

4

Metodologia

4.1

Participantes

Os participantes deste estudo foram recrutados pela Secretaria Municipal de Educação do Rio de Janeiro. O total de crianças avaliadas foi 45,650 dentro da faixa etária de 4 meses a 60 meses. Todas as crianças de creches públicas do município do Rio de Janeiro foram avaliadas pelos agentes auxiliares de creche (AAC), ou cuidadores. Os cuidadores são profissionais de regime estatutário cuja função é dar a atenção e participar das intervenções pedagógicas no ambiente de creche. Somente AACs preencheram o ASQ-BR.

Inicialmente, uma parceria entre o autor e a instituição em que esta dissertação foi feita com o Instituto de Estudos do Trabalho e Sociedade (IETS) foi consolidado, pois havia o interesse da prefeitura do Rio de Janeiro em utilizar o ASQ-BR como indicador de políticas públicas.

A prefeitura buscou o IETS e este que entrou em contato com o autor desta dissertação para estudar as propriedades psicométricas do instrumento e definir os critérios ideais para aplicação na rede municipal de atendimento educacional de creches.

Essa parceria teve financiamento de diversas fontes: Banco Mundial de Desenvolvimento, através do IETS; FAPERJ, através da PUC-Rio e da Universidade Estadual do Rio de Janeiro; da Secretaria Municipal de Educação (SME) do Rio de Janeiro. Os esforços financeiros e logísticos das quatro instituições culminaram na aplicação do ASQ. A gerência do projeto, da aplicação e da infra-estrutura ficou a cargo do IETS em parceria com o autor.

4.2

Adaptação Transcultural do ASQ Brasil (ASQ-BR)

O ASQ Brasil (ASQ-BR) foi adaptado para a cultura brasileira a partir da Terceira Edição do ASQ (ASQ-3; Squires et al., 2009). O primeiro passo foi

utilizar o procedimento de retrotradução como parte da metodologia de adaptação (Flaherty et al., 1988). Três tradutores bilingues diferentes fizeram a versão do ASQ-3 (Squires et al., 2009) para o português.

Um grupo de especialistas em desenvolvimento infantil e educação ($n = 15$) de diferentes instituições brasileiras foram solicitados a avaliar e determinar a melhor versão para cada um dos itens do ASQ-BR ($n = 540$). A versão preliminar em Português para o Brasil foi, então, retrotraduzida para o inglês por um tradutor nativo dos Estados Unidos. Ambas as versões em inglês, o original do ASQ-3 e a versão retrotraduzida, foram comparadas novamente pelo mesmo grupo de especialistas e mínimas diferenças semânticas foram encontradas.

O estudo piloto foi feito para determinar problemas de compreensão e entendimento dos itens com 120 professores e AACs. Peculiaridades culturais foram consideradas tanto quanto outras observações feitas pelos entrevistados no estudo piloto. Comentários e sugestões feitas neste período foram incorporados na versão final do ASQ-BR.

As adaptações culturais foram muito importantes para a versão brasileira do ASQ, o ASQ-BR. Palavras comumente usadas nos EUA foram modificadas para adequação à cultura brasileira: “polegadas” transformaram-se em “centímetros”, “pés” transformaram-se em “metros”, “libras” transformaram-se em “quilos”, e “Cheerio” transformaram-se em “pedacinhos de biscoito”.

Da mesma forma, não se ensina o sobrenome às crianças brasileiras senão por volta dos 6 anos de idade. Portanto, a expressão “Último Nome” foi substituída por “Nome da(o) Mãe/Pai” em alguns itens.

Outros itens culturalmente inapropriados foram modificados. Por exemplo: “alimenta-se com bolacha ou biscoito” transformou-se em “alimenta-se com uma fruta”, porque as crianças nas creches do Rio de Janeiro normalmente comem frutas e vegetais, evitando produtos alimentícios industrializados. Tais substituições foram feitas após consultar o grupo de especialistas em desenvolvimento infantil e educação para assegurar que o construto psicológico avaliado pelos itens se manteve o mesmo.

A versão final do ASQ-BR foi usada para mensurar o desenvolvimento global de bebês e crianças em creches públicas municipais do Rio de Janeiro.

4.3

Procedimentos

A versão final do ASQ-BR foi aplicada em 46 creches públicas do Rio de Janeiro. Para viabilizar logisticamente a aplicação completa, uma versão do ASQ-BR foi programada em versão *website*. Este ficou à disposição dos diretores das creches das respectivas instituições através do endereço: <http://www.pesquisadata.com.br/asq3>. Assim sendo, a Secretaria Municipal de Educação (SME) do município do Rio de Janeiro disponibilizou o *hyperlink* do sítio virtual permitindo acesso completo à versão *online* dos questionários do ASQ-BR.

