3
Análises de propriedades psicométricas

A Psicometria é o campo da Psicologia que busca analisar as características matemáticas constantes dos dados empíricos. A intersecção com as Ciências Estatísticas é típica dos estudos em Psicometria.Os indicadores gerados por diversas funções matemáticas podem revelar informações relevantes sobre os dados obtidos. O objetivo final da Psicometria é a busca do invariante, isto é, elementos característicos de uma variável psicológica que não mudam de sujeito para sujeito. Destarte, é universal.

Erthal (2001) afirma que a Psicometria é um grupo de técnicas que viabiliza a quantificação de fenômenos psíquicos. Medir consiste na atribuição de magnitudes através de valores numéricos a um objeto ou classe de objetos. Dessa forma, a medida apresenta características de quantificação, unidades relativamente constantes e um ponto de referência como marco inicial da medida.

Na física, o zero absoluto é usado como referência da medida – por exemplo, 0 Kelvin da temperatura (zero absoluto) ou 0 m/s da velocidade (estado físico de ausência de movimento). Na Psicologia, inteligência, emoção ou aptidão não possuem um ponto de zero absoluto. Logo, a solução dos pesquisadores foi atribuir um ponto de referência arbitrário, como as medidas de tendência central moda, média e mediana (Erthal, 2001).

Uma vez definido um ponto de referência e um modo de mensurar um traço psicológico, precisamos compreender o conceito de variabilidade e erro. Segundo Erthal (2001), toda medida, por mais cuidado que se tenha ao extraí-la, tem inúmeros fatores que a influenciam e alteram seus resultados, tornando-a menos confiável. Por essa razão, um instrumento proporciona resultados individuais diversos. A tal diversidade dá-se o nome de variabilidade.

Uma parte dessas variações é devida às características da própria medida. Mas outra importante parcela deve-se a erros cometidos ao longo do processo de mensuração. Por essa razão, erros da observação, erros do instrumento ou erros devidos à falta de uniformidade na mensuração são falhas que precisam ser conhecidas pelo pesquisador a fim de diminuí-las ou extingui-las (Erthal, 2001).

Com este objetivo, a psicologia vale-se de diversos métodos estatísticos para mensurar o erro e, para isso, utiliza medidas de variabilidade em função da tendência central, como o percentil ou o desvio padrão. O cálculo do erro é possível através de medidas de confiabilidade como o erro padrão da média (EPM) que é dado pelo teorema de Bernoulli (Pasquali, 2009) na expressão matemática:

$$\varepsilon = \frac{\sigma}{\sqrt{n}}$$

Nessa equação, o erro ε é dado pelo desvio padrão σ de uma média μ em uma amostra n. A inferência que se extrai desta função matemática é que, quanto maior o número de vezes uma medida é usada, maior a certeza que se tem sobre as características do instrumento. Em outras palavras, quanto maior a amostra, menor a variabilidade em função dos erros da observação ou da falta de uniformidade na mensuração. O erro que é extraído em uma grande amostra ocorrerá muito mais em função da variabilidade da medida no instrumento do que por erros no método de mensuração.

A medida psicométrica possui o pressuposto da validação para tornar-se confiável. Portanto, validar um instrumento constitui estudar duas características principais: validade e fidedignidade (Pasquali, 2009).

A validade consiste em conhecer se o instrumento realmente mede aquilo a que se propõe. Por exemplo, atesta se uma escala de ansiedade de fato está medindo ansiedade. Diversos métodos são utilizados para se conhecer a validade de uma medida. Dois dos métodos mais utilizados são: (1) a validade teórica e (2) a validade de construto convergente e divergente. Um teste pode ser validado de modo teórico quando especialistas na área em questão se reúnem e concordam que aquele instrumento mede o que está sendo proposto. A validade de construto convergente e divergente utiliza respectivamente instrumentos que já são usados para mensurar dado construto psicológico ou que não possuem este uso e correlaciona-os com a medida que se pretende avaliar (Erthal, 2001).

Fidedignidade é o índice de precisão da medida, ou seja, o quanto a medida está susceptível ao erro. Quanto maior a fidedignidade, maior a precisão de um instrumento. Duas formas básicas de calcular a fidedignidade são através dos coeficientes de confiabilidade com base: (1) nos teoremas da correlação de Pearson e Spearman, ou (2) no teorema de probabilidade de Bayes.

