

6

Discussão e Conclusão

O ASQ-BR tem sua gênese no ASQ-3 (Squires et al., 2009). O instrumento é uma ferramenta de detecção que mensura o desenvolvimento de crianças em relação à sua idade a partir de 1 mês de vida até 66 meses de idade.

Apesar de a proposta inicial do instrumento era ser aplicado pelos pais das crianças, Allen (2007) reportou que algumas pré-escolas e creches na Florida utilizaram-no como uma medida aplicada por professores e cuidadores. O ASQ-3, nesse tipo de contexto, deve ser respondido por uma pessoa que conheça bem os hábitos e habilidades da criança (Squires et al., 2009). Tal pessoa pode ser um dos pais, familiar, professor ou cuidador. Neste estudo, o ASQ-BR foi usado por professores e cuidadores de creches e pré-escolas do Rio de Janeiro.

Todavia, o ASQ-3 é validado para a cultura e realidade americanas. Nesse sentido, adaptações culturais foram necessárias para respeitar as diferenças e idiosincrasias da amostra brasileira. A metodologia de retrotradução (Flaherty et al., 1988) foi aplicada para traduzir do inglês para o português do Brasil. A partir daí, adaptações culturais foram feitas respeitando as diferenças entre os dois países e consultando um grupo de especialistas em desenvolvimento e educação infantil para confirmar eventuais adaptações. A versão final da adaptação transcultural produziu uma versão brasileira do ASQ-3 chamada ASQ-BR.

A eficácia do ASQ-BR como medida e indicador do desenvolvimento foi estudada a partir da perspectiva do sistema público de creches e pré-escolas do Rio de Janeiro. Diferentemente do ASQ-3 e outras versões do ASQ, os respondentes do ASQ-BR foram apenas professores e cuidadores. Os dados e resultados obtidos mostraram que a adaptação para a língua e cultura brasileira foi bem sucedida. Destarte, o ASQ-BR demonstrou que é uma ferramenta eficaz e que permite a avaliação de crianças no ambiente pré-escolar e de creche.

Allen (2007) já havia relatado alguns casos na Flórida, Estados Unidos, do uso do ASQ em creches. De acordo com Baggett et al. (2007), o ASQ é uma boa medida de detecção de problemas no desenvolvimento infantil para programas e políticas públicas quando combinado com outros instrumentos.

Análises em Teoria Clássica de Testes (TCT) e Teoria de Resposta ao Item (TRI) foram conduzidas para estudar as propriedades psicométricas do ASQ-BR. Na análise de TCT, a homogeneidade foi medida pelo coeficiente alfa de Cronbach (Cronbach, 1951) e os itens foram investigados um a um em relação à correlação item-total, como sugerido por Cohen et al. (1992).

A análise fatorial foi conduzida usando três diferentes técnicas: Teste *scree plot* de Cattell (Cattell, 1966), Teste MAP de Velicer e análise paralela (O'Connor, 2000). A análise em TRI foi feita a partir dos modelos de mensuração de Mokken (Mokken, 1971) e Rasch (Rasch, 1961) para compreender ambos os parâmetros dos itens e pessoas nas escalas do ASQ-BR.

Os índices de Mokken foram extraídos a partir do *rho* de confiabilidade da escala para verificar monotonicidade (Molenaar e Sijtsma, 1988) e *H* de Loevinger para itens politômicos para cada item e para a escala toda (Loevinger, 1948, Sijtsma et al., 1990). A medida de ajuste do modelo Rasch para o ASQ-BR usou duas diferentes estatísticas: média quadrática de *infite outfit*. A confiabilidade com base no modelo Rasch foi calculada para os parâmetros de pessoas e itens a fim de compreender a reprodutibilidade da localização do *logit* (Linacre, 2009).

De modo geral, a análise de TCT mostrou boa consistência interna quando analisada pelo coeficiente alfa de Cronbach, com exceção do domínio Pessoal-Social. Isso pode ser justificado pela combinação de dois diferentes construtos na mesma escala: Domínio Pessoal, relacionado ao reconhecimento e desenvolvimento do *self*, e Domínio Social, relacionado com habilidades sociais e características de autonomia e independência (Coll, Marchesi, e Palacios, 2004).

O mesmo domínio também mostrou problemas com relação aos itens. 16.7% dos itens mostraram correlação item-total menor que 0.30 e, conseqüentemente, pequenas contribuições para a medida global de desenvolvimento infantil no ASQ-BR. Contudo, as estatísticas de TCT encontradas no ASQ-BR são muito similares quando comparados com o ASQ-3 original (Squires et al. 2009) e outras adaptações culturais, como as versões do ASQ coreana (K-ASQ; Heo et al., 2008), holandesa (D_ASQ; Kerstjens et al., 2009), e chinesa (Tsai et al., 2006).

