALI, 2017). Segundo Cronbach e Meehl (1955), essa validade é definida como a ca-

racterística de um teste enquanto mensuração de um atributo, o qual não tenha sido definido operacionalmente. Porém, segundo Pasquali (2017), o problema não é descobrir o construto a partir de um teste existente, mas sim descobrir se o teste constitui uma representação legítima do traço latente. A validade de construto de um teste pode ser avaliada sob múltiplas visões, entre elas estão a Análise Fatorial, a Teoria Clássica dos Testes e a Teoria de Resposta ao Item.

2.3 Unidimensionalidade

A avaliação da unidimensionalidade de um instrumento pode ser realizada utilizando a Análise Fatorial. Essa técnica consiste em verificar se uma série de itens pode ser reduzida idealmente a uma única dimensão ou variável, chamada de fator, com o qual todas as variáveis da série estão relacionadas. Sendo esse o caso, conclui-se que o teste é unidimensional (PASQUALI, 2017).

Em termos gerais, esse método baseia-se na matriz de correlações. No caso em que os itens são dicotômicos podem ser utilizadas as correlações tetracórica ou *phi*. A correlação tetracórica é utilizada quando duas variáveis contínuas normalmente distribuídas foram artificialmente reduzidas a duas categorias.

Não há um critério exato para definir o número de fatores que devem ser utilizados na Análise Fatorial, mas a Análise Paralela é uma técnica bastante utilizada para esse fim. Esse é um procedimento de simulação que consiste na construção de um conjunto hipotético de matrizes de correlação de variáveis, utilizando como base a mesma dimensionalidade do conjunto de dados reais. São feitas fatorações dessas matrizes e as médias dos autovalores simulados são comparadas de forma pareada com os autovalores dos dados reais. Essa comparação é realizada a partir da análise do *scree plot* dos dados observados e simulados e o número de fatores corresponde ao número de autovalores reais maiores que a média dos simulados, ou superiores ao intervalo de confiança (DAMÁSIO, 2012).

2.4 Teoria Clássica dos Testes (TCT)

A Teoria Clássica dos Testes utiliza o escore total obtido por meio do instrumento de medida e o seu erro de medida como referência principal (HUTZ; BANDEIRA; TRENTINI, 2015). O escore total do *j*-ésimo indivíduo é definido por

$$T_j = \sum_{i=1}^I U_{ij}, \quad (2.4.1)$$

onde:

- I é o número de itens no questionário;
- $U_{ij} = \begin{cases} 0, & \text{se o } j\text{-ésimo indivíduo fornece resposta incorreta para o } i\text{-ésimo item;} \\ 1, & \text{se o } j\text{-ésimo indivíduo fornece resposta correta para o } i\text{-ésimo item.} \end{cases}$

Segundo Hutz; Bandeira; Trentini (2015), o experimento mental que fundamenta a TCT baseia-se no escore verdadeiro, que seria a média esperada para os escores de um indivíduo, caso fosse possível que ele respondesse ao instrumento infinitas vezes, todas elas sem se lembrar da resposta anterior. Na TCT, além da mensuração do traço latente por meio do escore total, são utilizadas algumas medidas para avaliar a qualidade do instrumento de medida.

2.4.1 Coeficiente de correlação ponto-bisserial (ρ_{pb})

O coeficiente de correlação ponto-bisserial (ρ_{pb}) mede a correlação linear entre o escore total do teste e as respostas a cada um dos itens. O coeficiente referente ao i -ésimo item pode ser escrito como

$$\rho_{pb_i} = \frac{\bar{T}_A - \bar{T}}{S_T} \sqrt{\frac{p_i}{1 - p_i}}, \quad (2.4.2)$$

onde:

- \bar{T}_A é a média dos escores dos respondentes que acertaram o i -ésimo item;
- \bar{T} é a média dos escores do teste;
- S_T é o desvio padrão dos escores do teste;
- p_i é a proporção de acertos observados no i -ésimo item.

Segundo Primi (2012), um alto valor do coeficiente (ρ_{pb_i}) indica que o i -ésimo item contribui para aumentar a variância dos escores, o que ajuda na discriminação dos indivíduos e é de grande importância para a qualidade da mensuração do traço latente. Valores negativos para correlação entre o item e o escore total apontam itens incoerentes com o teste, pois indicam uma tendência a observar mais respostas positivas para indivíduos com escores baixos do que para indivíduos com escores altos.

2.4.2 Coeficiente alfa de Cronbach (α)

O coeficiente alfa de Cronbach apresenta o grau de covariância dos itens entre si, servindo de indicador de consistência interna de um instrumento de medida. O alfa de

Cronbach de um teste é dado por

$$\alpha = \frac{I}{I - 1} \left(1 - \frac{\sum_{i=1}^I p_i(1 - p_i)}{S_T^2} \right), \quad (2.4.3)$$

onde:

- I é o número de itens;
- p_i é a proporção de acertos observados no i -ésimo item;
- S_T^2 é a variância dos escores do teste.

Esse coeficiente varia de 0 a 1, em que 0 indica ausência total de consistência interna dos itens e o 1 indica completa presença de consistência interna. Na literatura não há um consenso sobre os valores de referência para alfa de Cronbach, mas, em geral, considera-se a consistência interna dos itens inadequada quando o valor do coeficiente está abaixo de 0,7. Por outro lado, avalia-se que há uma redundância nos itens, isto é, itens que abordam o mesmo elemento do traço latente, para valores de alfa superiores a 0,9.

2.5 Teoria da Resposta ao Item (TRI)

Na Teoria Clássica dos Testes as estimativas do traço latente são dependentes do conjunto de itens que compõem o teste e a estimação dos parâmetros de cada item depende do conjunto de respondentes. A Teoria da Resposta ao Item foi desenvolvida para solucionar essas dependências. Ou seja, esse método permite a comparação de resultados de indivíduos pertencentes a populações diferentes mesmo que tenham realizado outro teste, desde que existam alguns itens em comum (ANDRADE; TAVARES; VALLE, 2000).

A TRI consiste em um grupo de modelos que tem como variável resposta a probabilidade de um indivíduo escolher uma dada resposta a um item condicionado ao seu nível do traço latente e aos parâmetros do item. Segundo Andrade; Tavares; Valle (2000), os modelos de TRI propostos na literatura são substancialmente dependentes de três aspectos: a natureza do item (dicotômicos ou politômicos), o número de populações envolvidas e a quantidade de traços latentes que estão sendo medidos. Neste estudo foram utilizados modelos para itens dicotômicos, com apenas uma população envolvida e apenas um traço latente a ser mensurado.

As duas principais suposições para a utilização desses modelos são a unidimensionalidade e a independência local. A unidimensionalidade garante que apenas um traço latente é responsável pelas respostas dos indivíduos aos itens, ou pelo menos que haja

um traço latente dominante responsável pelo desempenho em um teste. A independência local certifica que, mantido o traço latente constante, as respostas do indivíduo aos itens serão independentes, ou seja, a resposta dada a um item não deve influenciar uma resposta dada em outro item.

2.5.1 Modelo logístico unidimensional de três parâmetros

Os modelos logísticos são muito utilizados quando os itens têm respostas dicotômicas e esses modelos podem ter até três parâmetros por item. Considerando que o modelo com três parâmetros é o mais complexo e os demais podem ser obtidos facilmente a partir dele, ele foi o modelo detalhado neste estudo.

O modelo logístico de três parâmetros é definido por

$$P(U_{ij} = 1|\theta_j) = c_i + \frac{1 - c_i}{1 + \exp[-a_i(\theta_j - b_i)]}, \quad (2.5.1)$$

onde:

- $U_{ij} = \begin{cases} 0, & \text{se o } j\text{-ésimo indivíduo fornece resposta incorreta para o } i\text{-ésimo item;} \\ 1, & \text{se o } j\text{-ésimo indivíduo fornece resposta correta para o } i\text{-ésimo item;} \end{cases}$
- θ_j é o traço latente do j -ésimo respondente;
- c_i é o parâmetro de probabilidade de acerto ao acaso do i -ésimo item;
- a_i é o parâmetro de discriminação (ou de inclinação) do i -ésimo item;
- b_i é o parâmetro de dificuldade (ou de posição) do i -ésimo item.

O parâmetro a_i corresponde à capacidade do i -ésimo item de diferenciar indivíduos quanto à intensidade do traço latente. O parâmetro b_i representa o grau de dificuldade do i -ésimo item gerar uma resposta correta (SARTES; SOUZA-FORMIGONI, 2013) e é medido na mesma unidade do traço latente. O parâmetro c_i é considerado o parâmetro de acerto ao acaso do i -ésimo item, pois quando θ_j tende a $-\infty$, a probabilidade do j -ésimo indivíduo acertar o i -ésimo item é igual a c_i , tratando-se, portanto, de uma probabilidade e varia entre 0 e 1.