O maior problema é que a medida psicológica Ψ não tem como ser acessada diretamente. Ademais, as medidas em Psicologia são quantificadas de modo indireto. Mais especificamente, quantifica-se uma medida de natureza física θ, constituída pelas respostas do sujeito a um instrumento, considerando sua variância σ^2 como referência para o cálculo do erro ε da medida psicológica, com base no teorema de Bernoulli (Pasquali, 2009). A expressão matemática da medida psicométrica é expressa pela função geral:

$$\Psi = \theta + \frac{\sigma}{\sqrt{n}}$$

Toda medida psicológica, portanto, passa por uma variável física que considera o erro em função da variância dos dados empíricos. Uma escala de autopreenchimento, um dado coletado em galvanômetro ou tempo de resposta manual são forma diferentes de coleta de dados físicos que serão interpretados do ponto de vista de um construto psíquico.

Segundo Pasquali (2009), há dois métodos contemporâneos de construir e verificar medidas psicométricas. Tais métodos são complementares e conhecidos como Teoria Clássica de Testes (TCT) e Teoria de Resposta ao Item (TRI).Cada uma baseia-se em um dos teoremas de fidedignidade supracitados: Pearson e Bayes, respectivamente. Investigaremos os diferentes métodos a fim de compreender como são utilizados de modo complementar nos estudos de propriedades psicométricas contemporâneos.

3.1
Teoria Clássica dos Testes (TCT)

O conceito estruturante da Teoria Clássica dos Testes (TCT) é a possibilidade de vislumbrar uma escala ou grupo de itens como partes iguais de um construto psicológico. Cronbach (1951) propõe, inicialmente, que uma escala não deve apresentar itens monotônicos, mas deve ser constituída por itens que colaboram com o escore final na mesma proporção. Baseado nesse princípio, o autor cria um coeficiente alfa α que é a medida da relação entre os escores obtidos a partir dos dados empíricos nas respostas de um grupo de pessoas e um grupo de itens e seus escores reais. O alfa de Cronbach, portanto, é utilizado para verificar a confiabilidade de uma escala em função de seus escores obtidos empiricamente e seus escores reais.

Tal confiabilidade, ou homogeneidade, como afirma o autor em seu trabalho de 1951, serve para conhecer o quanto este grupo de itens contribui para o escore total (Cronbach, 1969). A função matemática que descreve o alfa de Cronbach é dada por:

$$\alpha = \frac{n}{n - 1}\left(1 - \frac{\sum_{i=1}^{n} \sigma^2_{\mathrm{i}}}{\sigma^2_{\mathrm{sum}}}\right)$$

Nessa equação, o coeficiente α é dado quando um número *i* de pessoas responde a um instrumento de *n* itens, que são calculados em função da variância σ^2_{i} entre os dados obtidos nas respostas e a soma da variância σ^2_{sum} dos itens. De acordo com a TCT, esse coeficiente tem singular utilidade, uma vez que valores abaixo de 0.70 mostram uma escala inconsistente e pouco confiável. Entretanto, valores entre $0.50 \leq \alpha \leq 0.70$ podem ser aceitos em escalas com poucos itens, uma vez que o valor do coeficiente depende proporcionalmente do número de itens.

Anteriormente falamos que a TCT possui base estruturante na correlação de Pearson (*r*) em dados paramétricos e Spearman (*rho*) em dados não-paramétricos. Partindo do pressuposto de Cronbach (1951, 1969) e de outros autores em psicometria que estudam a TCT, como Cohen, Swerdlik, e Douglas (1992), Erthal (2001), e Pasquali (2009), de que os itens de uma escala contribuem de modo equivalente para o escore total, a correlação entre o item e o total obtido na escala é uma ferramenta fundamental no estudo das características psicométricas de uma escala. O índice de correlação item-total *r* mostra o quanto o item está contribuindo para o escore final. De modo geral, o valor aceito é de $r > 0.30$.

A TCT contribui de modo singular para a compreensão das propriedades psicométricas através de diversos coeficientes. No presente estudo, porém, ater-nos-emos somente ao coeficiente alfa de Cronbach (Cronbach, 1951) e à correlação item-total (Cohen, Swerdlik, e Douglas, 1992, Cronbach, 1969, Erthal, 2001, Pasquali, 2009).

3.2
Análise Fatorial

Tucker e McCallum (1997) afirmam que o método de análise fatorial é o estudo estruturado de dados multivariados que inclui tanto a teoria quanto construtos e dinâmicas subjacentes cujos resultados empíricos dão gênese ao fenômeno observado. Basicamente, há dois métodos procedimentais para a extração das dimensões de dados empíricos: Análise Fatorial Exploratória (AFE) e Análise Fatorial Confirmatória (AFC). Neste trabalho, ater-nos-emos somente à AFE.