A análise fatorial do ASQ-BR revelou unidimensionalidade dos domínios, com poucas exceções. O domínio Pessoal-Social apresentou duas escalas com

dois fatores extraídos pela análise paralela. Ao verificar o conteúdo semântico dos itens, as dimensões podem ser explicadas a partir de dois construtos dentro do mesmo grupo de itens.

Na escala de 10 meses desse domínio, os itens 1, 2 e 3 estão relacionados a hábitos alimentares (por exemplo, o item 2 é “O bebê se alimenta sozinho com uma bolacha ou biscoito?”) e os demais itens – por exemplo, o item 6 é “Quando você estende a mão e pede ao bebê o brinquedo dele, ele o solta na sua mão?” – estão mais relacionados ao desenvolvimento do *self* (Coll, Marchesi, e Palacios, 2004).

Na escala de 54 meses, o problema permanece o mesmo. O item 4 apresentou um uma dimensão diferente dos outros itens, estando relacionado a comportamentos alimentares (o item 4 é: “A criança se serve, tirando comida de um recipiente para outro usando talheres? Por exemplo, ela utiliza uma colher grande para pegar comida da travessa e colocar no prato?”), enquanto os demais itens estão relacionados a atividades de autonomia (item 1: “A criança lava as mãos com água e sabão e depois se seca com uma toalha, sem ajuda?”, ou item 3: “A criança escova os dentes colocando a pasta de dente na escova e escovando todos os dentes sem ajuda?”)

O teste MAP de Vellicer extraiu também dois fatores da escala Pessoal-Social de 60 meses, mas as dimensões não mantiveram o padrão ressaltado anteriormente. Com base em Coll et al. (2004), os itens 1 e 4 parecem relacionar-se com a autonomia da criança. Por exemplo, o item 4 é: “A criança põe e tira a roupa sozinha, inclusive aquelas com botões de tamanho médio e zíperes na frente da roupa?”, e os itens 2, 3, 5 e 6 são mais relacionados ao desenvolvimento do *self* (o item 6 é: “A criança normalmente aguarda a sua vez e compartilha coisas com outras crianças?”).

Squires, Bricker e Twombly (2010) propuseram um novo instrumento de detecção para resolver esse problema do domínio Pessoal-Social: o Ages & Stages Questionnaire Social-Emotional (ASQ-SE). É recomendado o uso do ASQ-3 combinado com o ASQ-SE para uma detecção mais precisa de problemas no desenvolvimento infantil. Portanto, a partir da perspectiva da TCT, o ASQ-BR pode ser considerado uma boa medida e uma adaptação cultural bastante precisa. Assim como o ASQ-3, recomenda-se a adaptação e validação do ASQ-SE para melhorar os indicadores no Brasil.

O modelo de medidas de Rasch (Linacre, 2009; Rasch, 1961) revelou boas características psicométricas baseadas no índice na média quadrática dos índices de ajuste, com apenas 8.7% dos itens extrapolando o critério sugerido (Linacre, 2009). Contudo, o MnSq de *outfit* predominou com maiores valores em relação ao *infit*.

Provavelmente, como um instrumento de detecção, o ASQ-BR tende a apresentar uma maior concentração de respostas na categoria mais alta, “Sim”, por tratar-se de um instrumento de baixa dificuldade, com valores de MnSq para *outfit* em itens que receberam respostas de presença do traço, quando deveriam ser respondidas com ausência. Isso permite inferir que tais itens não são suficientemente bons para prever o traço latente do desenvolvimento infantil, pois são excessivamente fáceis e não discriminam.

De acordo com Linacre (2009), problemas de *outfit*, apesar de serem afetados pelos *outliers*, não evidenciam um padrão de resposta. Em outras palavras, a escala somente estará comprometida quando um item tiver coeficiente de *outfit* acima de 2.0, o que ocorre em 9 dos 570 itens do ASQ-BR. Apesar do critério de ajuste apresentar valores de MnSq entre 0.5 e 1.5, é possível ter itens com valores abaixo de 0.5, ou entre 1.5 e 2.0 sem comprometer a escala.