Nas situações em que o parâmetro de acerto ao acaso não existe ou pode ser desconsiderado, tem-se c_i igual a zero na equação 2.5.1, o que equivale ao modelo logístico de dois parâmetros. Em outros casos, além de desconsiderar a probabilidade de acerto ao acaso, o parâmetro de discriminação de todos os itens é considerado igual a um, correspondendo ao modelo logístico de um parâmetro, ou modelo de Rasch.

2.5.2 Curva Característica do Item (CCI)

A curva característica do item é uma representação gráfica da probabilidade do j -ésimo indivíduo com valor do traço latente igual a θ_j acertar o i -ésimo item. O modelo proposto baseia-se no fato de que a probabilidade de resposta correta para o item é função crescente não linear de θ_j .

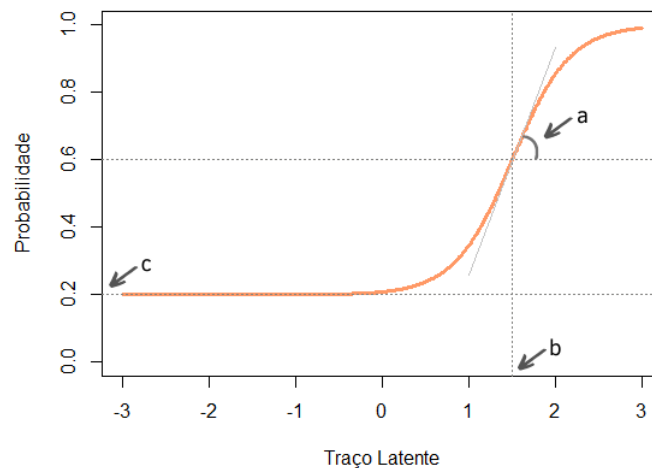


Figura 1: Exemplo de curva característica do item (CCI)

Como ilustrado na Figura 1, a curva característica do item tem forma de “S” com inclinação e deslocamento na escala do traço latente definidos pelos parâmetros do item. O valor de a_i é proporcional à inclinação da curva do i -ésimo item, sendo que as curvas mais inclinadas (com tendência a formar um ângulo de 90°) indicam maior discriminação e as menos inclinadas indicam menor discriminação (ANDRADE; TAVARES; VALLE, 2000). Já o valor de b_i , em termos gráficos, pode ser interpretado como o valor do traço latente necessário para obter uma probabilidade de acerto igual a $(1 + c_i)/2$ no i -ésimo item.

2.5.3 Função de informação

A representação gráfica da função de informação é outro recurso importante para a avaliação dos itens e do teste em geral. A função de informação do item é baseada na informação de Fisher e é utilizada para analisar quanto um item contém de informação para a medida do traço latente. A informação fornecida pelo i -ésimo item no nível θ do traço latente pode ser calculada por

$$I_i(\theta) = \frac{\left[\frac{d}{d\theta} P_i(\theta)\right]^2}{P_i(\theta) [1 - P_i(\theta)]}, \quad (2.5.2)$$

onde $P_i(\theta) = P(U_{ij} = 1|\theta)$ e $\frac{d}{d\theta}$ representa a derivada da função em relação à θ .

A função de informação do teste é dada pela soma das funções de informação de cada item que o compõem, ou seja,

$$I(\theta) = \sum_{i=1}^I I_i(\theta). \quad (2.5.3)$$

2.5.4 Estimação

Como mencionado anteriormente, na TRI a probabilidade de um indivíduo acertar uma resposta em um determinado item decorre do seu traço latente e dos parâmetros que configuram esse item. Em geral, esses parâmetros são desconhecidos e apenas as respostas dos indivíduos aos itens do teste estão disponíveis. Por isso, a estimação dos parâmetros dos itens e do traço latente dos respondentes é considerada uma das etapas mais importantes da TRI.

A estimação pode ser dividida, resumidamente, em três cenários:

- quando os parâmetros dos itens já são conhecidos e é necessário estimar os traços latentes dos indivíduos;
- quando os traços latentes dos indivíduos já são conhecidos e é necessário estimar os parâmetros dos itens;
- a situação mais comum, em que todos os parâmetros envolvidos são desconhecidos e é necessário estimá-los.

Em ambos os cenários citados, geralmente a estimação é realizada pelo Método da Máxima Verossimilhança através da aplicação de algum processo iterativo, como o algoritmo Newton-Raphson ou *Scoring* de Fisher, mas alguns procedimentos bayesianos também são utilizados com frequência (ANDRADE; TAVARES; VALLE, 2000).

2.6 Comparação e seleção de modelos

Utilizando a Teoria da Resposta ao Item, alguns modelos foram construídos e os seus parâmetros estimados, mas é preciso considerar que todos os modelos geram alguma perda de informação. Portanto, foram utilizados critérios para selecionar o melhor modelo possível dentre aqueles que foram ajustados para explicar o fenômeno sob estudo.

Os critérios de informação de Akaike (AIC) e Bayesiano (BIC) são possíveis opções para estimar a partir dos dados a incerteza de cada modelo. Em ambos os critérios os

modelos que apresentam menores valores de AIC e BIC são preferíveis.

O critério de informação de Akaike é calculado por

$$AIC = -2 \log(L(\hat{\theta})) + 2K, \quad (2.6.1)$$

onde L é a função de verossimilhança do modelo e K é o número de parâmetros do modelo.

Similarmente, o critério de informação Bayesiano é calculado por

$$BIC = -2 \log(L(\hat{\theta})) + K \log(n), \quad (2.6.2)$$

onde n é o tamanho da amostra em questão.

3 Resultados

Neste capítulo serão apresentados os resultados obtidos. Inicialmente, foi realizada a análise das propriedades psicométricas do Teste Olfatório da Universidade da Pensilvânia (UPSIT) com base nos dados coletados na última etapa da validação desse método para a população brasileira. Em seguida, foi feita a análise dos resultados do estudo preliminar do novo procedimento digital.

3.1 Avaliação das propriedades psicométricas do UPSIT

O banco de dados utilizado nesta etapa contém respostas de 1.578 pessoas para o UPSIT em português e para um questionário com informações sociodemográficas e de saúde. As informações foram obtidas em um estudo realizado em São Paulo (SP) no período de dezembro de 2011 a agosto de 2012 (FORNAZIERI, 2013).

Vale destacar que o UPSIT consiste em um teste com 40 itens com quatro opções de resposta, em que apenas uma é considerada correta. Mais especificamente, o teste é constituído por quatro livretos de papel contendo dez odores em cada, sendo um odor por página. O estímulo odorífero fica embebido em microcápsulas plásticas em uma faixa localizada no rodapé de cada página. Durante a realização da avaliação, a pessoa examinada raspa essa faixa, fazendo com que o odor seja liberado. Em seguida, a pessoa aproxima o livreto do nariz e responde, entre quatro opções de aromas, aquele que corresponde ao que sentiu (FORNAZIERI, 2013).

Os aromas correspondentes a cada um dos itens estão descritos a seguir e as opções de resposta de cada um dos itens podem ser vistas no Apêndice 6.1.

Quadro 1: Aromas correspondentes a cada um dos itens do UPSIT

Item	Aroma	Item	Aroma	Item	Aroma	Item	Aroma
1	Pizza	11	Cebola	21	Perfume	31	Solvente
2	Chiclete	12	Suco de frutas	22	Flor	32	Gramma
3	Mentol	13	Talco de bebê	23	Pêssego	33	Fumaça
4	Cereja	14	Jasmin	24	Pneu	34	Madeira
5	Óleo de motor	15	Canela	25	Picles	35	Uva
6	Menta	16	Gasolina	26	Abacaxi	36	Alho
7	Banana	17	Morango	27	Framboesa	37	Sabão
8	Cravo	18	Café	28	Laranja	38	Gás Natural
9	Couro	19	Pão de Mel	29	Nozes	39	Rosa
10	Coco	20	Maçã	30	Melancia	40	Amendoim

Observou-se que alguns participantes não responderam alguns itens e outros selecionaram mais de uma opção em um mesmo item. Essas respostas não apresentaram um padrão por item ou por respondentes. Apenas 94 das 63.120 respostas obtidas (0,15%) foram em branco ou com dupla marcação. Além disso, 95,56% dos integrantes da amostra responderam todos os itens de acordo com o esperado. Portanto, os itens com esse tipo de resposta foram considerados como incorretos para as análises.

Fornazieri (2013) descreve com mais detalhes a amostra e os resultados obtidos. Neste capítulo serão retratadas a análise da unidimensionalidade do teste e a análise dos itens por meio da TCT e dos modelos de TRI.

3.1.1 Análise de unidimensionalidade

A Figura 2 mostra as correlações tetracóricas para cada par de itens do questionário. Não foram observadas correlações negativas e o maior valor para esse coeficiente foi igual a 0,6. Ou seja, apesar de, em geral, os itens do teste estarem correlacionados positivamente entre si, essas relações não são tão fortes. Os itens 5, 19 e 34 foram os que apresentaram menores correlações.