No procedimento da AFE, cada superfície ou variável diferente pressupõe alta correlação com um atributo interno ou latente disponível para explicar a variância encontrada no modelo (Long, 1983, Tucker e McCallum, 1997). Em outras palavras, num grupo de itens que se propõe avaliar diferentes construtos psicológicos, cada subgrupo a examinar um construto específico terá maior correlação entre si que com outros itens do grupo maior. Essa correlação deve explicar a variância contida no subgrupo de itens. Caso contrário, deve-se recorrer novamente ao grupo para encontrar novos itens que melhor se correlacionem e expliquem a variância.

Tucker e McCallum (1997) sintetizam que o objetivo básico da AFE é a redução dos dados empíricos, quando um grupo grande de covariâncias pode ser explicado por um número menor de atributos internos ao determinar sua natureza e padrão de influência nas variáveis latentes.

A Análise de Componentes Principais (ACP) é o tipo de fatoramento que extrai o máximo da variância obtida através de uma combinação linear das covariâncias (Hair et al., 2009). Nesta, cada fator é contado com a variância restante dos dados na ordem da extração. Com base na ACP, há diversos procedimentos para identificar o melhor modelo fatorial exploratório (Hair et al., 2009).

O primeiro critério utilizado na análise fatorial, normalmente, é verificar o valor do coeficiente Kaiser-Meyer-Olkin (KMO). Esse coeficiente verifica se há autovalores superiores a 1. Em outras palavras, se a covariância de um subgrupo de dados puder ser explicada pelo fator, o autovalor desse subgrupo será maior que 1.

O critério da comunalidade também deve adotar os valores maiores que 0.50 para cada item. Nesse caso, a covariância dos itens deve estar fortemente correlacionada para mostrar as superfícies em comum.

O critério mais conhecido na AFE, entretanto, é o estudo da porcentagem da variância explicada pelo fator. Para que uma escala seja considerada unidimensional, o valor da variância explicada pelos dados empíricos dos itens deve ser maior que 55% e a diferença para o segundo fator no quadrado das variâncias deve ser menor que 5% (Linacre, 2009).

No caso do ASQ-BR, utilizamos diferentes métodos baseados em ACP. Esses métodos serão aprofundados posteriormente neste trabalho. De qualquer forma, porém, todos pressupõem estudar a unidimensionalidade das escalas do instrumento considerando cada grupo de 6 perguntas referentes a cada um dos cinco domínios do desenvolvimento uma escala referente a uma faixa etária.

3.3 Teoria de Resposta ao Item Não-paramétrica (NIRT): Escalonamento de Mokken

O conceito de escalonamento representado matematicamente nas análises estatísticas de Mokken (Mokken, 1971) nasce antes do primeiro artigo de Mokken em 1971, acerca de um método de construção de escalas centradas no indivíduo. A idéia de escalonar um grupo de pessoas e itens em uma escala ordinal unidimensional inicia-se com a proposição de Guttman de que nenhum item é idêntico ao outro e de que sempre estarão ranqueados em alguma ordem hierárquica. Esta pode ser que pode ser: dificuldade, qualidade ou quantidade da presença de um traço latente (Guttman, 1944). Portanto, quaisquer constructos psicológicos poderiam ser escalonados a partir de itens que revelam menos ou mais do traço contido em uma pessoa.

Do mesmo modo, tanto itens quanto pessoas podem ser ordenadas. A partir desta premissa, pode-se inferir que a probabilidade de uma pessoa respondendo a um item difícil corretamente acertar um item mais fácil é maior do que uma pessoa que não respondeu corretamente ao item difícil. Na mesma medida, uma pessoa que responde positivamente a um item que revela mais de um traço psicológico, tenderá a responder positivamente a um item que dê menos informação sobre aquele constructo do que uma pessoa que respondeu negativamente ao item mais revelador.

As premissas supracitadas são essenciais para compreender dois alicerces da Teoria de Resposta ao Item (TRI): probabilidade da resposta e diferença entre itens e pessoas. Os modelos de TRI apresentam a vantagem de permitir a comparação entre um modelo matematicamente calculado e os dados empíricos obtidos pelo pesquisador. Em outras palavras, há um ideal matemático a ser atingido (Hambleton, Sireci e Zumbo, 2011). A partir dessa idéia, Loevinger (1948) propôs um procedimento estatístico para comparar dados empiricamente inconsistentes com a propositura teórica de Guttman. O resultado desse modelo passou a ser conhecido como *H* de Loevinger, cuja função matemática é capaz de descrever a capacidade geral de monotonicidade de uma escala, assim como o escalonamento de um item.

