

Propriedades psicométricas da versão brasileira do Denver II: teste de triagem do desenvolvimento

Psychometric properties of the Brazilian version of the Denver II: developmental screening test

Janaina Araujo Teixeira Santos (<https://orcid.org/0000-0002-5514-6264>)¹
 Kennea Martins Almeida Ayupe (<https://orcid.org/0000-0002-4410-8748>)²
 Amanda Larissa Oliveira Lima (<https://orcid.org/0000-0001-9418-2690>)³
 Karolina Alves de Albuquerque (<https://orcid.org/0000-0002-6044-3570>)⁴
 Fabiane Frota da Rocha Morgado (<https://orcid.org/0000-0002-3969-9029>)⁵
 Paulo José Barbosa Gutierrez Filho (<https://orcid.org/0000-0001-7753-0825>)¹

Abstract *The scope of the Denver II Developmental Screening Test is to assess the development of children between 0 and 6 years of age. The aim of this study was to verify evidence of intra- and inter-examiner reliability, concurrent validity, sensitivity and specificity of the Brazilian version of Denver II. It involved a cross-sectional methodological study. A total of 254 children, between 0 and 72 months with risk of developmental delay, participated in the study. Two examiners verified the Denver II intra- and inter-examiner reliability. Concurrent validity, sensitivity and specificity were checked against the Brazilian version of the Ages & Stages Questionnaires as a criterion-referenced test. Statistical analysis used the Intraclass Correlation Coefficient, Spearman's Correlation Test and Contingency Table, the level of significance being $\alpha=0.05$. The results identified that intra and inter examiner reliability was excellent in all age groups. Concurrent validity showed moderate to very strong rates in the 13-to-60-month age group. The sensitivity and specificity indices ranged from 73-99% and 58-92%, respectively. The Brazilian version of Denver II has good rates of psychometric properties and is a reliable and valid instrument to be applied to Brazilian children at risk of developmental delay.*

Key words *Screening test, Validation study, Evaluation, Child development*

Resumo *O Denver II tem como objetivo avaliar o desenvolvimento de crianças entre 0 e 6 anos de idade. O objetivo deste estudo foi verificar a confiabilidade intra e interexaminadores, validade concorrente, sensibilidade e especificidade da versão brasileira do Denver II. Estudo metodológico transversal. Participaram 254 crianças, entre 0 e 72 meses, com risco para atraso no desenvolvimento. Dois examinadores verificaram a confiabilidade intra e interexaminadores do Denver II. Validade concorrente, sensibilidade e especificidade foram verificadas com relação à versão brasileira do Ages & Stages Questionnaires como teste de critério. Análise estatística utilizou o Intraclass Correlation Coeficiente, Teste de Correlação de Spearman e Tabela de Contingência, nível de significância $\alpha=0.05$. Os resultados identificaram que a confiabilidade intra e interexaminadores foi excelente em toda a amostra. A validade concorrente apresentou índices moderados a muito fortes entre 13 e 60 meses. Índices de sensibilidade e especificidade variaram de 73-99% e 58-92%, respectivamente. A versão brasileira do Denver II apresenta bons índices de propriedades psicométricas sendo um instrumento confiável e válido para ser aplicado em crianças brasileiras sob risco de atraso no desenvolvimento.*

Palavras-chave *Teste de triagem, Estudos de validação, Avaliação, Desenvolvimento infantil*

¹ Programa de Pós-Graduação em Educação Física, Universidade de Brasília. Campus Universitário Darcy Ribeiro, Faculdade de Educação Física. 70910-900 Brasília DF Brasil.
janafisiot@gmail.com

² Colegiado de Fisioterapia, Universidade de Brasília. Ceilândia DF Brasil.

³ Programa de Pós-Graduação em Ciências da Reabilitação, Universidade de Brasília. Brasília DF Brasil.

⁴ Centro de Ciências da Saúde, Universidade Federal do Espírito Santo. Vitória ES Brasil.

⁵ Instituto de Educação, Universidade Federal Rural do Rio de Janeiro. Seropédica RJ Brasil.

Introdução

O desenvolvimento é um processo dinâmico, transacional e contextualista, que se inicia na concepção, segue por toda vida e envolve transformações nos aspectos físicos, psicológicos, linguísticos e sociais^{1,2}. Esse processo é multifacetado, amplamente constituído pela transação das características herdadas geneticamente e experiências oferecidas pelo ambiente^{2,3}. Por sua complexidade, a avaliação do desenvolvimento é um desafio para pesquisadores e estudiosos, em especial, quando o enfoque é o público de crianças na primeira infância em países de baixa e média renda⁴.

Lactentes e crianças até os cinco anos de idade possuem reconhecidos fatores de risco de atraso no desenvolvimento, dentre os quais se destacam os biológicos e ambientais^{5,6}. Os fatores biológicos incluem prematuridade, complicações no parto, fatores intrauterinos, nutrição da criança, infecções da infância e exposição a toxinas ambientais^{7,8}. Já os fatores de risco ambientais abarcam depressão, baixa escolaridade, altos níveis de estresse dos cuidadores, restrições econômicas, exposição à violência, precariedade dos serviços de saúde e sociais⁸. Tais fatores são preditores fundamentais no diagnóstico de atraso do desenvolvimento, que significa o atraso na aquisição de uma habilidade na idade esperada. Portanto, necessitam de acompanhamento contínuo, a fim de auxiliar no diagnóstico precoce e encaminhamento para intervenções eficazes e assertivas⁹.