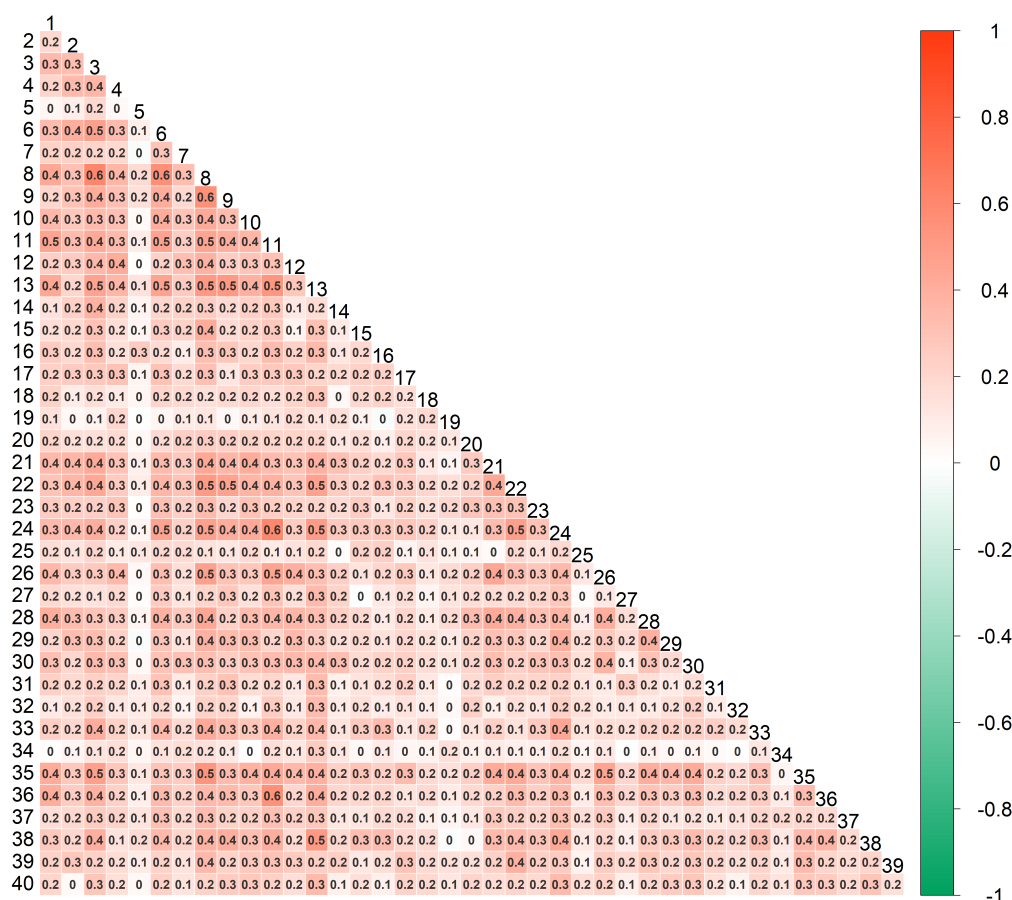


Figura 2: Correlações tetracóricas entre os itens do UPSIT

De acordo com os resultados da Análise Paralela mostrados na Figura 3, é possível perceber que há uma dimensão dominante, apesar do gráfico mostrar o valor do segundo fator um pouco superior ao intervalo de confiança para os dados reamostrados.

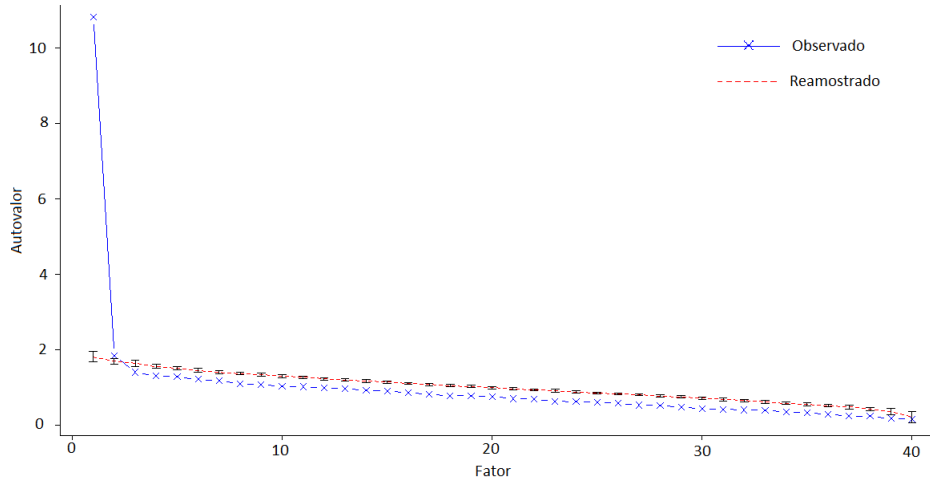


Figura 3: Screeplot resultante da Análise Paralela dos itens

Aplicando a Análise Fatorial e utilizando apenas um fator no modelo, tem-se que apenas os itens 5, 19, 25 e 34 apresentam carga fatorial menor do que 0,3, sendo que esse fator explica 25% da variância dos dados. Portanto, foi assumida a suposição de unidimensionalidade e os itens que não tiveram um desempenho tão satisfatório dentro desse aspecto foram analisados com maior detalhe nos demais procedimentos.

3.1.2 Análise utilizando a TCT

A Tabela 4 resume o comportamento dos itens com base em medidas da Teoria Clássica dos Testes. O menor valor para a proporção de acertos foi 0,47, ou seja, todos os itens foram acertados por pelo menos 47% dos respondentes. Apesar de nenhum item ter sido acertado por todos os respondentes, apenas 3% dos indivíduos erraram o item 13. Os valores do coeficiente de correlação ponto-bisserial variaram de 0,2 a 0,48, ou seja, não foi possível identificar itens incoerentes com o teste.

Tabela 1: Análise dos itens do UPSIT utilizando a TCT

Item	Proporção de Acertos	ρ_{pb}	Item	Proporção de Acertos	ρ_{pb}
1	0,69	0,44	21	0,88	0,42
2	0,80	0,39	22	0,87	0,45
3	0,91	0,44	23	0,75	0,41
4	0,84	0,38	24	0,92	0,43

Continua

Continuação					
Item	Proporção de Acertos	ρ_{pb}	Item	Proporção de Acertos	ρ_{pb}
5	0,47	0,20	25	0,47	0,28
6	0,91	0,42	26	0,91	0,36
7	0,80	0,35	27	0,83	0,27
8	0,92	0,48	28	0,84	0,43
9	0,85	0,43	29	0,68	0,41
10	0,83	0,43	30	0,82	0,41
11	0,92	0,43	31	0,72	0,33
12	0,82	0,39	32	0,47	0,29
13	0,97	0,34	33	0,90	0,34
14	0,79	0,32	34	0,76	0,20
15	0,81	0,32	35	0,89	0,45
16	0,88	0,30	36	0,86	0,39
17	0,87	0,33	37	0,82	0,32
18	0,82	0,28	38	0,9	0,37
19	0,75	0,23	39	0,66	0,40
20	0,58	0,36	40	0,88	0,29

3.1.3 Modelos de TRI

Foram ajustados aos dados três modelos logísticos unidimensionais utilizando a função *tpm* do pacote *ltm* do *software* R (RIZOPOULOS, 2006). Os três modelos são baseados na equação 2.5.1, sendo que no primeiro são estimados os três parâmetros, no segundo apenas os parâmetros de discriminação e dificuldade considerando a probabilidade de acerto ao acaso fixada em 0,25 e no terceiro esses mesmos parâmetros desconsiderando a probabilidade de acerto ao acaso. De acordo com a Tabela 2, tem-se que o segundo é o que melhor se ajusta, já que apresenta os menores valores de AIC e BIC. Portanto, esse foi o modelo selecionado para as análises a seguir.

Tabela 2: Valores de AIC e BIC dos modelos de TRI

Modelo	AIC	BIC
Modelo 1	52963,43	53607,10
Modelo 2	52935,36	53364,47
Modelo 3	52941,38	53370,50

A Tabela 3 apresenta as estimativas dos parâmetros para cada um dos itens no

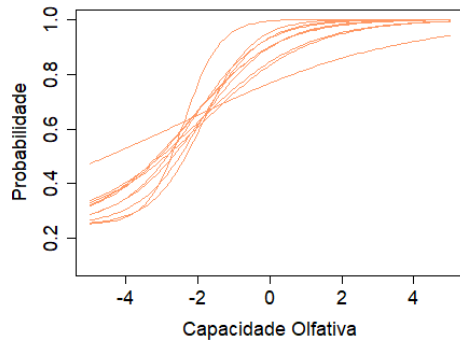
modelo selecionado. Cabe ressaltar que a_i corresponde ao parâmetro de discriminação do i -ésimo item e b_i corresponde ao parâmetro de dificuldade. Em geral, as estimativas mostram um resultado satisfatório em relação à qualidade do teste.