Em 1971, Mokken propôs o uso do coeficiente *H* como critério para boas escalas baseadas em traços latentes. A extensão de um traço é considerada a probabilidade da resposta certa a um item em comparação do escore total da escala (Sijtsma, Debets e Molenaar, 1990). Essa probabilidade tende a um crescimento linear proporcional ao traço latente.

Por exemplo, se o item i é mais fácil que o item j, o padrão de resposta da escala deve ser descrita pela equação $j = i + 1$. Uma boa escala possui monotonicidade constante, em outras palavras, a dificuldade cresce de forma monotônica, não permitindo dupla monotonicidade. Sendo a monotonicidade descrita pela diferença igual entre as probabilidades de acerto dos itens, uma escala monotônica é aquela que mantém o padrão $j = i + 1$ ao longo de toda a escala para todos os itens. Se imaginarmos que uma escala possui três itens (i, j, l) e é idealmente monotônica, a ordem de dificuldade dos itens é considerada pelo item i como sendo o mais fácil, o item j intermediário e o item l o mais difícil. Consequentemente, as expressões matemáticas que descrevem os padrões de resposta desta escala devem ser: $l = j + 1$; $j = i + 1$.

A limitação do trabalho de Mokken está na análise de items dicotômicos. Há, portanto, apenas dois graus de discriminação: certo e errado. Para sanar este problema, Molenaar (1982) propôs um modelo para inferir a monotonicidade de uma escala de itens politômicos. Ambos os aspectos são descritos pela função matemática:

$$Ht = 1 - \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} O_{ij} \Bigg/ \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} E_{ij}$$

Nessa equação, *Ht* é o coeficiente de homogeneidade de uma escala com *k* itens baseados na relação entre a soma dos erros observados (O_{ij}) e a soma de erros esperada (E_{ij}) entre dois itens *i* e *j*. Um erro ocorre quando o item *i* é mais fácil que o item *j*, mas a pessoa responde corretamente ao item *j*. Entretanto, erra o item *i*. O valor máximo do coeficiente *H* é igual a 1 e não há valores mínimos. A soma dos erros também é conhecida como erros de Guttman, pois, são empiricamente verificados com base na proposição de Guttman para escalonamento de traços latentes (Guttman, 1944, Molenaar, 1982). De acordo com Sijtsma et al. (1990), uma escala boa ou forte apresenta $H \geq 0.50$, uma escala média ou medíocre tem o valor entre: $0.40 \leq H < 0.50$, e uma escala ruim ou fraca entre: $0.30 \leq H < 0.40$.

A análise do coeficiente *H* de cada item segue uma regra um pouco diferente. Mokken e Lewis (1982) propuseram, primeiramente, identificar os dois itens com maior *H* de Loevinger. Então, são acrescentados os itens restantesum aum, seguindo o princípio de escalonamento de coeficientes *H*, do maior para o menor e até o fim, quando $H < 0.30$.

3.3.1
Escalonamento de Itens Politômicos: A noção de Gradação do Item

Itens politômicos são fundamentados pela noção de gradação do item. O exemplo dado por Sijtsma et al. (1990) para ilustrar a gradação do item é simples. Um dado item de uma escala de atitudes apresenta respostas ordenadas em Likert com três níveis de concordância: (1) discordo, (2) neutro, (3) concordo. Os intervalos entre as categorias são gradações segundo a noção de Sijtsma et al. (1990). Dessa forma, para construir um item com respostas politômicas que apresentam três graus de discriminação, as distâncias entre as categorias são consideradas gradações, ou seja, há duas gradações, cada uma com dois níveis. Isso transforma um item politômico em dois itens dicotômicos seguindo o procedimento de resposta descrito a seguir.

Inicialmente, a pessoa escolhe entre (1) ‘discordo’ e (2) ‘neutro’. Se a resposta for negativa, a escolha será ‘discordo’ e se a resposta for positiva, a resposta será ‘neutro’. Uma vez que (2) ‘neutro’ tiver sido escolhido, a próxima gradação se iniciará: se a resposta da pessoa for negativa, (2) ‘neutro’ permanecerá. Se for positiva, a pessoa responderá (3) ‘concordo’.