Todos os dados foram adicionados a uma base de dados diretamente pelos diretores e professores através dos computadores instalados dentro das próprias creches. O estudo foi conduzido entre 12 de outubro de 2010 e 17 de dezembro de 2010.

4.4

Análise dos Dados

4.4.1

Teoria Clássica dos Testes (TCT)

Para analisar os índices da Teoria Clássica dos Testes (TCT) do ASQ-BR, dois diferentes aspectos foram considerados: homogeneidade do teste e investigação dos itens.

A homogeneidade foi estudada a partir de sua consistência interna usando o coeficiente Alfa de Cronbach (Cronbach, 1951). Esse coeficiente usa a técnica de divisão ao meio (*split-half*) para estimar a correlação entre duas amostras aleatórias de itens do universo de itens totais do teste. Cronbach (1951) ressalta que o coeficiente alfa é um índice interessante para verificar a concentração do primeiro fator de um teste. O valor aceitável do alfa é acima de 0.70 (Cronbach, 1969).

A investigação das características dos itens na TCT foi analisada pela correlação item-total conforme sugerido por Cohen, Swerdlik, e Douglas (1992).

Uma boa correlação item-total situa-se acima de 0.30. Esse índice informa quanto um item contribui para o escore total de uma escala. Se um item do ASQ-BR apresentar $r < 0.30$, provavelmente o item corresponderá a uma dimensão que não é a da escala que está alocado.

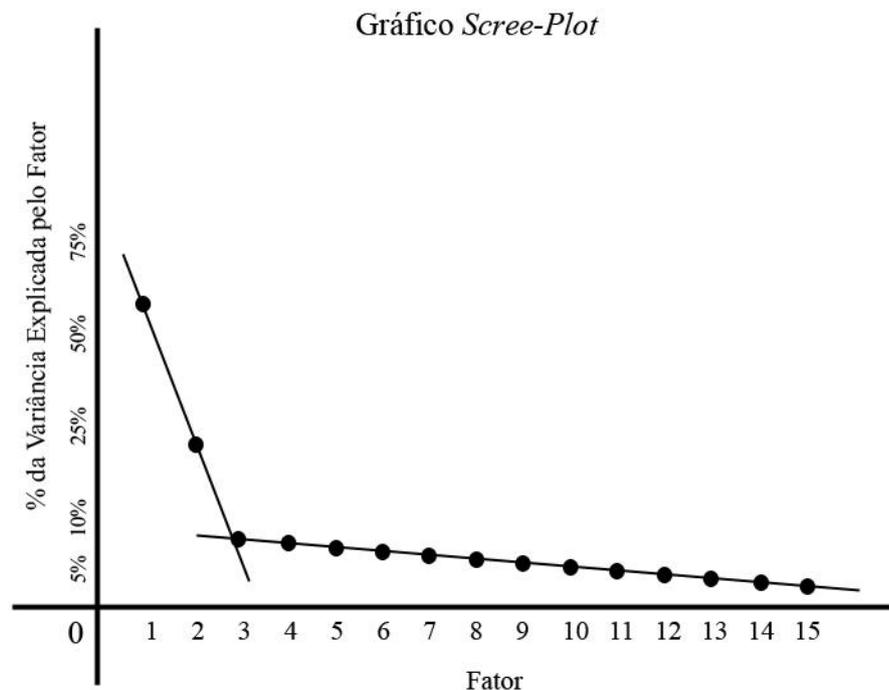
A análise do ASQ-BR deve respeitar cada domínio do desenvolvimento como um fator em separado dentro da respectiva faixa etária. Coeficientes da TCT foram calculados pelo IBM SPSS 18.0 (IBM, 2009).

4.4.2

Análise Fatorial

Análise fatorial é usada para extração do número de dimensões da escala. Cattell (1966) propôs o teste de *scree plot* para identificar fatores de uma escala. O procedimento é simples: a partir de um gráfico que plota os autovalores da escala no eixo cartesiano calculados pelo método de Análise de Componentes Principais (ACP), uma linha reta é traçada a partir dos pontos plotados com menor valor. Os pontos de maior valor que ficaram fora da reta traçada são o número de fatores da escala (Cattell, 1966).

Figura - Exemplo do método de extração de fatores de Cattell (1966)



Legenda: Ao plotar os autovalores (*fator*) em um gráfico cartesiano, começamos a traçar uma linha reta a partir do décimo-quinto fator (15). Esta linha reta claramente vai até o terceiro fator (3), porém a tendência da reta não sobrepõe o segundo (2) e nem o primeiro fator (1). Isso significa que a escala, de acordo com o método de Cattell, apresenta duas dimensões: os dois maiores autovalores que estão fora da reta.