Em conjunto com a detecção de fatores de risco e do diagnóstico de atraso no desenvolvimento, tem sido utilizado, em diferentes países, o Denver II: Teste de triagem do desenvolvimento^{10,11}. Este é um teste de triagem, cujo objetivo é detectar alterações/atrasos e acompanhar o desenvolvimento de crianças na faixa etária de zero a seis anos^{11,12}. O teste foi criado em 1967 e revisado na década de 1990, passando a se chamar *The Denver Developmental Screening Test - Denver II*. O teste apresenta 125 itens que avaliam quatro áreas, quais sejam: pessoal-social, motor fino-adaptativo, motor grosso e linguagem^{11,13}. Sua amostra normativa contou com 2.096 crianças norte-americanas¹². Estudos que verificaram suas propriedades psicométricas demonstram que o Denver II apresenta alta taxa de sensibilidade (83%) e excelente confiabilidade (0.80 – 0.90)¹⁴⁻²⁰.

No Brasil, a avaliação do desenvolvimento em crianças tem sido pautada em testes padronizados que foram desenvolvidos em outros países e

necessitam de verificação de suas propriedades psicométricas na população brasileira²¹. O Denver II é o teste padronizado de triagem mais utilizado na prática clínica e em pesquisas brasileiras, devido à facilidade e pouco tempo de aplicação, idade de abrangência e baixo custo^{10,22,23}. Em 2017 o Denver II foi traduzido e adaptado para o Brasil com publicação do manual na língua portuguesa¹¹. Estudo de validação psicométrica da versão brasileira do Denver II identificou que o teste apresenta bons índices de validade interna e de constructo em crianças entre 0 e 28 meses de idade do nordeste brasileiro¹⁵. Entretanto, ainda são necessários estudos metodológicos que verifiquem evidências das demais propriedades psicométricas do teste em crianças das demais faixas etárias e em crianças sob risco de atraso no desenvolvimento. O objetivo desse estudo foi verificar evidências de confiabilidade intra e interexaminadores, validade concorrente, sensibilidade e especificidade da versão brasileira do Denver II em crianças sob risco de atraso no desenvolvimento.

Método

Este estudo metodológico de temporalidade transversal consistiu na verificação da confiabilidade da versão brasileira do Denver II: Teste de Triagem do Desenvolvimento, bem como da validade concorrente, sensibilidade e especificidade em relação à versão brasileira do *Agès & Stages Questionnaires (ASQ-3BR)*²⁴.

Participantes

Foram selecionadas 254 crianças com risco de atraso no desenvolvimento infantil nos ambulatórios de acompanhamento do Hospital Universitário de Brasília, Hospital de Ceilândia do Distrito Federal e nos Programas de Educação Precoce do DF, entre maio de 2019 a fevereiro de 2020.

Foram incluídas crianças de 1 a 72 meses de idade com histórico de fatores de risco biológicos ou ambientais para atraso no desenvolvimento infantil, como prematuridade, baixo peso, complicações gestacionais e menor nível socioeconômico. Foram excluídas as crianças com diagnóstico comprovado de condições de saúde genéticas, congênitas ou neurológicas, como Síndrome de Down e Paralisia Cerebral. Tais condições de saúde podem causar, além de atraso na aquisição de habilidades ao longo da infância, alterações de

tônus muscular, deformidades articulares, atrofia muscular e deficiência mental, assim sendo, não possuem indicação de serem avaliadas por testes de triagem, e sim instrumentos baseados em critérios específicos de cada condição²³. O tamanho da amostra foi calculado pelo software G* Power. Considerando $rpb > 0.7$, $\alpha = 0.05$ e poder estatístico de 80%, foram necessárias, no mínimo, 11 crianças por grupo etário para os estudos de confiabilidade e validade²⁵.

Dois examinadores, fisioterapeutas, também participaram do estudo. O Examinador I (J.T.) possui experiência na área de desenvolvimento infantil e realizou treinamento de 30 horas com o autor da tradução do Denver II no Brasil¹¹. O Examinador II (A.L.O.L.) recebeu treinamento teórico e prático. Durante as sessões de treinamento, os examinadores estudaram cuidadosamente o manual de treinamento, assistiram ao vídeo explicativo e realizaram exercícios práticos com crianças que não foram incluídas na amostra.

Instrumentos

O Denver II é um teste normativo que avalia a capacidade da criança de 0 a 6 anos em realizar atividades, esperadas para a idade, descritas em 125 itens distribuídos em quatro áreas do desenvolvimento infantil: pessoal-social (25 itens), motor fino-adaptativo (29 itens), motor grosso (32 itens) e linguagem (39 itens). Os itens são organizados em ordem crescente de dificuldade no formulário de aplicação e são administrados diretamente com a criança ou, alguns deles, podem ser respondidos pelos responsáveis^{11,13}. Cada item é representado por uma barra que indica as idades em que 25%, 50%, 75% e 90% das crianças da amostra padronizada conseguiram realizar a tarefa especificada^{11,13}. Para avaliação, deve-se traçar uma linha de cima a baixo no formulário de aplicação, na idade em que a criança se encontra, e administrar os itens intersectados pela linha, além de três itens à esquerda e itens à direita, até que sejam registradas três falhas consecutivas em cada área¹¹.