Tabela 3: Estimativas dos parâmetros dos itens - Modelo com dois parâmetros e probabilidade de acerto ao acaso igual a 0,25

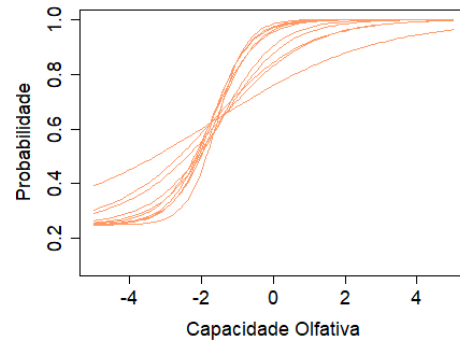
Item	a_i	b_i	Item	a_i	b_i	Item	a_i	b_i	Item	a_i	b_i
1	1,26	-0,38	11	1,82	-1,75	21	1,64	-1,43	31	0,81	-0,75
2	1,26	-1,06	12	1,19	-1,24	22	1,74	-1,32	32	0,82	1,21
3	1,78	-1,63	13	2,01	-2,39	23	1,12	-0,76	33	1,05	-2,19
4	1,10	-1,48	14	0,82	-1,34	24	1,92	-1,72	34	0,33	-2,41
5	0,42	2,27	15	0,80	-1,53	25	0,76	1,32	35	1,75	-1,47
6	1,56	-1,75	16	0,84	-2,24	26	1,45	-1,86	36	1,28	-1,50
7	0,94	-1,23	17	0,95	-1,96	27	0,67	-2,00	37	0,77	-1,72
8	2,41	-1,57	18	0,67	-1,86	28	1,47	-1,24	38	1,21	-1,93
9	1,46	-1,33	19	0,44	-1,71	29	1,19	-0,35	39	1,14	-0,21
10	1,37	-1,21	20	1,24	0,28	30	1,38	-1,13	40	0,82	-2,28

Nota-se que os itens 8 e 13 foram, respectivamente, os itens mais discriminativos e os itens 34, 5 e 19, respectivamente, os menos discriminativos. Em relação ao parâmetro de dificuldade, o item 5 foi o que apresentou maior valor desse parâmetro e o item 34 o menor valor.

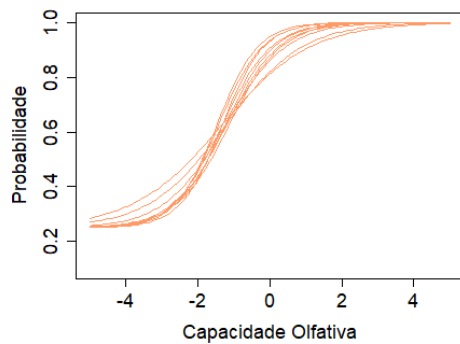
Analisando o comportamento das curvas características dos itens agrupadas por ordem de dificuldade, ilustradas na Figura 4, percebe-se que os itens apresentaram, em geral, um comportamento bem semelhante, tanto em relação à dificuldade, quanto à discriminação. Em grande parte, esse comportamento dos itens está centrado na mensuração de traços latentes baixos. É possível verificar a curva característica de cada um dos itens separadamente no Apêndice 6.2.



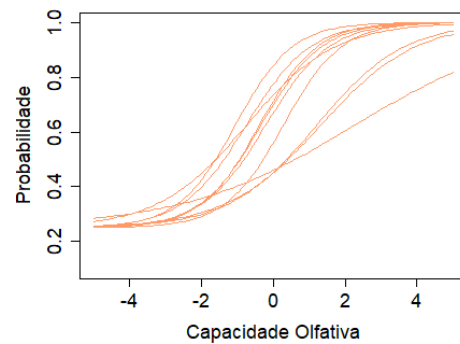
(a) Curvas características dos itens 34, 13, 40, 16, 33, 27, 17, 38, 26 e 18



(b) Curvas características dos itens 11, 6, 37, 24, 19, 3, 8, 15, 36 e 4



(c) Curvas características dos itens 35, 21, 14, 9, 22, 28, 12, 7, 10 e 30



(d) Curvas características dos itens 2, 23, 31, 1, 29, 39, 20, 32, 25 e 5

Figura 4: Curvas características dos itens por ordem de dificuldade

Comparando o comportamento das curvas de informação dos itens de acordo com a Figura 5, tem-se que alguns itens possuem um pico de informação ao longo da escala de capacidade olfativa, enquanto outros possuem menores valores de informação, mas em um intervalo maior da escala do traço latente. O item 8 foi o responsável pelo maior pico mostrado no gráfico. Alguns itens, como o 5, 19 e 34, apresentam valores muito baixos de informação ao longo de toda a escala. É possível verificar a curva de informação de cada um dos itens separadamente no Apêndice 6.3.

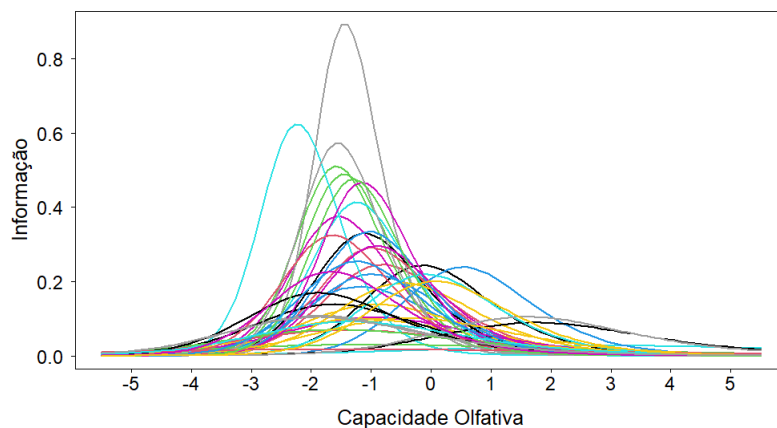


Figura 5: Curvas de informação dos itens do teste

A Figura 6 mostra que a informação do teste está bem concentrada no intervalo de capacidade olfativa de -2 a 0, assim como foi observado no gráfico dos itens. Além disso, é possível perceber que o teste apresenta maiores valores de informação para menores valores de capacidade olfativa, ou seja, esse instrumento tem uma capacidade maior em diferenciar os indivíduos com baixa capacidade olfativa.

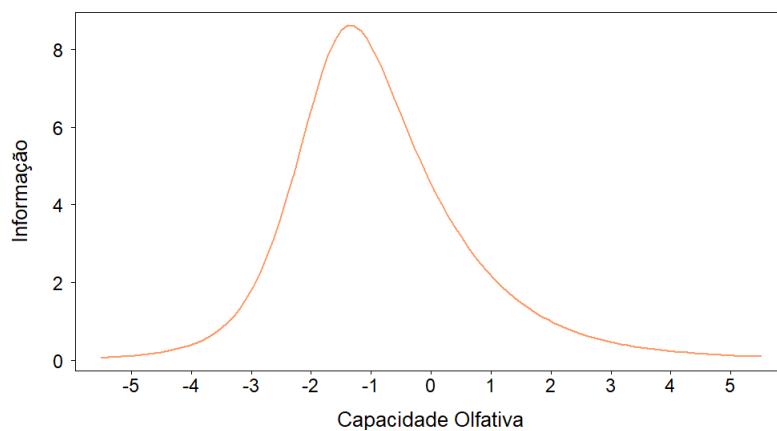


Figura 6: Curva de informação do teste

O coeficiente alfa de Cronbach do teste foi igual a 0,817, o que sinaliza uma boa consistência interna do teste e é um bom indicativo da fidedignidade do teste. Retirando os itens 5, 19 e 34 que não apresentaram um comportamento tão satisfatório comparado com os demais, esse coeficiente foi igual a 0,824. Ou seja, a retirada desses itens não impacta na fidedignidade do teste.

De acordo com os resultados obtidos, é possível assumir que o teste apresenta evidência de validade, pois pode ser interpretado como apenas um fator, expressou um indicador de fidedignidade dentro dos padrões esperados e a análise pela TRI apresentou bom desempenho dos itens. Entretanto, os itens 5, 19 e 34, que representam, respecti-

vamente os aromas de óleo de motor, pão de mel e madeira, são pontos de atenção e a exclusão desses itens poderia ser considerada em uma possível reformulação do teste.

3.2 Estudo preliminar do teste digital

Nesta etapa, o banco de dados utilizado é referente ao estudo preliminar para avaliação do teste digital realizado em Brasília (DF) no período de março a dezembro de 2021 e é composto pelas respostas de 179 indivíduos a um questionário sociodemográfico e clínico de distúrbios do olfato e aos testes UPSIT e olfatométrica digital (NAKANISHI et al., 2022).

O teste de olfatométrica digital, nesse contexto, refere-se a um teste composto pelos mesmos itens, com as mesmas opções de resposta e na mesma ordem que o UPSIT, porém utilizando um dispositivo digital que emite odores. O dispositivo, também chamado de demonstrador digital de fragrâncias, trata-se de um aparelho utilizado na indústria de cosméticos para testes de perfumes. As fragrâncias são armazenadas em microcartuchos e são liberadas através de um pequeno orifício do dispositivo quando a pessoa examinada aperta um botão com esse comando no tablet que funciona como suporte para o exame. Após a liberação desse odor, a pessoa tem a opção de solicitar que o aroma seja liberado mais uma vez ou marcar a opção que considera ser a correspondente àquele aroma (NAKANISHI et al., 2022).