Todavia, outros métodos de extração de fatores de uma escala são mais rigorosos que o teste *scree-plot* de Cattell. O'Connor (2000) recomenda usar dois diferentes métodos para acessar e confirmar a dimensionalidade de uma escala: a Média Parcial Mínima de Velicer (MAP) e o método de Análises Paralelas.

O MAP de Velicer permite combinar uma ACP completa com uma série de matrizes parciais de correlação. A Análise Paralela extrai autovalores de dados empíricos aleatoriamente escolhidos que correspondem a partes iguais de itens e pessoas da amostra total (O'Connor, 2000).

Neste estudo, a análise fatorial foi feita a partir do método de Cattell (1966) e dos métodos mais rigorosos propostos por O'Connor no IBM SPSS 18.0 (IBM, 2009), usando o gráfico *scree plot* padrão do *software* e duas diferentes sintaxes para o SPSS programadas por O'Connor (2000).

4.4.3

Análise Não-Paramétrica de Mokken

A análise de Mokken é conhecida como Teoria de Resposta ao Item Não-Paramétrica (NIRT). A tendência das respostas do ASQ-3 a apresentar respostas concentradas em uma das extremidades nas respostas ordinais categóricas de Likert demonstra que o instrumento possui características de probabilidade que diferem das distribuições normais, pressupostos de parametria (Hair et al., 2007).

Há evidências na literatura (Squires et al., 2009) de que o ASQ-3 é utilizado como instrumento de *screening*, pois a maior parte das respostas está alocada no “Sim”, portanto, em uma escala Likert de 3 eixos categóricos. Esse fato indica que a maior probabilidade de resposta está na categoria de maior valor.

O modelo univariado paramétrico de Rasch pressupõe que a probabilidade de resposta da menor e da maior categoria é igual, o que pode não ocorrer nesta medida. Portanto, a análise de Mokken poderá contribuir bastante na compreensão da escalabilidade do ASQ-BR e na capacidade dos itens em se diferenciarem ao

longo dos cinco traços latentes, ou seja, ao longo das cinco dimensões do desenvolvimento infantil do ASQ-BR.

Utilizamos o *software* MSP 5.0 (Molenaar, Sijtsma e Boer, 2000, Sijtsma et al., 1990) na análise de Mokken para o presente estudo calculando o coeficiente de confiabilidade *rho* para cada escala (Molenaar e Sijtsma, 1988), assim como o coeficiente *H* de Loevinger (Loevinger, 1948, Mokken, 1971) para a escala global. Os itens foram tratados da mesma forma, baseados no algoritmo dos coeficientes para itens politômicos (Molenaar, 1982, Sijtsma et al., 1990).

Os valores limítrofes de *rho* situam-se acima de 0.30. Valores menores apontam para ausência de escalabilidade dos dados. Em outras palavras, caso o valor de *rho* seja menor que 0.30, a escala não respeita a monotonicidade de Guttman (1944).

Para o *H* da escala, os valores dos resultados são divididos em três forças em função da escalabilidade: uma escala boa ou forte apresenta $H \geq 0.50$; uma escala média ou medíocre tem o valor entre $0.40 \leq H < 0.50$; uma escala ruim ou fraca, entre: $0.30 \leq H < 0.40$. Abaixo de 0.30 não há monotonicidade e a análise de Mokken não é recomendada (Sijtsma et al., 1990).

O *H* do item, por sua vez, funciona com um ponto de corte mínimo em 0.30. Itens abaixo desse valor provavelmente apresentam um padrão de resposta probabilística diferente do padrão da escala. O item, portanto, provavelmente não acompanha a monotonicidade do instrumento de que faz parte e é um item extremamente fácil ou difícil. Isso significa que o item não diferencia as pessoas e não se comporta corretamente ao longo da curva do traço latente (Molenaar et al., 2000, Sijtsma et al., 1990).

4.4.4

Modelo de Mensuração de Rasch

A análise de Rasch é uma abordagem baseada no item, sendo considerada um método de Teoria de Resposta ao Item (TRI). A TRI fornece uma visão diferente da TCT, refletindo os traços latentes que estão subjacentes em um instrumento.

O foco principal da análise de Rasch é o uso de uma estatística centrada no sujeito Sua análise opõe-se diametralmente à maioria dos estudos psicométricos

que usam a TCT (Rasch, 1961). De acordo com Olsen et al. (2010), modelos de TRI permitem testar o ajuste (*fit*) entre dados e modelo teórico, apresentando a vantagem de serem modelos empíricos explícitos.