A pontuação do item é realizada conforme o comportamento observado, da seguinte forma: “P” para passou; “F” para falhou; “SO” para sem oportunidade e “R” para recusou. Posteriormente, cada item é classificado como: avançado (passar em um item completamente à direita da linha da idade), normal (passar ou falhar em um item entre as porcentagens 25-75%), cautela (falhar em um item entre as porcentagens 75-90%) ou atraso (falhar em um item à esquerda da linha

da idade)¹¹. Para os estudos de confiabilidade e validade concorrente, os escores qualitativos foram transformados em escores numéricos, sendo atribuído um ponto para os itens creditados como “P” e zero ponto para itens “F” ou “R”. Para determinar o escore em cada área e o escore total, foram somados os pontos de cada item avaliado, além de acrescentar um ponto para cada item à esquerda do menos complexo que a criança realizou^{15,26}. Para o estudo de sensibilidade e especificidade, o resultado final do teste foi interpretado conforme consta no manual, como: “Normal”, quando a criança não apresentou atrasos e, no máximo, uma cautela; “Suspeito”, quando a criança apresentou duas ou mais cautelas e/ou um ou mais atrasos nos itens¹¹.

O *Ages & Stages Questionnaires* terceira edição (ASQ-3) é um instrumento de triagem norte-americano que avalia o desenvolvimento global de crianças até os cinco anos de idade. É composto por 21 questionários, com 30 itens ou perguntas cada um, divididos em cinco áreas do desenvolvimento: comunicação, motor grosso, motor fino, resolução de problemas e pessoal-social^{27,28}. Cada item possui três possibilidades de resposta: “sim” (10 pontos), “às vezes” (5 pontos) ou “não” (0 pontos) e a soma destas respostas fornece o escore total para cada área²⁸. Cada área tem um ponto de corte de referência, sendo considerada “Normal” aquela criança que estiver acima ou no ponto de corte, e sendo considerada “Suspeita” a criança que pontuar abaixo do ponto de corte^{24,28}. A versão original do ASQ-3 apresenta parâmetros adequados de validade e confiabilidade, além de bons índices de sensibilidade e especificidade para identificar crianças com suspeita de atraso no desenvolvimento²⁷. No Brasil, o teste passou pelo processo de adaptação transcultural e a ser reconhecido por ASQ-3BR²⁴. Também teve suas propriedades psicométricas aferidas em uma amostra de 45.640 crianças em 468 creches e pré-escolas do Rio de Janeiro²⁴. O ASQ-3BR apresentou ótimas evidências psicométricas, com índices de correlação item-total razoáveis, consistência interna adequada e um bom ajuste ao modelo proposto²⁴. O ASQ-3BR apresenta baixo custo, é de fácil aplicação e avalia as mesmas áreas do desenvolvimento que o Denver II, justificando sua escolha para o estudo de verificação da validade concorrente, sensibilidade e especificidade. O ASQ-3BR pode ser administrado diretamente na criança ou em forma de questionário com os pais. Neste estudo, optou-se pela aplicação direta na criança uma vez que seria necessário garantir a precisão das avaliações.

Procedimentos

Inicialmente, foram coletadas, dos prontuários ou em entrevista com o responsável, informações sobre fatores biológicos, como sexo, idade gestacional (prematureo ou a termo), peso ao nascer (baixo peso <2500g; peso adequado >2500g) e fatores ambientais, como escolaridade da mãe e classe socioeconômica das famílias, determinada pelo Critério de Classificação Econômica Brasil²⁹.

Para verificação da confiabilidade do Denver II, 97 crianças, entre 1 e 72 meses, foram alocadas em 8 grupos etários: 1-6; 7-12; 13-18; 19-24; 25-36; 37-48; 49-59 e 60-72 meses. Para crianças nascidas prematuras, foi utilizada a idade corrigida até os 24 meses. Cada criança foi avaliada pelo examinador I, que aplicou e pontuou o Denver II, e pelo examinador II, que pontuou a criança por meio da observação da avaliação, de forma independente. Todas as avaliações foram gravadas com uma câmera (Handcam Sony, modelo DCR-SX43) fixada em um tripé. Para a confiabilidade intraexaminador, o examinador I realizou nova pontuação das avaliações assistindo às filmagens em um intervalo médio de 14 dias após a primeira avaliação³⁰.

Para verificação da validade concorrente, sensibilidade e especificidade, 248 crianças, entre 1 e 60 meses, foram distribuídas em sete grupos: 1-6; 7-12; 13-18; 19-24; 25-36; 37-48; 49-60 meses. Crianças entre 61 e 72 meses não foram incluídas, porque o ASQ-3BR, diferentemente do Denver II, não avalia crianças nessa faixa etária. As crianças foram avaliadas no mesmo dia pelo examinador I, que aplicou o Denver II, e pelo examinador II, que aplicou o ASQ-3BR, de forma independente. Os resultados obtidos pelos testes foram comparados entre as áreas de mesmo domínio, quais sejam: pessoal-social (Denver II) com pessoal-social (ASQ-3BR); motor fino-adaptativo (Denver II) com coordenação motora fina (ASQ-3BR); motor grosso (Denver II) com coordenação motora ampla (ASQ-3BR); linguagem (Denver II) com comunicação (ASQ-3BR). Os resultados da área resolução de problemas do ASQ-3BR não foram utilizados na análise.

Análise estatística

Os dados foram armazenados e analisados por meio do programa *Statistical Package for the Social Science* (SPSS) para Windows versão 20.0. O nível de significância adotado nos testes estatísticos foi $\alpha \leq 0,05$. Foi realizada estatística des-

critiva para caracterizar a amostra com cálculo das frequências absolutas, relativas, média e desvio-padrão.

A confiabilidade foi avaliada pelo *Intraclass Correlation Coeficiente* (ICC). Para confiabilidade de interexaminadores, foi realizada medida de concordância absoluta, com análise bidirecional entre os escores numéricos obtidos pelos examinadores I e II, para cada área do Denver II e entre as faixas etárias (ICC 2.1; two-wayrandom). Para a confiabilidade intraexaminador, foi utilizada medida de consistência, com análise bidirecional (ICC 3.1; two-waymixed) entre os escores das avaliações realizadas pelo examinador I (presencial e filmagem)³¹. Os resultados do ICC foram interpretados como: 0-0.25 pequena correlação; 0.25-0.50 regular; 0.50-0.75 moderada; >0.75 excelente correlação³¹.