No experimento, cada participante foi considerado seu próprio controle, uma vez que todos realizaram os dois testes. A ordem de aplicação dos testes foi realizada de forma cruzada, com intervalo de no mínimo um e no máximo sete dias entre eles. Nas realizações dos testes consideradas nessa amostra não foram identificadas respostas fora do padrão, ou seja, itens em branco ou com dupla marcação.

Um problema relatado pelo pesquisador é que o demonstrador digital não permite a liberação de alguns odores que podem ser nocivos aos participantes. Além disso, grande parte dos odores utilizados não tinham sido fabricados até então pelo fabricante do equipamento. Ou seja, o teste não é exatamente igual ao UPSIT e os aromas ainda não foram validados. Dessa forma, alguns problemas encontrados podem estar relacionados com a fragrância utilizada e não com o teste em si.

3.2.1 Caracterização da amostra

Os testes permitem separar os indivíduos em classes de capacidade olfativa. Aqueles que não apresentam perda olfatória são classificadas como normosmia. Já as pessoas com alguma perda nesse sentido são separadas em anosmia, que representa a perda com-

pleta do olfato e em hiposmia, correspondente à perda parcial do olfato. A hiposmia pode ser dividida em leve, moderada ou severa, de acordo com a intensidade da redução da capacidade olfativa. Além disso, existe a classificação de provável simulador, que corresponde aos indivíduos que provavelmente estão forjando o resultado do teste com um escore propositalmente muito baixo.

Além do escore total no teste, o sexo e a idade do indivíduo são considerados na sua classificação em relação à capacidade olfativa. De acordo com a Figura 7, é possível perceber que a amostra é composta por indivíduos com idade entre 18 e 60 anos e predominantemente do sexo feminino.

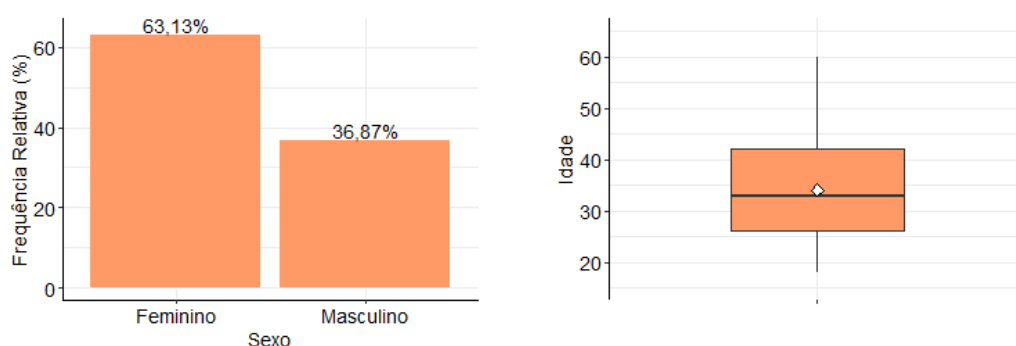


Figura 7: Análise do perfil dos indivíduos - Sexo e idade

Separando os sujeitos de acordo com a classificação olfativa, utilizando o escore total no UPSIT e os critérios definidos na validação do UPSIT para a população brasileira, nota-se, de acordo com a Figura 8, que não foram identificados prováveis simuladores nessa amostra, o que é adequado para a avaliação do instrumento.

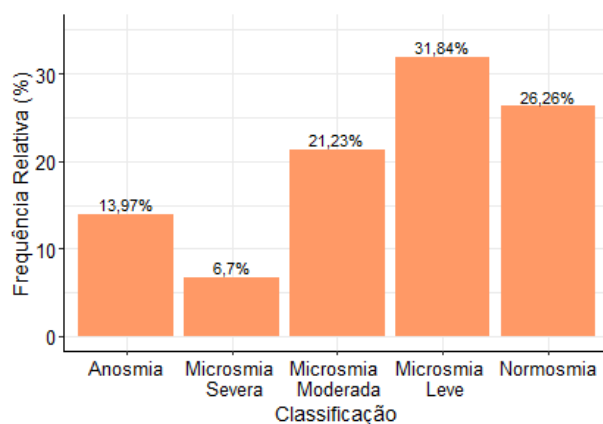


Figura 8: Análise do perfil dos indivíduos - Classificação quanto à capacidade olfativa

3.2.2 Correlação entre os resultados dos testes

Considerando que cada indivíduo realizou ambos os testes, é possível comparar esses resultados para verificar os pontos concordantes e discordantes entre esses métodos. A Figura 9 mostra que os escores dos indivíduos nos testes foram bem semelhantes entre si e apresentaram uma correlação linear positiva. O coeficiente de correlação de Spearman igual a 0,73 entre essas variáveis e o p-valor menor que 0,01 para o teste de significância reforçam a correlação forte e significativa entre os escores totais no teste digital e no UPSIT. Em geral, os escores obtidos nos testes concentraram-se em valores elevados, mas os escores no UPSIT foram mais dispersos.

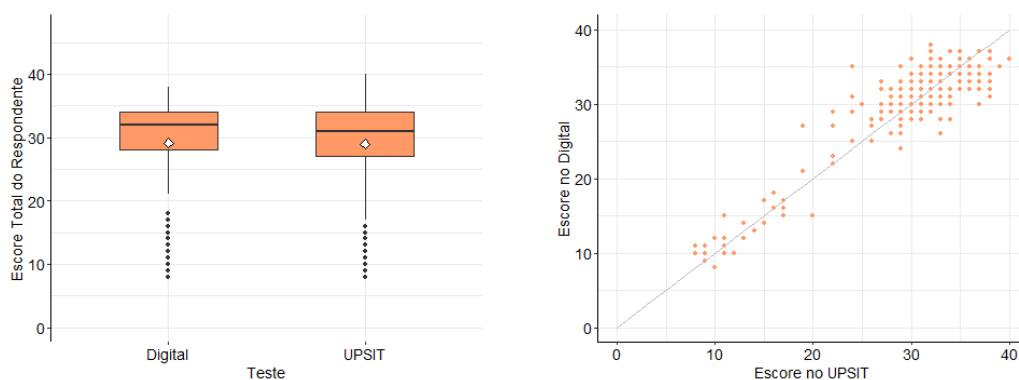


Figura 9: Comparação dos escores totais dos indivíduos em cada um dos testes

Comparando o tempo para realização de cada um dos testes, de acordo com a Figura 10, é possível perceber que a duração do teste digital em geral foi menor do que a do UPSIT. Esses tempos apresentaram um coeficiente de correlação de Spearman igual a 0,14 ($p\text{-valor} > 0,05$), ou seja, observa-se uma correlação positiva, fraca e não significativa entre a duração dos testes.

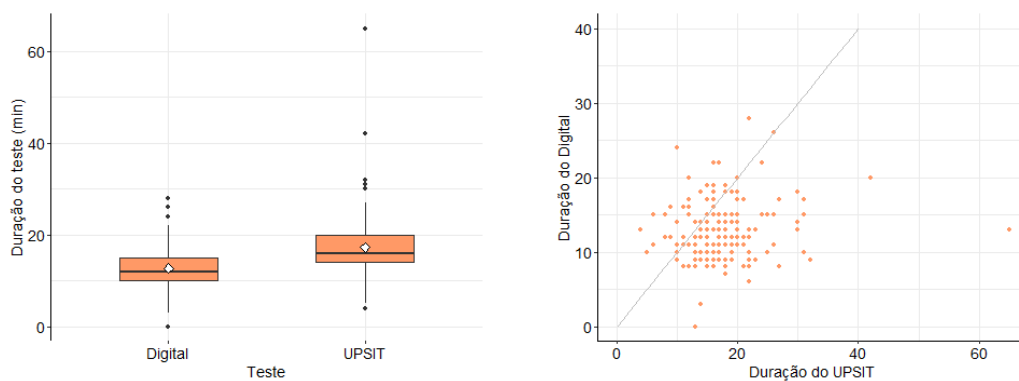


Figura 10: Comparação das durações das realizações de cada um dos testes

3.2.3 Análise dos itens

Verificando o comportamento de cada um dos itens de acordo com a Tabela 4, nota-se que apesar de não terem sido identificados itens com coeficiente de correlação ponto-bisserial negativo, que seria um indicativo de incoerência com o teste, os itens 31 e 25 apresentaram coeficiente de correlação ponto-bisserial menor do que 0,1 no teste digital. Esse resultado muito baixo para o coeficiente não foi observado no UPSIT e mostra que esses itens não contribuem muito para aumentar a variância dos escores, o que reduz a capacidade de discriminação dos indivíduos e a qualidade da mensuração do traço latente.