Consequentemente, a estrutura ordinal da resposta para itens politômicos é mantida e descrita pela regra: uma categoria e está relacionada à sua próxima categoria f pela relação $e = f + 1$. Uma vez que a categoria f é escolhida, a notação da próxima gradação passa a ser $f + 1 = g + 2$.

Essas relações permitem inferir um procedimento estrutural para a gradação do item, onde uma gradação anterior p é igual à próxima gradação n mais 1: $p = n + 1$. A noção de gradação do item, então, viabiliza o uso do H de Loevinger em escalas de itens politômicos, permitindo acessar os valores dos coeficientes de itens e escalas.

Todavia, uma questão ainda se impunha para a compreensão da probabilidade de respostas dentro do escalonamento de Guttman: quanto se pode confiar nos dados empíricos para compará-los ao modelo matemático descrito por Loevinger?

Nesse sentido, Molenaar e Sijtsma (1988) propuseram um coeficiente de confiabilidade *MS* para compreender a capacidade da escala em ordenar corretamente os itens politômicos de uma escala cujos dados foram empiricamente obtidos. Esse coeficiente verifica a confiabilidade e dupla monotonicidade não desejada na escala, permitindo tomar decisões sobre itens e verificar o quanto se pode confiar na escalabilidade de um grupo de itens (Sijtsma et al., 1990).

A análise de Mokken, portanto, apresenta importantes contribuições para a noção de escalonamento: itens são apresentados em um grupo hierárquico único, o modelo proporciona encontrar índices de dificuldade e discriminação dos itens, e, além disso, fundamenta a eliminação de itens que não se encaixam nos critérios estabelecidos.

Para a aplicabilidade da análise de Mokken, no entanto, um pressuposto deve ser sempre verificado previamente: a unidimensionalidade da escala (Molenaar, 1982, Sijtsma et al., 1990). A análise fatorial é o procedimento estatístico utilizado para identificar a dimensionalidade de um grupo de itens. Teoricamente, se itens representam o mesmo construto psicológico, devem também apresentar correlações altas e, consequentemente, ficar estruturados sob uma mesma dimensão ou fator.

Uma série de métodos é utilizada para detectar o número de fatores de um grupo de dados empíricos. Porém, a mais recomendada para obter a dimensionalidade dentro da análise de Mokken é a Análise Fatorial Exploratória (AFE) baseada na Análise de Componentes Principais (ACP). Os dados são considerados unidimensionais quando o primeiro fator explica mais que 55% da Variância encontrada e a diferença entre a Variância explicada pelo primeiro fator e a Variância do fator seguinte é menor que 5% (Sijtsma e Verweij, 1992).

No presente estudo, utilizamos o *software Mokken Scaling Procedure* (MSP; Sijtsma et al., 1990) para mensurar três indicadores das medidas de NIRT: (1) coeficiente de confiabilidade *rho* ou *MS* (Molenaar e Sijtsma, 1988), (2) *H* de Loevinger para a escala, e (3) *H* de Loevinger para cada item.

O índice *rho* detecta o quanto a amostra responde dentro do esperado dada a probabilidade calculada a partir do modelo matemático, comparando-a com os dados empíricos. Os valores de *MS* devem ser maiores que 0.70. Os valores de *H*, como previamente elucidado, revelam características de força da escala: uma boa escala apresenta $H \geq 0.50$, uma escala média tem o valor entre $0.40 \leq H < 0.50$, e uma escala ruim, entre $0.30 \leq H < 0.40$. Por outro lado, para os itens específicos, os valores de *H* devem ser maiores que 0.30.

3.4
Modelo de Análise de Medidas de Rasch

3.4.1
Análises Univariadas

Análises estatísticas univariadas, isto é, de apenas uma variável dependente, são classicamente usadas para compreender diferentes construtos da Psicologia, estudando-os como um fator geral com diversas subdimensões. Essa divisão intrínseca dos construtos permite usar diversas técnicas de inferência e análise estatísticas univariadas, sem perder a capacidade explanatória do construto.

Hair et al. (2009) afirmam que, muitas vezes, o melhor modelo estatístico não é o mais simples.Entretanto, seu ganho pode ser muito pequeno em relação a um modelo mais simples de ser mensurado. Este, por sua vez, pode apresentar grande capacidade explanatória. O modelo mais simples para explicar uma variável pode ser o mais eficaz: o Princípio da Parcimônia.