Para o estudo de validade concorrente, foi verificada a correlação entre os escores numéricos das áreas correspondentes do Denver II e ASQ-3BR pelo teste de Spearman, uma vez que os dados não apresentaram normalidade. A correlação foi interpretada como: ≤ 0.19 muito fraca; 0.20–0.39 fraca; 0.40–0.59 moderada; 0.60–0.79 forte; > 0,80 muito forte³². Tabela de contingência 2x2 entre os resultados qualitativos (*Normal* ou *Suspeito*) dos instrumentos foi utilizada para calcular as seguintes propriedades: “sensibilidade”, capacidade do Denver II de identificar as crianças suspeitas; “especificidade”, capacidade do Denver II de identificar crianças normais; “valor preditivo positivo (VPP)”, proporção de verdadeiros suspeitos no Denver II entre todas as crianças identificadas como suspeitas no ASQ-3BR; “valor preditivo negativo (VPN)”, proporção de verdadeiros normais entre todas as crianças identificadas como normais no ASQ-3BR. Quanto maior a sensibilidade e especificidade mais confiável se torna o teste. Em testes ou exames de rastreamento, o mais adequado é aquele que apresenta alta especificidade, porque terá mais impacto no valor preditivo positivo. Ou seja, se o teste for positivo é muito provável que a pessoa presente, de fato, a condição investigada³³.

Aspectos éticos

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa em Seres Humanos da Universidade de Brasília (Protocolo nº 2.826.111/9358 4218.9.0000.0030). Os pais ou responsáveis das crianças participantes foram informados sobre os procedimentos do estudo e convidados a participar. Aqueles que voluntariamente concordaram

ram com a participação de seus filhos no estudo assinaram um Termo de Consentimento Livre e Esclarecido. Todas as crianças foram avaliadas na presença dos pais durante o período em que estiveram presentes nas instituições.

Resultados

Foram avaliadas 254 crianças com idade média de 20 meses (DP=18,3), maioria prematura, baixo peso ao nascer, de classe econômica média. Na amostra total, a avaliação pelo Denver II identificou 61.4% das crianças com suspeita de atraso no desenvolvimento, enquanto a avaliação com o ASQ-3BR identificou 69%. As características da amostra estão descritas na Tabela 1.

Confiabilidade intraexaminador e interexaminadores

As Tabelas 2 e 3 apresentam os resultados dos índices de confiabilidade intra e interexaminadores das quatro áreas do Denver II, nos grupos de idades e na amostra total. Todas as áreas do Denver II apresentaram resultados satisfatórios de confiabilidade na amostra total e na maior parte das faixas etárias. Apenas a faixa etária de 19-24 meses, no domínio pessoal social, apresentou confiabilidade interexaminadores moderada (ICC=0.68) (Tabela 3). Outros quatro resultados apresentaram confiabilidade adequada, porém os intervalos de confiança foram muito amplos, conforme consta nas Tabelas 2 e 3.

Validade concorrente

A Tabela 4 apresenta os resultados da correlação entre os escores brutos do Denver II e escores brutos do ASQ-3BR nos domínios correspondentes em cada faixa etária. Foram encontradas correlações muito fraca a forte entre 1 e 12 meses. A partir dos 13 meses, a correlação encontrada foi de moderada a muito forte.

Sensibilidade e especificidade

A Tabela 5 apresenta os valores de sensibilidade, especificidade, VPP e VPN do resultado qualitativo do Denver II com relação ao resultado do ASQ-3BR (Normal ou suspeito). Os valores da sensibilidade do Denver II variaram de 73% a 92%, valores de especificidade variaram de 58% a 92%, VPP de 55% a 89% e VPN 59% a 78%.

Tabela 1. Descrição da amostra dos estudos de confiabilidade, validade, sensibilidade e especificidade.

Fatores Biológicos	N (%) ou média (±DP)
Sexo	
Masculino	104 (57.9)
Feminino	107 (42.1)
Idade Gestacional	
A termo	92 (36.2)
Prematuro	162 (62.2)
Peso ao nascer	
Peso adequado	128 (50.4)
Baixo peso	126 (62.2)
Idade (meses)	20 (±18.3)
Fatores Ambientais	
Escolaridade materna	
Primário	29 (11.4)
Fundamental	44 (17.3)
Ensino médio	126 (49.6)
Superior	55 (21.7)
Classe Econômica	
Classe A	4 (1.6)
Classe B	13 (5.1)
Classe C	126 (49.6)
Classe D	56 (22.0)
Classe E	55 (21.7)

Legenda: N: Número de sujeitos; %: Frequência; DP: Desvio Padrão.

Fonte: Autores.

Discussão

O presente estudo demonstrou que a versão adaptada para o Brasil do Denver II apresenta adequada confiabilidade intra e interexaminadores, desde o nascimento até os seis anos de idade. O Denver II apresenta ainda bons índices de validade concorrente e índices satisfatórios de sensibilidade e especificidade na triagem de atrasos no desenvolvimento de crianças sob risco, principalmente entre 18 meses e 5 anos de idade. Os resultados desse estudo indicam que a versão brasileira apresenta semelhança com os resultados da versão original e revisada do teste¹².