Tabela 4: Análise dos itens dos testes utilizando a TCT

Item	UPSIT		Digital		Item	UPSIT		Digital	
	Proporção de Acertos	ρ_{pb}	Proporção de Acertos	ρ_{pb}		Proporção de Acertos	ρ_{pb}	Proporção de Acertos	ρ_{pb}
1	0,55	0,30	0,82	0,35	21	0,81	0,55	0,46	0,16
2	0,82	0,46	0,87	0,59	22	0,80	0,59	0,72	0,40
3	0,87	0,54	0,86	0,59	23	0,82	0,68	0,82	0,77
4	0,84	0,61	0,86	0,66	24	0,90	0,54	0,84	0,70
5	0,78	0,43	0,43	0,16	25	0,44	0,40	0,11	0,06
6	0,87	0,55	0,84	0,69	26	0,75	0,43	0,75	0,47
7	0,67	0,42	0,84	0,53	27	0,84	0,56	0,69	0,55
8	0,87	0,72	0,79	0,77	28	0,61	0,32	0,85	0,60
9	0,80	0,57	0,84	0,60	29	0,53	0,38	0,56	0,26
10	0,79	0,57	0,87	0,54	30	0,80	0,61	0,82	0,45
11	0,92	0,55	0,89	0,47	31	0,79	0,49	0,40	0,05
12	0,58	0,32	0,69	0,42	32	0,62	0,42	0,82	0,53
13	0,85	0,70	0,93	0,50	33	0,75	0,45	0,88	0,65
14	0,66	0,42	0,75	0,53	34	0,40	0,16	0,33	0,25
15	0,60	0,45	0,64	0,45	35	0,67	0,41	0,87	0,62
16	0,35	0,21	0,41	0,11	36	0,79	0,69	0,83	0,64
17	0,88	0,51	0,79	0,56	37	0,80	0,43	0,89	0,66
18	0,38	0,10	0,85	0,33	38	0,82	0,69	0,69	0,46
19	0,73	0,52	0,84	0,63	39	0,68	0,52	0,71	0,55
20	0,77	0,60	0,59	0,39	40	0,77	0,41	0,67	0,45

É possível perceber, de acordo com a Figura 11, uma leve tendência de correlação linear positiva entre a proporção de acertos de cada item em cada teste, mas que alguns itens não seguem essa tendência. Os itens 18 e 25 destacam-se por apresentarem uma proporção de acertos muito maior em um teste do que no outro, sendo que no item 18 o desempenho é melhor no teste digital e no item 25 é no UPSIT. Também foi possível perceber nos itens 31, 21 e 5 uma discrepância com a proporção de acertos no UPSIT

menor que a do digital.

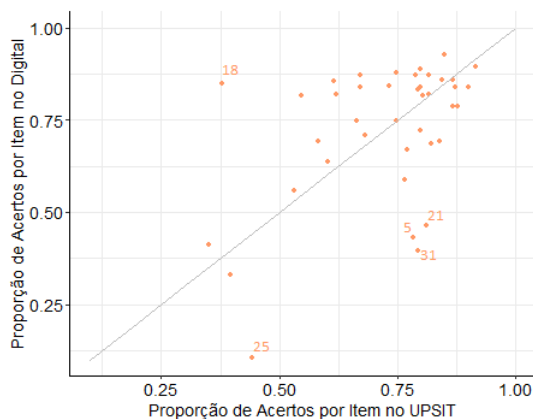


Figura 11: Comparação das proporções de acertos dos indivíduos por itens em cada um dos testes

Utilizando o estudo preliminar foi possível perceber um grande potencial no teste digital para mensuração do olfato, visto que os resultados mostraram, em geral, uma compatibilidade do teste digital com o UPSIT, principalmente em relação ao escore total. Apesar disso, os itens 25 e 31, referentes, respectivamente, aos aromas de picles e solvente, se destacaram dos demais por não apresentarem um resultado tão satisfatório tanto em relação à correlação com o escore total do teste (correlação ponto-bisserial), quanto à correlação com a proporção de acertos no UPSIT. Portanto, esses itens devem ser analisados com mais cautela e a retirada dessas questões pode ser considerada em uma nova versão do teste. Os itens referentes aos aromas de perfume e óleo de motor também apresentaram proporções de acertos bem mais baixas no teste digital do que no UPSIT, por isso qualidade dessas fragrâncias deve ser averiguada.

4 Conclusão

Os testes de mensuração do olfato são de extrema importância, pois esse sentido interfere em diversos aspectos da qualidade de vida de um indivíduo e a sua mensuração pode auxiliar na prevenção, no acompanhamento e no tratamento de diversas doenças. Os métodos psicofísicos, sobretudo, são relevantes por serem mais práticos e comercialmente mais acessíveis, apesar de não mensurarem diretamente o sentido. Nesse caso, é essencial avaliar as propriedades psicométricas dos instrumentos usados na mensuração da capacidade olfativa, já que ela pode ser vista como um traço latente. Neste trabalho, portanto, buscou-se, por meio da análise das propriedades psicométricas, fundamentar a utilização do Teste Olfatório da Universidade da Pensilvânia como referência dentro dos testes psicofísicos. Permitindo assim, explorar os resultados de um estudo preliminar de um novo procedimento digital para mensuração do olfato com base no UPSIT.

O estudo das propriedades psicométricas do UPSIT foi realizado a partir dos dados coletados na última etapa da validação do método para a população brasileira e os resultados apontaram evidências de validade e fidedignidade para o teste. Verificou-se, contudo, que o método apresenta questões centradas na mensuração de traços latentes inferiores, ou seja, esse instrumento tem uma capacidade maior em diferenciar os indivíduos com baixa capacidade olfativa. Considerando que o teste é utilizado, principalmente, para identificar e classificar pessoas com o olfato ruim, essa informação estaria de acordo com os objetivos. No entanto, o teste não se mostra tão adequado para identificar a perda olfatória no início, afim de facilitar o tratamento e evitar o agravamento dessa disfunção. Além disso, os itens 5, 19 e 34, relativos aos aromas de óleo de motor, pão de mel e madeira, respectivamente, se destacaram negativamente em diversos aspectos na validação do UPSIT e a exclusão ou reformulação desses itens poderia ser considerada em uma possível alteração do teste.

Considerando as evidências de validade do UPSIT, ele foi considerado como referência na análise do novo procedimento. Utilizando os dados coletados em um estudo preliminar para implantação do teste digital, foi possível perceber, em geral, uma compatibilidade do teste digital com o UPSIT, principalmente no que se refere ao escore total. Além de inúmeras vantagens operacionais do teste digital, foi possível perceber que, em geral, o procedimento apresentou duração menor que o UPSIT. No entanto, a reformulação ou exclusão dos itens 25 e 31, que representam, respectivamente, os aromas de picles e solvente, pode ser avaliada, visto que eles não apresentaram resultados tão satisfatório quanto os demais. Em razão do estudo de comparação do teste digital com o UPIST ter contado com um número reduzido de indivíduos, não foram utilizados os procedimentos de Análise Fatorial e Teoria da Resposta ao Item, apesar dessas técnicas serem importantes na avaliação das propriedades psicométricas de um instrumento.

Assim sendo, em termos gerais, o teste digital se mostra com grande potencial para mensuração do olfato, mas deve ser feito um estudo com amostra maior para validação desse método. Nesse estudo, pode ser analisada uma possibilidade de redução do número de questões, considerando as observações realizadas neste trabalho em relação aos itens com desempenho fora do esperado e da semelhança entre itens em relação à dificuldade e à discriminação no UPSIT. Além disso, podem ser buscadas modificações em questões existentes ou inserção de novos itens para aumentar o intervalo em que são observados bons resultados na mensuração, além das capacidades olfativas inferiores.

5 Referências

- ANDRADE, D. F. de; TAVARES, H. R.; VALLE, R. da C. Teoria da resposta ao item: conceitos e aplicações. *ABE, Sao Paulo*, 2000.
- ANJOS, A. dos; ANDRADE, D. F. de. Teoria da resposta ao item com uso do r.
- BAKER, F. B. *The basics of item response theory*. [S.l.]: ERIC, 2001.
- BAKER, F. B.; KIM, S.-H. *The basics of item response theory using R*. [S.l.]: Springer, 2017.
- BASTOS, L. O. D. et al. Acurácia dos testes de identificação de odores no diagnóstico da doença de parkinson. [sn], 2019.
- BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, Sage Publications Sage CA: Thousand Oaks, CA, v. 33, n. 2, p. 261–304, 2004.
- CRONBACH, L. J.; MEEHL, P. E. Construct validity in psychological tests. *Psychological bulletin*, v. 52 4, p. 281–302, 1955.
- DAMÁSIO, B. F. Uso da análise fatorial exploratória em psicologia. *Avaliação Psicológica: Interamerican Journal of Psychological Assessment*, Instituto Brasileiro de Avaliação Psicológica (IBAP), v. 11, n. 2, p. 213–228, 2012.
- DOTY, R. L. Studies of human olfaction from the university of pennsylvania smell and taste center. *Chemical senses*, Oxford University Press, v. 22, n. 5, p. 565–586, 1997.
- ERTHAL, T. C. S. *Manual de psicometria*. [S.l.]: Zahar, 2010.
- FORNAZIERI, M. A. Validação do teste de identificação do olfato da universidade da pensilvânia (upsit) para brasileiros. 2013.
- GUYTON, A. C.; HALL, J. E. et al. *Tratado de Fisiologia Médica (9ª Edição)*. [S.l.]: Ed. Guanabara Koogan: Rio de Janeiro, 1997.
- HUTZ, C. S.; BANDEIRA, D. R.; TRENTINI, C. M. *Psicometria*. [S.l.]: Artmed Editora, 2015.
- JOHNSON, R.; WICHERN, D. Applied multivariate statistical analysis.. prenticehall international. *INC., New Jersey*, 2007.
- NAKANISHI, M. et al. The digital scent device as a new concept for olfactory assessment. In: WILEY ONLINE LIBRARY. *International Forum of Allergy & Rhinology*. [S.l.], 2022.
- PASQUALI, L. *Psicometria: teoria dos testes na psicologia e na educação*. [S.l.]: Editora Vozes Limitada, 2017.
- PASQUALI, L.; PRIMI, R. Fundamentos da teoria da resposta ao item: Tri. *Avaliação Psicológica: Interamerican Journal of Psychological Assessment*, Instituto Brasileiro de Avaliação Psicológica (IBAP), v. 2, n. 2, p. 99–110, 2003.