Podemos exemplificar o Princípio da Parcimônia no presente estudo, com o ASQ-3. O módulo Resolução de Problemas possui diversas dimensões. O conceito deste construto, segundo Squires et al. (2009) consiste na capacidade de criança resolver novos problemas. Tal habilidade exige, principalmente, o processamento de informações através do córtex pré-frontal (Diamond, 2011) demandando diferentes funções cognitivas: (1) controle inibitório, (2) flexibilidade de pensamento e (3) memória de trabalho.

Se a dimensão Resolução de Problemas do ASQ-3 está basicamente dividida em três habilidades executivas, pode-se imaginar que ao realizar-se uma análise fatorial das seis perguntas deste fator, três dimensões seriam contempladas. Diante dessa possibilidade, podemos indagar se o melhor ajuste terá as três dimensões independentes, correlacionadas, ou três fatores com uma quarta dimensão geral de segunda ordem.

Myiake e seus colaboradores (2000) usaram cinco conhecidos procedimentos para avaliar funções executivas tentando responder a esta pergunta: Geração Aleatória de Números (GAN), Torre de Hanoi, *Span* operatório, tarefas duplas e *Wisconsin Sorting Card Test* (WSCT).

Os autores modelaram as três soluções por Equação Estrutural com base nos escores dos testes e demonstraram que a resposta para essa questão residia no Princípio da Parcimônia. Esse princípio consiste em que a melhor análise considera tanto o primeiro como o segundo modelo. Apesar de o modelo com ajuste de três dimensões correlacionadas ter apresentado valor de χ^2 e *Goodness of Fit Index* (GFI) melhor que o modelo de quatro dimensões com segunda ordem, ao serem considerados isoladamente apenas os escores de qualquer teste como fonte dos dados para a modelagem, o modelo de um fator geral de segunda ordem geravam mais informações que o modelo de três fatores.

Cada teste carregou em um diferente fator, mas manteve os resíduos distribuídos pelos demais, o que reforçou a hipótese de um fator geral em segunda ordem. Portanto, o exemplo de Myiake et al. (2000) mostra que o Princípio da Parcimônia deve ser usado a cada momento em análise de dados, pois nem sempre o pesquisador poderá apoiar-se em uma solução matematicamente ideal. Deverá, entretanto, respeitar as particularidades dos dados empiricamente obtidos.

As soluções univariadas que consideram uma dimensão para o construto psicológico são as mais simples. Mas, como visto no exemplo de Myiake et al. (2000), têm enorme utilidade em pesquisas que analisam variáveis psicológicas que a literatura consistentemente aponta como uma unidade fatorial.

O ASQ-3 apresenta cinco medidas diferentes, separando-as em cinco dimensões que claramente possuem subdivisões intrínsecas. Porém, ao seguir o Princípio da Parcimônia, podemos considerá-las dimensões únicas e separá-las apenas em cinco sub-escalas, como o faz o próprio instrumento. A partir dessa divisão, a análise dos itens deve considerar os escores das seis questões de cada subescala como univariada.

3.4.2
Teoria de Resposta ao Item Univariada em Modelo de Créditos Parciais

A Teoria de Resposta ao Item (TRI) busca basicamente avaliar o impacto de dois parâmetros de uma curva de habilidades ou traços latentes que constituem uma medida: os itens e as pessoas que respondem ao instrumento. A Análise de Rasch (Rasch, 1961) é uma abordagem em que os escores empiricamente obtidos para cada um dos itens observados são transformados em dados ordinais e, posteriormente, em medidas lineares que representam um traço ou habilidade latente. A invariante da medida é o elemento buscado no modelo de Rasch (Olsen et al., 2010). Nesse paradigma, a invariância considera que os escores para pessoas devem representar uma função para itens específicos e que pessoas específicas dentro da amostra não devem influenciar os resultados de um item.

No modelo Rasch, pessoas e itens estão alocados numa mesma escala cuja unidade recebe diversos nomes: *logit*, *theta* (θ) e medida logística. Todavia, essa unidade representa o coeficiente médio de respostas para pessoas e itens ao longo de uma curva de características do item (CCI).

Rasch (1961) propôs um modelo matemático ideal para habilidades ou traços latentes de um construto. Segundo esse ideal matemático, os itens mais fáceis ou que revelam menos de um traço latente tendem a apresentar altos índices de acerto ou concordância ao longo de uma medida. Quando o item fica mais difícil ou revela mais de um traço latente, menos pessoas acertam ou concordam com o item. Tais itens são, portanto, mais discriminativos e permitem diferenciar melhor as pessoas mais proficientes ou com maior quantidade do traço latente ao longo da curva.