No primeiro caso, o item é pouco produtivo para a escala, discrimina pouco e, portanto, o padrão de respostas tem pouca variância, não discriminando as pessoas. No segundo caso, o item é improdutivo para a escala ao passo que não discrimina em virtude da distorção da localização da dificuldade e do item na escala. Porém, apenas valores acima de 2.0 são capazes de degradar totalmente a configuração da escala (Linacre, 2009).

Após compreender as razões para o MnSq de *outfit* acima de 2.0 nos itens do ASQ-BR, dois principais problemas foram revelados: itens com médias muito baixas –que provavelmente revelam uma tendência dos avaliadores em responder negativamente a uma questão que não testaram ou não observaram de modo eficaz. Por exemplo, o item 6 do domínio de Comunicação na escala de 6 meses de idade mostrou *outfit* igual a 9.23: “Se você imita os sons que o bebê costuma fazer, ele repete os mesmos sons para você?”.

Por outro lado, houve itens cujo escore médio foi extremamente alto, o que provavelmente aponta para a possibilidade de crianças que acertaram o item não deveria mostrar aquele tipo de habilidade em função de as outras respostas dadas

em outros itens. Por exemplo, o item 3 do domínio Pessoal-Social, na escala de 10 meses teve MnSq *outfit* igual a 2.52 (“O bebê come sozinho um biscoito?”).

Esses itens provavelmente degradam as escalas de seus respectivos domínios, diminuindo a capacidade do teste em separar pessoas. No entanto, seus impactos no instrumento de modo global são muito pequenos e, conseqüentemente, aceitáveis.

Valores de *infit* acima de 2.0 apareceram apenas em dois itens do ASQ-BR, num total de 570 itens. Linacre (2009) ressalta que o MnSq de *infit* é problemático a partir de valores acima de 2.0. Isso significa que a amostra deveria responder ao item erradamente apresentando ausência daquele traço latente, mas os participantes receberam respostas corretas. Portanto, estaria relacionado a um padrão de resposta que vai de encontro a toda a escala, distorcendo sua capacidade de discriminação e investigação da localização do traço.

Se o padrão de respostas estiver comprometido, Linacre (2009) ressalta que provavelmente há um comprometimento também da validade do instrumento. Os itens que apresentaram este tipo de problema foram: o item 6 de Motora Ampla na escala de 60 meses (“A criança salta alternando os pés? (Você pode mostrar como fazer.”), e o item 2 do domínio Resolução de Problemas na escala de 30 meses (“Se a criança quer alguma coisa que não consegue alcançar, ela procura uma cadeira ou uma caixa para subir e alcançar o objeto (por exemplo, para pegar um brinquedo sobre um balcão ou para “ajudar” você na cozinha?”).

A estatística de confiabilidade para a análise Rasch também demonstrou boas propriedades psicométricas, assim como os índices de confiabilidade que também mostraram bons resultados. Valores baixos de confiabilidade da pessoa foram encontrados, como previamente hipotetizados, devido ao teste ser curto, haver poucas categorias de discriminação por item e baixos valores de variância. Os valores de confiabilidade para a pessoa apontam apenas 1 nível de discriminação (Linacre, 2009), o que é um dado importante, uma vez que o propósito do ASQ-BR é detectar atrasos no desenvolvimento, sem comprometer com diferentes níveis de discriminação. A confiabilidade do item apresentou valores altos para todas as escalas, mostrando que o tamanho da amostra foi bom para o modelo de Rasch.

Portanto, em virtude de todos os dados coletados e as medidas calculadas em TCT e TRI, o ASQ-BR demonstrou-se um bom instrumento de medida para o

desenvolvimento infantil. Sua eficácia reside em ser um instrumento de detecção de atrasos no desenvolvimento das crianças e permitir acessá-las dentro do sistema público escolar. Portanto, O ASQ-BR pode servir como indicador para quantificar a eficácia de políticas públicas e programas governamentais, justificando investimentos e, paralelamente, firmando-se como alicerce sólido no encaminhamento das crianças detectadas pelo ASQ-BR para profissionais da saúde.

Para estudos futuros, recomendamos a replicação do ASQ-BR com itens alterados baseando-se nas questões que apresentaram algum tipo de problema em relação à correlação item-total, H de Loevinger e estatísticas de ajuste ao modelo Rasch.

A adaptação transcultural de um instrumento tão extenso é um processo lento e que demanda diversas e repetidas análises, como a modelagem por análise fatorial confirmatória (AFC), Teoria de Resposta ao Item Multidimensional (TRIM) e equações estruturais que permitam conhecer as possíveis variáveis intervenientes e espúrias que podem atravessar os resultados e explicar melhor a variância e o padrão de resposta obtidos.