O Denver II foi traduzido para diferentes idiomas, em diversos países e as propriedades psicométricas das versões traduzidas e adaptadas têm sido consistentemente verificadas^{14,16-20,26}. No Brasil, o processo de adaptação transcultural do Denver II gerou uma versão do teste semelhante à original, sendo trocado apenas o item referen-

Tabela 2. Resultado da confiabilidade intraexaminador dos domínios do Denver II por faixa etária e na amostra total.

Domínios do Denver II		Pessoal Social	Motor-fino-adaptativo	Linguagem	Motor grosso
Faixa Etária (meses)	N	ICC (IC)	ICC (IC)	ICC (IC)	ICC (IC)
01-06	17	0.98(0.94-0.99)	0.99(0.98-0.99)	0.97(0.92-0.99)	0.98(0.96-0.96)
07-12	11	0.96(0.87-0.99)	0.95(0.85-0.98)	0.99(0.96-0.99)	0.97(0.92-0.99)
13-18	15	0.99(0.99-0.99)	0.99(0.98-0.99)	0.98(0.95-0.99)	0.99(0.99-0.99)
19-24	10	0.99(0.97-0.99)	0.98(0.95-0.99)	0.99(0.98-0.99)	0.98(0.93-0.99)
25-36	10	0.99(0.98-0.99)	0.98(0.94-0.99)	0.96(0.96-0.99)	0.99(0.98-0.99)
37-48	10	0.99(0.97-0.99)	0.99(0.96-0.99)	0.99(0.99-0.99)	0.97(0.90-0.99)
49-60	11	0.87(0.58-0.96)*	0.98(0.93-0.99)	0.99(0.99-0.99)	0.96(0.88-0.99)
61-72	13	0.78(0.43-0.92)*	0.97(0.91-0.99)	0.96(0.90-0.99)	0.98(0.95-0.99)
Total (01-72)	97	0.99(0.99-0.99)	0.99(0.99-0.99)	0.99(0.99-0.99)	0.99(0.99-0.99)

Legenda: ICC: *Intraclass Correlation Coefficient*; IC: intervalo de confiança; N: número de sujeitos; *resultados com intervalos de confiança muito amplos.

Fonte: Autores.

Tabela 3. Resultado da confiabilidade interexaminadores dos domínios do Denver II por faixa etária e na amostra total.

Domínios do Denver II		Pessoal Social	Motor-fino-adaptativo	Linguagem	Motor grosso
Faixa Etária (meses)	N	ICC (IC)	ICC (IC)	ICC (IC)	ICC (IC)
01-06	17	0.97(0.92-0.99)	0.97(0.92-0.99)	0.97(0.92-0.99)	0.93(0.82-0.97)
07-12	11	0.93(0.78-0.98)	0.98(0.96-0.99)	0.79(0.41-0.94)*	0.96(0.86-0.98)
13-18	15	0.96(0.89-0.98)	0.97(0.91-0.99)	0.99(0.97-0.99)	0.98(0.96-0.99)
19-24	10	0.68(0.18-0.91)**	0.97(0.91-0.99)	0.98(0.94-0.99)	0.98(0.94-0.99)
25-36	10	0.99(0.99-0.99)	0.99(0.97-0.99)	0.78(0.34-0.94)*	0.99(0.96-0.99)
37-48	10	0.97(0.91-0.99)	0.98(0.95-0.99)	0.98(0.93-0.99)	0.98(0.96-0.99)
49-60	11	0.90(0.68-0.97)	0.98(0.88-0.99)	0.99(0.94-0.99)	0.98(0.95-0.99)
61-72	13	0.89(0.68-0.96)	0.95(0.85-0.98)	0.98(0.95-0.99)	0.97(0.92-0.99)
Total (01-72)	97	0.98(0.97-0.98)	0.99(0.98-0.99)	0.99(0.98-0.99)	0.99(0.99-0.99)

Legenda: ICC: *Intraclass Correlation Coefficient*; IC: intervalo de confiança; N: número de sujeitos; *resultados com intervalos de confiança muito amplos; **confiabilidade moderada.

Fonte: Autores.

te ao “preparo de cereal”, na área pessoal-social, que não corresponde a uma tarefa habitual no Brasil¹¹. Apesar de ter sido traduzido, foi mantida a mesma normatização da folha de pontuação original norte-americana, em que constam as barras que correspondem às idades de aquisição de cada habilidade, tornando necessária a verificação da aplicabilidade dessa versão nas crianças brasileiras¹¹. Nesse sentido, o grupo que traduziu o Denver II identificou, recentemente, em 2.755 crianças de 0 a 28 meses do Nordeste brasileiro, que o teste apresenta boa evidência de valida-

de de constructo, verificada por análise fatorial confirmatória e correlações dos resultados do Denver II com fatores de risco biológico e social dessa população¹⁵. Boo e colaboradores (2020) também identificaram bons índices de confiabilidade, assim como o presente estudo, e bons índices de consistência interna da versão brasileira do Denver II¹⁵.

Nossos resultados identificaram que a versão brasileira do Denver II, aplicada em toda a faixa etária de abrangência do teste, apresenta bons índices de confiabilidade, assim como a confia-

Tabela 4. Resultado da correlação entre os escores das áreas do Denver II e áreas correspondentes do ASQ-3BR por grupo etário.

Faixa Etária (meses)	N	PS/PS	MFA/MF	LING/ COM	MG/MG	Total
01-06	72	0.488	0.698	0.371	0.363	0.607
07-12	54	0.135	0.348	0.305	0.470	0.248
13-18	27	0.868	0.675	0.560	0.771	0.814
19-24	17	0.672	0.760	0.721	0.911	0.820
25-36	32	0.695	0.662	0.645	0.877	0.855
37-48	29	0.801	0.787	0.714	0.876	0.901
49-60	17	0.499	0.689	0.824	0.450	0.826

Legenda: ASQ-3BR: *Ages & Stages Questionnaires* – 3ª Edição verão brasileira; PS: Pessoal-Social; MFA: Motor-fino-adaptativo; MF: Motor Fino; LING: Linguagem; COM: Comunicação; MG: Motor Grosso.