O modelo de Créditos Parciais (Andrich, 1978, Linacre, 2005, Masters, 1982) é o mais utilizado em relação às possíveis análises de Rasch por tratar-se de um modelo que considera o dado obtido das respostas a um item como categorias ordinais de classificação. Isso é aplicável a itens politômicos. Em outras palavras, a solução matemática da análise de Rasch para Créditos Parciais é o melhor modelo para escalas Likert (1932) a partir de três níveis de discriminação.

Andrich (1978) e Masters (1982),todavia, possuem abordagens divergentes no tratamento da função matemática que descreve a probabilidade de resposta para um traço latente. A principal diferença entre os modelos é que as Medidas de Escalonamento de Andrich (1978) consideram um grupo único de itens. Assim sendo, todos os itens são agrupados dentro de uma mesma categoria ordenada e os intervalos entre as categorias no modelo de Andrich são iguais. Em Masters (1982) os intervalos fazem parte do mesmo eixo de classificações. Nem sempre, porém, os intervalos são iguais e as categoriais ordinais baseiam-se na média entre as categorias adjacentes.

3.4.2.1
O Modelo de Escalonamento de Andrich (1978)

Andrich contextualiza, pela primeira vez, uma proposta de avaliação da resposta a itens politômicos. Neste sentido, considera que todos os itens de uma escala compartilham do mesmo intervalo ordinal entre as categorias discriminatórias das respostas. A notação matemática de Andrich (1978) é expressa pelo algoritmo:

$$P\,\{X_{ni} = x\} = \frac{e^{\left(x(\beta_n - \sigma_i) - \sum^{j} \tau_{ki}\right)}}{\sum_{j=0}^{m} e^{\left(j(\beta_n - \sigma_i) - \sum_{k=0} \tau_{ki}\right)}}$$

Nessa equação, $P\ \{X_{ni} = x\}$ é a probabilidade de endossamento da pessoa *n* com o parâmetro pessoal β_n recebendo *x* pontos em relação ao item *i* com o parâmetro do item σ_i no qual x = {0,1,2, ..., *m*}. Nesse contexto, o parâmetro β_n é a medida do nível de traço latente apresentado pela pessoa e σ_i é expressado por quão difícil ou o quanto de traço latente é revelado pelo item *i* quando endossado pela pessoa (Rasch, 1961, Andrich, 1978).

Um item *j* cujas respostas estão ordenadas apresenta *m* limiares ou intervalos τ_k no qual *k* = {0, 1, 2, 3, ..., m}. É possível interpretar os limiares τ_k como parâmetros de σ_{ik} para cada categoria de discriminação das respostas para o item *i*. Assim sendo, todos os itens da escala compartilham a mesma escala ordinal de *m* + 1, para uma diferença de limiares de k = 0.

Os limiaresque definem o valor dos escores da escala ordinal Likert, também chamados de intervalos, são considerados pontos da escala latente quando as probabilidades condicionais de pontuação em uma das duas categorias de discriminação adjacentes são iguais a ½ (Andrich, 1982). O valor do limiar τ_1 corresponde ao nível de endossamento do item, no qual a probabilidade de uma pessoa em pontuar 1 é 50% e 0 também é 50%.

O limiar τ_2 corresponde à probabilidade condicional de 50% de resposta do escore 1 e 50% do escore 2, e assim por diante. Para uma escala Likert de 5 graus de discriminação, por exemplo, haveria quatro intervalos a serem ordenados seguindo o princípio $\tau_{ki} + 1 > \tau_{ki}$.

No ASQ-3 os escores foram distribuídos da seguinte forma: 0 para “Ainda Não”, 5 para “Às Vezes”, e 10 para “Sim”. O número de limiares é constituído pelos intervalos entre os escores das categorias. Nesse caso há,portanto, 2 limiares: de 0 a 5, e de 5 a 10. Logo, para o ordenamento dos limiares das categorias de discriminação do item 4 da escala de Comunicação de 6 meses, teremos arepresentação matemática dada pela notação: $\tau_{26m4com} + 1 > \tau_{16m4com}$,onde $\tau_{26m4com}$ é o limiar dos escores 5 e 10 e $\tau_{16m4com}$, o intervalo entre os escore 0 e 5.