- PEREIRA, I. I. d. C. *Relação entre depressão e perda de capacidade olfativa*. Tese (Doutorado), 2020.
- PRIMI, R. Psicometria: fundamentos matemáticos da teoria clássica dos testes. *Avaliação Psicológica*, Instituto Brasileiro de Avaliação Psicológica, v. 11, n. 2, p. 297–307, 2012.
- RIZOPOULOS, D. ltm: An r package for latent variable modeling and item response theory analyses. *Journal of statistical software*, v. 17, n. 5, p. 1–25, 2006.
- SARTES, L. M. A.; SOUZA-FORMIGONI, M. L. O. d. Avanços na psicometria: da teoria clássica dos testes à teoria de resposta ao item. *Psicologia: Reflexão e Crítica*, SciELO Brasil, v. 26, p. 241–250, 2013.
- SEDAGHAT, A. R.; GENGLER, I.; SPETH, M. M. Olfactory dysfunction: a highly prevalent symptom of covid-19 with public health significance. *Otolaryngology–Head and Neck Surgery*, SAGE Publications Sage CA: Los Angeles, CA, v. 163, n. 1, p. 12–15, 2020.
- VIEIRA, M. M. C. d. S. *A doença de Alzheimer e a perda de olfato*. Tese (Doutorado), 2020.

6 Apêndice

6.1 Detalhes dos itens do UPSIT

Quadro 2: Opções de resposta correspondentes a cada um dos itens do UPSIT

Item	Alternativas	Item	Alternativas
1	A. Gasolina B. Pizza C. Amendoim D. Flor	21	A. Perfume B. Cravo C. Gasolina D. Fumaça
2	A. Pepino B. Chiclete C. Nozes D. Melancia	22	A. Flor B. Pão de Mel C. Maçã D. Morango
3	A. Tomate B. Gasolina C. Morango D. Mentol	23	A. Chocolate B. Pêssego C. Couro D. Pizza
4	A. Peixe B. Limão C. Alho D. Cereja	24	A. Pneu B. Melancia C. Banana D. Pão de Mel
5	A. Grama B. Pizza C. Óleo de Motor D. Abacaxi	25	A. Abacaxi B. Picles C. Melancia D. Flor
6	A. Cachorro B. Menta C. Peixe D. Refrigerante de Cola	26	A. Fumaça B. Cerveja C. Abacaxi D. Cebola
7	A. Banana B. Alho C. Cereja D. Óleo de Motor	27	A. Refrigerante de Cola B. Alho C. Solvente D. Framboesa
8	A. Talco de Bebê B. Cravo	28	A. Queijo B. Laranja

Continua

Continuação

Item	Alternativas	Item	Alternativas
8	C. Espaguete D. Banana	28	C. Espaguete D. Solvente
9	A. Cravo B. Flor C. Couro D. Maçã	29	A. Framboesa B. Nozes C. Peixe D. Chiclete
10	A. Cachorro B. Coco C. Tomate D. Mel	30	A. Cerveja B. Mentol C. Laranja D. Melancia
11	A. Chocolate B. Banana C. Cebola D. Pêssego	31	A. Melancia B. Amendoim C. Rosa D. Solvente
12	A. Couro B. Suco de Frutas C. Mentol D. Alho	32	A. Talco de Bebê B. Maçã C. Grama D. Morango
13	A. Talco de Bebê B. Abacaxi C. Queijo D. Cereja	33	A. Pepino B. Grama C. Fumaça D. Pêssego
14	A. Solvente B. Pneu C. Coco D. Jasmim	34	A. Madeira B. Couro C. Queijo D. Pneu
15	A. Óleo de Motor B. Canela C. Abacaxi D. Talco de Bebê	35	A. Pizza B. Solvente C. Cravo D. Uva
16	A. Rosa B. Alho C. Pêssego D. Gasolina	36	A. Óleo de Motor B. Baunilha C. Rosa D. Alho
17	A. Morango B. Picles	37	A. Sabão B. Pimenta

Continua

Continuação

Item	Alternativas	Item	Alternativas
17	C. Pipoca D. Gasolina	37	C. Laranja D. Amendoim
18	A. Café B. Chiclete C. Alho D. Cravo	38	A. Laranja B. Perfume C. Refrigerante de Cola D. Gás Natural
19	A. Alho B. Pão de Mel C. Pneu D. Pimenta	39	A. Baunilha B. Rosa C. Menta D. Chiclete
20	A. Mentol B. Talco de Bebê C. Maçã D. Queijo	40	A. Amendoim B. Alho C. Maçã D. Pneu

6.2 Curva Característica dos Itens

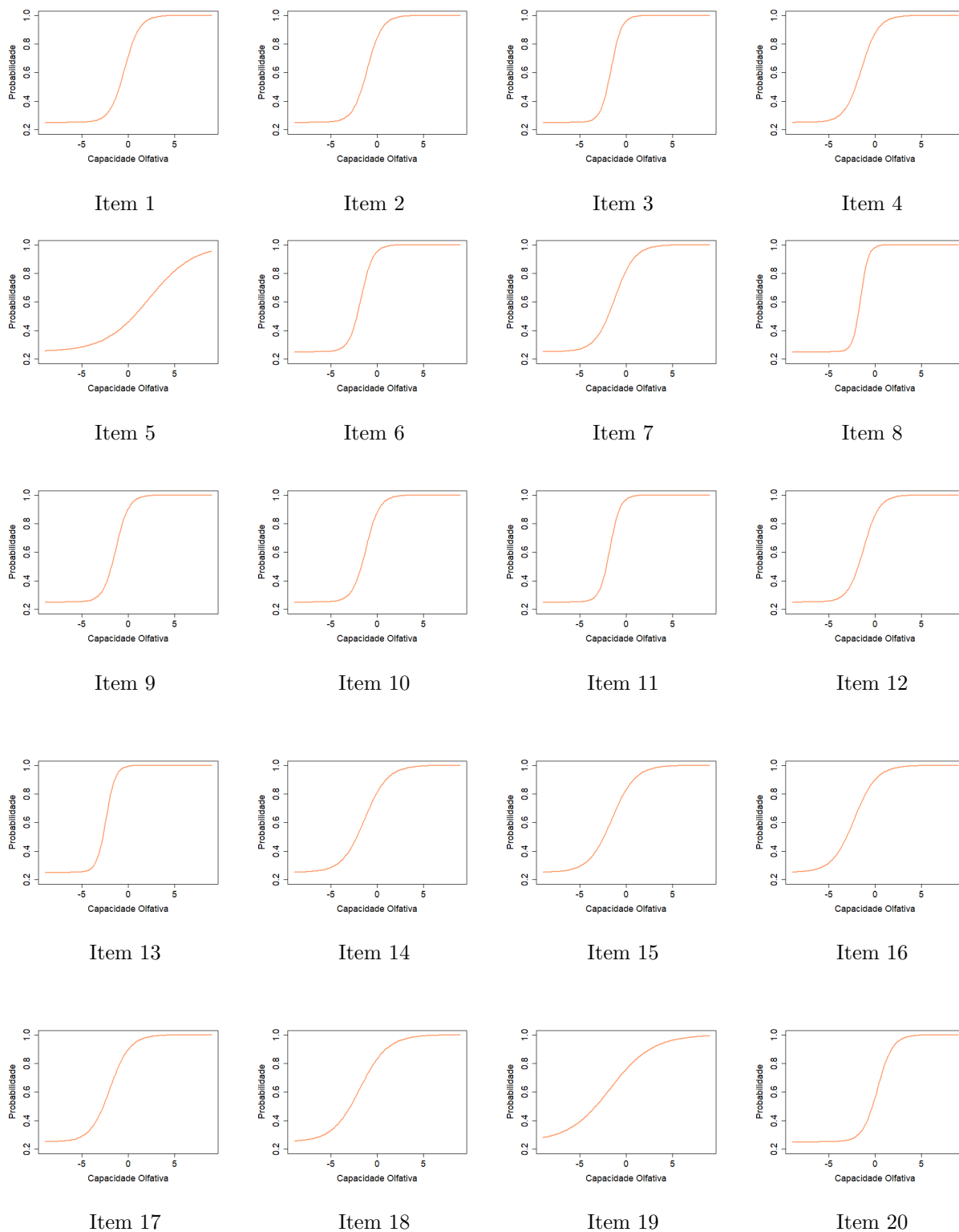
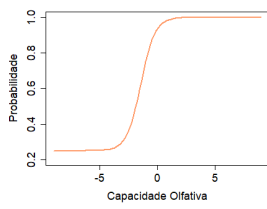
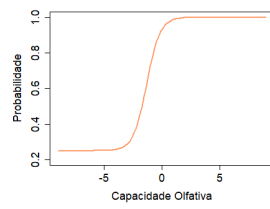