Os resultados correspondem ao Coeficiente de Spearman (r), todos os resultados apresentaram nível de significância $p > 0.05$.

Fonte: Autores.

Tabela 5. Resultados da sensibilidade, especificidade, valor preditivo positivo e valor preditivo negativo do Denver II em relação ao ASQ-3BR.

Faixas Etárias (meses)	Sensibilidade	Especificidade	VPP	VPN
1-6	73%	62%	55%	59%
7-12	88%	76%	78%	75%
13-18	82%	66%	70%	62%
19-24	99%	87%	81%	77%
25-36	83%	58%	72%	58%
37-48	99%	92%	89%	78%
49-60	86%	66%	74%	63%
Total (0-60)	89%	78%	79%	78%

Legenda: ASQ-3BR: *Ages & Stages Questionnaires* – 3ª Edição verão brasileira; VPP: Valor preditivo positivo; VPN: valor preditivo negativo.

Fonte: Autores.

bilidade da versão traduzida e adaptada para a população iraniana ($ICC > 0.86$)¹⁶, da versão traduzida e adaptada na Etiópia ($ICC > 0.95$)¹⁴, da versão do Sri Lanka ($kappa > 0.80$)¹⁹ e a confiabilidade interexaminadores da versão espanhola ($kappa > 0.88$)²⁰. A confiabilidade é a capacidade do instrumento de reproduzir um resultado de forma consistente e livre de erros³⁴. Uma das justificativas para os bons índices de confiabilidade encontrados nos estudos, além da clareza e objetividade dos itens, pode ser o treinamento que foi realizado para os examinadores e reportado nos estudos^{15,16,26}. O Denver II é um teste de fácil aplicação e, mesmo que o examinador não seja experiente, quando há treinamento é possível obter bons índices de confiabilidade¹⁵.

Uma exceção encontrada no presente estudo de confiabilidade foi na área pessoal-social, que

apresentou confiabilidade moderada na faixa etária de 19 a 24 meses ($ICC = 0.68$). Outro estudo brasileiro identificou que a área pessoal social do Denver II apresentou a menor proporção de concordância entre os comportamentos registrados e os esperados para a idade em crianças de 24 e 60 meses da região de São Paulo¹⁰. Estudo de confiabilidade entre oito examinadores da versão espanhola do Denver II também identificou índices menores na área pessoal social ($ICC = 0.49$) em uma amostra de crianças entre 0 e 72 meses²⁰. Outros estudos de tradução e adaptação do Denver II apontaram para a necessidade de exclusão de alguns itens, principalmente na área pessoal social, devido à dificuldade de administração e interpretação pelos examinadores^{10,35}. Além disso, a área pessoal-social do Denver II é a que apresenta o maior número de itens obtidos pelo relato dos

pais, que pode ocasionar diferentes interpretações das respostas por parte dos examinadores, devendo ser aplicada com cautela para identificação precisa dos comportamentos investigados.

Além de ser confiável, um teste precisa também apresentar evidências dos diferentes tipos de validade, para se determinar o grau de confiança que pode ser dado às conclusões descritas por seu resultado³⁴. A validade de construto da versão brasileira do Denver II foi verificada por meio de análise fatorial confirmatória, cujos resultados demonstram que as cargas fatoriais dos itens do teste são estatisticamente significantes ($p < 0,01$) e apresentam excelentes índices nas quatro áreas do teste¹⁴. No presente estudo foi verificada a validade de critério (concorrente) da versão brasileira do Denver II, pela relação de seus escores com os escores do ASQ-3BR. Os resultados identificaram que a correlação variou de muito fraca a muito forte, com os menores índices nas faixas etárias de crianças mais novas, o que corrobora com outro estudo de validade concorrente do Denver II com relação às Escalas Bayley III²⁶. Essa grande variabilidade na correlação entre as faixas etárias nos escores obtidos pelo Denver II e ASQ-3BR pode ser justificada pela diferente forma de pontuação dos testes. No Denver II, os itens são organizados em uma ordem crescente de dificuldade e o formato de pontuação utilizado faz com que o escore aumente com a idade²⁶, enquanto no ASQ-3BR cada idade possui uma pontuação base e uma pontuação teto, independente do escore obtido nas faixas etárias anteriores²⁴.

De outro modo, por meio dos resultados qualitativos (normal ou suspeito) do Denver II e ASQ-3BR, este estudo identificou que o Denver II apresenta bons índices percentuais de sensibilidade e especificidade, principalmente com o avançar da idade. Resultado semelhante foi encontrado no estudo de acurácia da versão espanhola do Denver II, que apresentou índices de sensibilidade de 89%, especificidade 92%, VPP 0.91% e VPN 89%, quando comparado com avaliação e diagnóstico médico de atraso no desenvolvimento²⁰. A versão traduzida para o turco do Denver II também apresentou excelentes índices de sensibilidade (100%) e especificidade (95%), quando comparada com resultado de ultrassonografia craniana de crianças com suspeita de déficits neurológicos, acompanhadas até os dois anos de idade¹⁷, assim como a versão coreana, que apresentou bons índices de sensibilidade (83%) e índices aceitáveis de especificidade (51%) quando comparada a Escala Bayley de Desenvolvimento Infantil II¹⁸.