3.4.2.2
O Modelo de Créditos Parciais de Masters (1982)

O modelo de Masters (1982), por sua vez, considera que cada item possui uma escala ordinal de m_i + 1 única. A expressão matemática do modelo de Créditos Parciais de Masters (1982) é expressa pela função:

$$P\,\{X_{ni} = x\} = \frac{e^{\,x\beta_n - (\sum_{k=0}^{x} \sigma_{ki})}}{\sum_{k=0}^{m_i} e^{\,j\beta_n - (\sum_{k=0}^{j} \sigma_{ki})}}$$

Nessa função, a probabilidade $P\,\{X_{ni} = x\}$ usa a diferença entre o parâmetro β_n da pessoa como referência e o parâmetro σ_{ki} para cada item *i* em função do parâmetro *k* das categorias. Essa mudança no modelo de Créditos Parciais permite fazer a inferência dos limiares e dos parâmetros da pessoa e dos itens para cada item ou grupo de itens que compartilha dos mesmos limiares.

3.4.2.3
Vantagens e Desvantagens dos Modelos de Créditos Parciais e o Algoritmo de Linacre (2005) para o Winsteps 3.72

O modelo de Masters (1982) diferencia-se do de Andrich (1978) pois dá um tratamento particular a cada um dos itens, respeitando os dados empíricos e fornecendo-lhe um modelo matemático específico ao considerar que pode ter seu próprio grupo de respostas. Em outras palavras, a principal diferença entre os dois modelos é que Andrich considera que todos os itens compartilham uma mesma escala Likert e os limiares entre as categorias são os mesmos, enquanto Masters considera que cada item tem seu escalonamento único e os limiares entre as categorias são particulares a cada item.

Logo, enquanto Andrich trabalha com limiares de *m* + 1, Masters trabalha com limiar de m_i+ 1, onde *i* é um item da escala que pode ou não compartilhar dos mesmos limiares de outros itens.

A principal vantagem do modelo de Andrich é universalizar o modelo para todos os itens da escala. Com isso, o pesquisador pode conhecer a configuração geral da escala comparando-a com um modelo global. Os itens são considerados de um mesmo grupo, e, portanto a escala é considerada como um todo, o que faz mais sentido para um modelo univariado.

A desvantagem do escalonamento de Andrich é não respeitar as diferenças entre o conjunto de dados empíricos de itens específicos, pois, ao considerar que os limiares são iguais para todos os itens, não permite que algumas partes da escala tenham respostas ordinais categóricas com modelos diferentes à escala global.

O modelo de Créditos Parciais de Masters, por sua vez, traz como vantagem permitir diferentes modelos de escalonamento a partir das idiossincrasias do item. Isto facilita a padronização da escala, mesmo que os itens tenham diferentes dados empíricos, pois os modelos comportam diferenças entre os limiares. Todavia, o modelo de Masters agrupa itens com as mesmas características nos limiares, o que pode ser prejudicial para a análise de dados univariados, uma vez que a análise de Rasch pressupõe unidimensionalidade.

Apesar das diferenças entre os modelos matemáticos, Linacre (2005) demonstrou, ao analisar o conceito de dificuldade do item nos dois modelos, que o parâmetro de dificuldade de Masters (1982) σ_{ki} e o de Andrich (1978) σ_i possuem propriedades matemáticas idênticas. A interpretação da dificuldade do item nesses dois modelos apresenta probabilidades iguais nas categorias de respostas menores e maiores, em face da localização do item no traço latente. Para confirmar a equivalência das equações, Linacre (2005) propôs que a probabilidade da presença do traço latente β_n de uma pessoa *n* é igual na menor e na maior categoria de discriminação de um item *i* cuja dificuldade é dada por σ_i.

O algoritmo de Linacre (2005) faz uma média entre os algoritmos de Masters e Andrich propondo que as probabilidades são iguais, isto é: $P\ \{X_i = 0\} = P\ \{X_i = x\}$. Com isto, o *software Winsteps* Versão 3.72.30 (Linacre, 2009) utilizado neste estudo, utiliza o algoritmo médio de Créditos Parciais (Linacre, 2005) para fazer a análise dos itens do ASQ-3 com base no modelo de análises de Rasch (1961).

O *software* parametriza o *logit* dos itens em função da dificuldade do item definida por σ_{ij} para categorias adjacentes *i* e *j* para σ_i do algoritmo igual à média de σ_{ij} (Linacre, 2005, 2009). Algebricamente, as três probabilidades são equivalentes (Linacre, 2009).