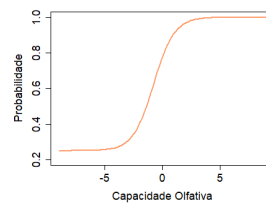
Figura 12: Curvas características dos itens 1 a 20



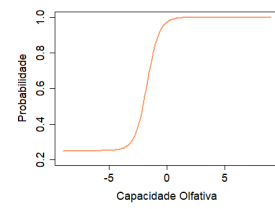
Item 21



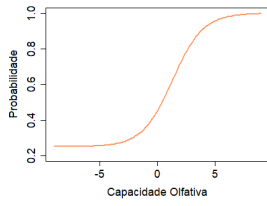
Item 22



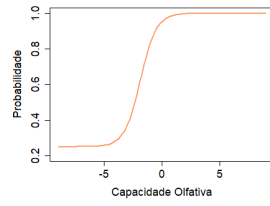
Item 23



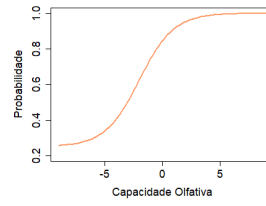
Item 24



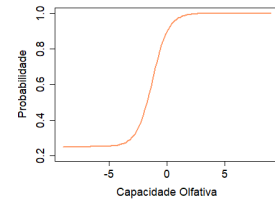
Item 25



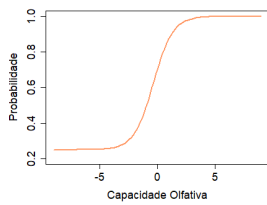
Item 26



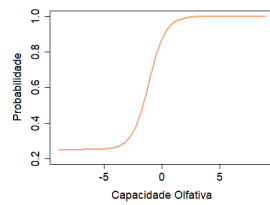
Item 27



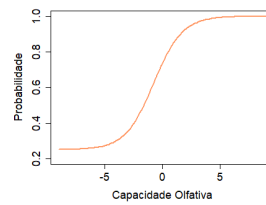
Item 28



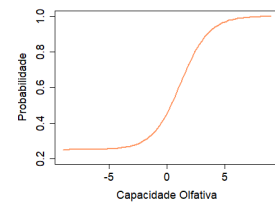
Item 29



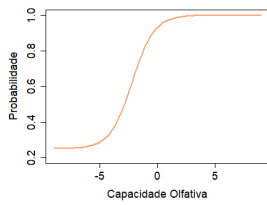
Item 30



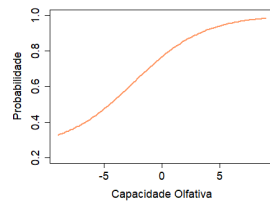
Item 31



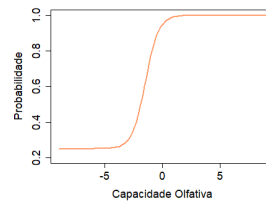
Item 32



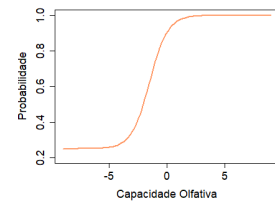
Item 33



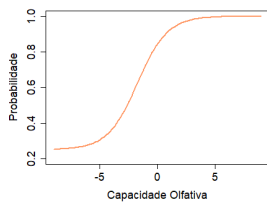
Item 34



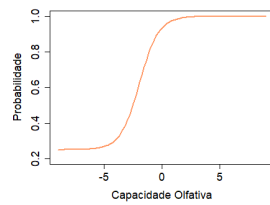
Item 35



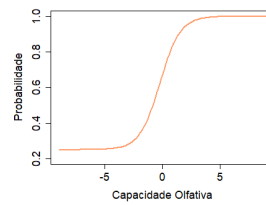
Item 36



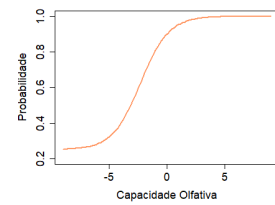
Item 37



Item 38



Item 39



Item 40

Figura 13: Curvas características dos itens 21 a 40

6.3 Curva de Informação dos Itens

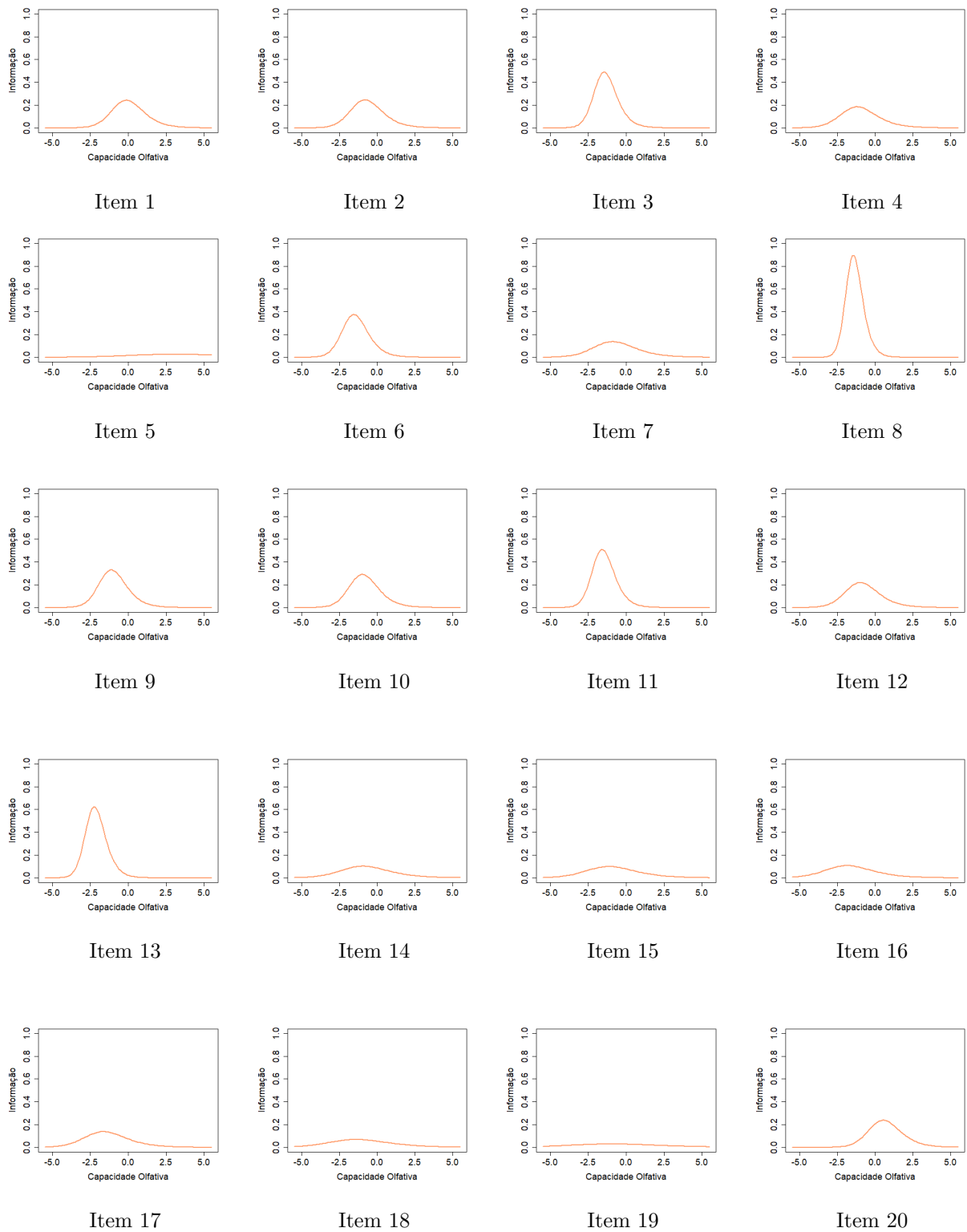


Figura 14: Curvas de informação dos itens 1 a 20