Considerando que o Denver II é um teste de triagem que tem por objetivo discriminar crianças que apresentam suspeita de atraso no desenvolvimento, a verificação de sua acurácia é de extrema relevância. Neste estudo, a aplicação do Denver II discriminou 61,4% das crianças com atraso no desenvolvimento, o que significa que as crianças apresentam atraso na aquisição de uma ou mais habilidades que já deveriam ter sido adquiridas na faixa etária esperada. Este resultado, provavelmente, se deve ao fato de os participantes terem um histórico de fatores de risco biológicos ou ambientais, como prematuridade, baixo peso, complicações gestacionais e menor nível socioeconômico que podem estar associados aos atrasos nas habilidades de linguagem, motoras e sociais⁷.

Utilizar testes de triagem válidos, precisos e confiáveis, como o Denver II, otimiza o acompanhamento do desenvolvimento infantil, a identificação precoce de atrasos ou alterações e o encaminhamento para serviços de reabilitação o mais breve possível, visando aproveitar ao máximo a neuroplasticidade, que é maior quanto mais nova for a criança^{36,37}. O Denver II é de fácil aplicação, baixo custo, avalia crianças em uma ampla faixa etária e pode ser aplicado por diferentes profissionais de saúde após treinamento e leitura do manual, o que facilita sua utilização nos serviços públicos de saúde do Brasil¹¹. Cabe ressaltar que o manual de treinamento da versão em português utiliza os pontos de corte da amostra normativa norte-americana, não tendo sido publicados dados normativos para a população brasileira. Desse modo, ao identificar que uma criança apresenta atraso no desenvolvimento infantil, pela aplicação do Denver II, o profissional precisa considerar os fatores contextuais nos quais esta criança está inserida. Pesquisas futuras poderão verificar a validade preditiva do Denver II, para complementação dos resultados encontrados nos estudos até o presente momento e definição de pontos de corte para atraso no desenvolvimento infantil da população brasileira.

Uma limitação deste estudo foi a utilização de duas escalas de triagem, Denver II e ASQ-3BR, para verificação da validade concorrente, sensibilidade e especificidade. A comparação com um teste de diagnóstico seria mais apropriada para avaliar esses parâmetros do Denver II. Contudo, não há um teste validado para a população brasileira que avalie a mesma faixa etária e mesmas áreas do desenvolvimento avaliadas pelo Denver II ou que seja reconhecido como padrão-ouro para uso com a população brasileira.

O presente estudo demonstrou que a versão Brasileira do Denver II, além de ser de fácil aplicação e baixo custo, apresenta excelente confiabilidade e boas evidências de validade concorrente, sensibilidade e especificidade quando utilizada em crianças sob risco de atraso no desenvolvimento, podendo ser utilizado na prática clínica e em pesquisas com crianças brasileiras. Estudos futuros são necessários para desenvolvimento de amostra normativa e pontos de corte para a população brasileira.

Colaboradores

JAT Santos realizou a coleta de dados, concepção e redação do manuscrito. KMA Ayupe fez a concepção e redação do manuscrito e revisão final. ALO Lima participou da coleta de dados e escrita do manuscrito. KA Albuquerque, FFR Morgado e PJBG Filho fizeram a revisão do manuscrito.

Agradecimentos

A FAP-DF pelo apoio financeiro concedido por meio da bolsa de iniciação científica.

Referências

1. Keenan T, Evans S, Crowley K. *An introduction to child development*. 3ª ed. Thousand Oaks: Sage; 2016.
2. Vélez-Agosto NM, Soto-Crespo JG, Vizcarrondo-Oppenheimer M, Vega-Molina S, García Coll C. Bronfenbrenner's bioecological theory revision: Moving culture from the macro into the micro. *Perspect Psychol Sci* 2017; 12(5):900-910.
3. Sameroff A. A unified theory of development: A dialectic integration of nature and nurture. *Child Dev* 2010; 81(1):6-22.
4. Sabanathan S, Wills B, Gladstone, M. Child development assessment tools in low-income and middle-income countries: how can we use them more appropriately? *Archives of Disease in Childhood* 2015; 100(5):482-488.
5. Santana RR, Silva MCL, Moreira TF, Lima ACD, Silva ÂCD. Fatores associados ao desenvolvimento motor de pré-escolares de uma escola pública de João Pessoa, Paraíba. *Rev Ter Ocupacional* 2017; 28(3):299-308.
6. Brasil. Ministério da Saúde (MS). Secretaria de Atenção à Saúde. *Diretrizes de estimulação precoce: crianças de zero a três anos com atraso no desenvolvimento neuropsicomotor decorrente de microcefalia*. Brasília: MS; 2016.
7. McDonald S, Kehler H, Bayrampour H, Fraser-Lee N, Tough S. Risk and protective factors in early child development: Results from the All Our Babies (AOB) pregnancy cohort. *Res Dev Disabil* 2016; 58:20-30.
8. Walker SP, Wachs TD, Grantham-McGregor S, Black MM, Nelson CA, Huffman SL, Baker-Henningham H, Chang SM, Hamadani JD, Lozoff B, Gardner JMM, Powel CA, Rahman A, Richter L. Inequality in early childhood: risk and protective factors for early child development. *The Lancet* 2011; 378(9799):1325-1338.
9. Neves KR, Souza RLM, Teixeira RA, Pinto PAF. Growth and development and their environmental and biological determinants. *J Pediatr* 2016; 92(3):241-250.