O modelamento de Rasch especifica uma função de distribuição das respostas potenciais de uma pessoa específica para um item específico da escala. Essa distribuição depende dos parâmetros da pessoa e do item. Portanto, o modelo matemático permite a comparação dos dados empíricos em face do modelo teórico para ambos os parâmetros: itens e pessoas (Rasch, 1961).

A medida de ajuste (*fit*) é uma medida do modelo Rasch dada por dois coeficientes de desajuste: *infit* e *outfit* (Linacre, 2009). *Infit* determina o padrão dos parâmetros dentro do traço latente, que corresponde à probabilidade do padrão de resposta corroborar o padrão de Andrich (1978) e Masters (1982). Em outras palavras, as categorias possuem respostas de igual probabilidade para os maiores e menores escores. Portanto, o coeficiente de *infit* dá a informação que compara o padrão de respostas nos dados empíricos e o modelo teórico, ou seja, permite discernir se um item mais fácil recebe mais respostas corretas do que um item mais difícil.

Outfit, de maneira inversa, corresponde às respostas fora do padrão, isto é, a parâmetros de itens e pessoas que não correspondem ao padrão dos dados empíricos. O coeficiente de *outfit* reflete os itens que deveriam apresentar certo tipo de resposta, mas não o fazem. Indica a probabilidade de resposta errada a um item que deveria ter a resposta correta. Os coeficientes de *infit* e *outfit* são dados pela média quadrática (MnSq) para cada uma das medidas em separado, tendo valor aceitável entre 0.5 e 1.5 (Linacre, 2009).

De acordo com Linacre (2009), o coeficiente de confiabilidade no modelo Rasch é a reprodutibilidade de uma localização relativa de um traço latente. Isso é o mesmo que afirmar a possibilidade de avaliar se uma pessoa ou um item realmente tem as medidas que foram extraídas pelos dados empíricos a partir de sua variabilidade e probabilidade de erro.

O objetivo do modelo é conhecer se o instrumento é capaz de discriminar e classificar as pessoas ao longo do traço latente, ou seja, se uma pessoa possui maior quantidade de um construto psicológico que outra pessoa. Valores altos de confiabilidade da pessoa mostram importantes características do modelo Rasch: a quantidade de variância presente na habilidade medida, se o tamanho do teste está

adequado, se o número de categorias equivale ao recomendado, se os dados da amostra são suficientes para discriminar sujeitos.

O coeficiente de confiabilidade da pessoa demonstra a boa capacidade de discriminação entre os sujeitos da amostra. Quando está acima de 0.9, significa de 3 a 4 níveis de discriminação; quando varia entre 0.8 a 0.9, significa 2 a 3 níveis de discriminação; quando varia entre 0.5 a 0.8, significa 1 ou 2 níveis; e, quando encontra-se abaixo de 0.5, significa 1 ou nenhum nível de discriminação (Linacre, 2009).

A confiabilidade do item mostra outras características em relação ao parâmetro do estímulo. Apesar de independe da extensão do teste ou do ajuste do modelo, a confiabilidade do item varia de acordo com duas variáveis: a dificuldade do item e o tamanho da amostra. Quanto maior for a extensão do traço latente e maior o tamanho da amostra, maior será a confiabilidade do item.

Entretanto, amostras pequenas não apontam somente para uma confiabilidade do item baixa, mas também criam problemas com relação à precisão da localização do item ao longo de um traço (Linacre, 2009). A confiabilidade do item abaixo de 0.8 é considerada baixa para este estudo, mas classicamente este valor é dado por 0.9.

Espera-se que os coeficientes de confiabilidade do ASQ-BR sejam diferentes entre si, uma vez que as características peculiares ao instrumento não são as mesmas de outros testes de aptidão. O ASQ é um instrumento de detecção baseado em pontos-de-corte (Squires et al., 2009, Heo et al., 2008). Consequentemente, a confiabilidade para as pessoas ficará provavelmente entre 0.5 e 0.8 para 1 ou 2 níveis de discriminação.

Isso ocorre, porque o ASQ-BR é um teste de curta extensão, possuindo apenas 6 itens por escala, baixo número de categorias por item (Likert de 3 categorias) e baixa variância das respostas (os sujeitos tendem a responder corretamente aos itens). Todavia, a confiabilidade do item deve ser muito alta, uma vez que a amostra do ASQ-BR possui tamanho acima do recomendado. O modelo de medidas de Rasch foi analisado usando o *Winsteps 3.72.3* (Linacre, 2009).