10. Pinto FCA, Isotani SM, Sabatés AL, Perissinoto J. Denver II: proposed behaviors compared to those of children from São Paulo. *Rev CEFAC* 2015; 17(4):1262-1269.
11. Sabatés AL. *Denver II: teste de triagem do desenvolvimento: manual de treinamento*. 1ª ed. São Paulo: Hogrefe; 2017.
12. Frankenburg WK, Dodds J, Archer P, Shapiro H, Bresnick B. The Denver II: a major revision and restandardization of the Denver Developmental Screening Test. *Pediatrics* 1992; 89(1):91-97.
13. Frankenburg WK. *Denver II: technical manual*. Denver Developmental Materials, Incorporated; 1996.
14. Abessa TG, Worku BN, Kibebew MW, Valy J, Lemmens J, Thijs H, Yimer WK, Kolsteren P, Granitzer M. Adaptation and standardization of a Western tool for assessing child development in non-Western low-income context. *BMC Public Health* 2016; 16(1):652.
15. Boo FL, Mateus MC, Sabatés AL. Initial psychometric properties of the Denver II in a sample from Northeast Brazil. *Infant Behav Dev* 2020; 58:1-17.
16. Shahshahani S, Vameghi R, Azari N, Sajedi F, Kazemnejad A. Validity and Reliability Determination of Denver Developmental Screening Test-II in 0-6 Year-Olds in Tehran. *Iran J Pediatr* 2010; 20(3):313-322.
17. Hallioglu O, Topaloglu AK, Zenciroglu A, Duzovali O, Yilgor E, Saribas S. Denver developmental screening test II for early identification of the infants who will develop major neurological deficit as a sequela of hypoxic-ischemic encephalopathy. *Pediatr Int* 2001; 43(4):400-404.
18. Shin H, Kwon B, Lim S. Validity of Korean Version of Denver II in Screening Children with Developmental Risk. *Korean J Child Health Nurs* 2005; 11(3):316-321.
19. Wijedasa D. Developmental screening in context: adaptation and standardization of the Denver Developmental Screening Test-II (DDST-II) for Sri Lankan children. *Child Care Health Dev* 2012; 38(6):889-899.
20. De-Andrés-Beltrán B, Rodríguez-Fernández ÁL, Güeita-Rodríguez J, Lambeck J. Evaluation of the psychometric properties of the Spanish version of the Denver Developmental Screening Test II. *Eur J Pediatr* 2015; 174(3):325-329.
21. Coster WJ, Mancini MC. Recomendações para a tradução e adaptação transcultural de instrumentos para a pesquisa e a prática em Terapia Ocupacional. *Rev Ter Ocup Univ São Paulo* 2015; 26(1):50-57.
22. Silva ML, Cavalcante LIC, Heumann S, Lima TVR. Relação entre gênero e desempenho neuropsicomotor de crianças em Belém, Brasil. *Cien Saude Colet* 2018; 23:2721-2730.
23. Albuquerque KA, Cunha ACB. New trends in instruments for child development screening in Brazil: a systematic review. *J Hum Growth Dev* 2020; 30(2):188-196.
24. Filgueiras A, Pires P, Maissonette S, Landeira-Fernandez J. Psychometric properties of the Brazilian-adapted version of the Ages and Stages Questionnaire in public child daycare centers. *Early Hum Dev* 2013; 89(8):561-576.
25. Pedrotti LG. *Tamanho de amostra e poder no Software R* [monografia]. Porto Alegre: UFRS; 2018.
26. Rubio-Codina M, Araujo MC, Attanasio O, Muñoz P, Grantham-McGregor S. Concurrent validity and feasibility of short tests currently used to measure early childhood development in large scale studies. *PLoS One* 2016; 11(8):1-17.
27. Squires J, Bricker DD, Twombly E. *Ages & stages questionnaires*. USA: Paul H. Brookes Baltimore, MD; 2009.
28. Instituto de Estudos do Trabalho e Sociedade (IETS). *Manual de uso do ASQ-3: Guia rápido para aplicação do ASQ-3*. Rio de Janeiro: IETS; 2010.
29. Associação Brasileira de Empresas de Pesquisa (ABEP). *Critério Brasil-Classificação Brasileira de Ocupações* [Internet]. São Paulo: ABEP; 2018 [acessado 2019 maio 16]. Disponível em: <http://www.abep.org/criterio-brasil>.
30. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, Vet HC. *COSMIN checklist manual*. Amsterdam: University Medical Center; 2012.
31. Perinetti G. Selection, interpretation and reporting of the intraclass correlation coefficient. *Eur J Orthod Dentofac* 2018; 5(1):3-5.
32. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use*. 3ª ed. Oxford: Oxford University Press; 2015.
33. Ferreira JC, Patino CM. Entendendo os testes diagnósticos. *J Bras Pneumol* 2017; 43(5):330-330.
34. Pasquali L. *Psicometria: Teoria dos testes na psicologia e na educação*. Petrópolis: Vozes Limitada; 2017.
35. Halpern R, Barros AJD, Matijasevich A, Santos IS, Victora CG, Barros FC. Estado de desenvolvimento aos 12 meses de idade de acordo com peso ao nascer e renda familiar: uma comparação de duas coortes de nascimentos no Brasil. *Cad Saude Publica* 2008; 24(Supl. 3):S444-S450.
36. Halle T, Zaslow M, Wessel J, Moodie S, Darling-Churchill K. Understanding and choosing assessments and developmental screeners for young children: Profiles of selected measures. Washington, DC: Office of Planning, Research, and Evaluation; 2011.
37. Halpern R. *Manual de pediatria do desenvolvimento e comportamento*. 1ª ed. São Paulo: Manole; 2015.

Artigo apresentado em 17/07/2020

Aprovado em 07/01/2021

Versão final apresentada em 09/01/2021

Editores-chefes: Romeu Gomes, Antônio Augusto Moura da Silva