

Concordância entre dois testes de triagem na avaliação da linguagem em crianças nascidas prematuras e de baixo peso

Paulo Roberto Bezerra de Mello ¹
Claudia de Souza Ozores Caldas ²
Sandra Coenga de Souza ³
Ageo Mário Cândido da Silva ⁴

^{1,3} Departamento de Pediatria. Universidade Federal de Mato Grosso. Av. Fernando Corrêa da Costa, 2367. Boa Esperança. Cuiabá, MT, Brasil. CEP: 78.068-600.
E-mail: prbmello@uol.com.br

² Secretaria Estadual de Saúde de Mato Grosso. Cuiabá, MT, Brasil.

⁴ Instituto de Saúde Coletiva. Universidade Federal de Mato Grosso. Cuiabá, MT, Brasil.

Resumo

Objetivos: avaliar a concordância dos resultados da linguagem de dois testes de triagem do desenvolvimento infantil - Denver II e Early Language Milestone Scale (ELM) em crianças de dois a três anos de idade cronológica, nascidas prematuras e com baixo peso.

Métodos: estudo observacional descritivo transversal através da aplicação dois instrumentos de triagem de desenvolvimento: Denver II e a ELM (Early Language Milestone Scale). A concordância entre o teste de Denver II e seu setor de linguagem e a ELM foi avaliada pelo coeficiente Kappa.

Resultados: das 77 crianças avaliadas, 36,3% tiveram prejuízo global do desenvolvimento pelo Denver II e 32,5% prejuízo da linguagem pela ELM. A concordância entre os resultados do teste de Denver II considerando todos os setores versus ELM mostrou coeficiente Kappa de 0,856 ($p < 0,001$) e considerando somente o setor da linguagem do Denver II versus ELM o coeficiente Kappa foi de 0,886 ($p < 0,001$).

Conclusões: o comprometimento do desenvolvimento observado nas crianças estudadas através da avaliação do Denver II e através do seu setor de linguagem mostrou concordância com as alterações de habilidades de linguagem observadas na ELM.

Palavras-chave Prematuro, Neonatologia, Desenvolvimento da linguagem, Coeficiente Kappa

Introdução

O aumento da sobrevivência de bebês prematuros nas últimas décadas tem aberto um desafio quanto ao seguimento destas crianças. Crianças nascidas prematuras e com baixo peso são consideradas de risco para atraso no desenvolvimento, em especial alterações de fala e linguagem. Os desvios de linguagem podem ser observados nas funções auditiva receptiva, auditiva expressiva e visual, assim como o domínio gramatical e o vocabulário por elas desenvolvido é significativamente menor. Os prejuízos no setor da linguagem podem causar problemas no aprendizado da leitura e da escrita, prejudicando o acompanhamento escolar e levando a problemas sociais e afetivos.^{1,2}

Testes de triagem do desenvolvimento infantil, sejam de desenvolvimento global ou específicos de linguagem, têm como finalidade identificar crianças que possam necessitar de uma avaliação diagnóstica mais específica do seu desenvolvimento.³ O uso de ferramentas de triagem do desenvolvimento padronizadas, quando comparado à observação clínica do desenvolvimento isolada, apresenta maior acurácia e favorece a intervenção precoce.¹ A escolha do instrumento dependerá das limitações e das características de aplicação do mesmo, assim como da população de estudo e objetivos a serem alcançados pelo profissional que os utilizar.⁴

O reconhecimento precoce de alterações no desenvolvimento, uma maior padronização e a validação em nosso meio de testes de avaliação de desenvolvimento, podem propiciar a melhor identificação e a inclusão de crianças em programas específicos de intervenção e estimulação.^{5,6}

O Denver II é usado e reconhecido internacionalmente, difundido em nosso país e recomendado pela Sociedade Brasileira de Pediatria para o acompanhamento do desenvolvimento infantil. É de uso rotineiro no seguimento de crianças, de fácil aplicação, baixo custo e rápido treinamento, podendo ser aplicado por profissionais de diferentes áreas que atendem a criança.^{7,8} Apesar de não ter sido validado para a criança brasileira, vem sendo utilizado em meios acadêmicos com adaptações para nossa população.^{5,9}

Divergências quanto à capacidade de identificar problemas na linguagem expressiva e problemas de articulação foram mencionadas na forma original do *Denver Developmental Screening Test* (Denver II).¹⁰ Já o Denver II difere do seu predecessor pela adição de 20 itens, a maioria dos quais englobam linguagem expressiva e destreza.¹¹ Pesquisas mostram ainda baixa especificidade (um número significativo de

crianças normais foi classificado como tendo atraso de desenvolvimento) e baixa sensibilidade (um número significativo de crianças com atrasos não foi classificado) deste teste.^{8,11} Alguns estudos apontam falhas na validade preditiva e sensibilidade do Denver II em detectar casos de atraso leve e, dentre os fatores que restringem a validade do teste em diferentes países, tem-se destacado a influência da cultura em alguns itens individuais.⁹

No Brasil, o teste de Denver II tem sido utilizado em estudos de avaliação do desenvolvimento de crianças, incluindo prematuras,^{6,12} e naqueles que abordam especificamente a linguagem.¹³ Apesar de seus autores chamarem atenção que conclusões de sua utilização devem vir de sua aplicação como um todo e não de suas partes,¹⁴ em nosso meio, diferentes estudos tem avaliado este teste de forma desmembrada, seja por seus itens,⁸ por seus setores, ou mesmo a partir da apresentação dos resultados de avaliação de setores isoladamente.¹³ Assim faz-se necessário investigar a adequação dos itens e o padrão de resposta em crianças brasileiras.⁹

Assim como o Denver II, a *Early Language Milestone Scale* (ELM) foi desenvolvida para ser um teste de triagem, mais especificamente para detecção de atrasos de fala e linguagem.¹⁵ É um instrumento de execução rápida, apresenta uma forma de fácil aplicação semelhante ao teste de Denver II, podendo ser utilizada nas formas de pontuação por escore ou por aprovação simples nos itens avaliados. Permite uma triagem detalhada de qual função da linguagem está comprometida, com marcadores agrupados em três áreas: função auditiva receptiva, função auditiva expressiva e função visual. A ELM é baseada largamente na observação dos pais; seus resultados são bons se os pais são informados e bons observadores; sua faixa de utilização vai de zero a 36 meses.¹⁶

A frequência de alterações de desenvolvimento pelo Denver II em crianças nascidas prematuras com idade entre dois e três anos foi mostrada em estudo anterior de nosso serviço, assim como frequências similares de alterações na linguagem por pela ELM.¹⁷ Considerando o predomínio de elementos de linguagem existente no Denver II nesta faixa etária, os resultados obtidos sugerem que as frequências próximas obtidas poderiam ser devido à capacidade similar dos dois testes de triarem alterações de linguagem em crianças. Desta forma este estudo tem como objetivo avaliar a concordância entre o setor de linguagem do teste de Denver II e a ELM em crianças nascidas prematuras em uma faixa etária de dois a três anos.

Métodos

Trata-se de um estudo epidemiológico analítico, prospectivo, de corte transversal, em crianças nascidas prematuras e com baixo peso, com idade cronológica entre dois e três anos incompletos.

As crianças foram triadas a partir do registro de internação de UTIs de quatro hospitais do município de Cuiabá (MT), região Centro-Oeste do Brasil. Destes hospitais, três atendiam ao Sistema Único de Saúde (SUS), e o quarto ao sistema de saúde complementar. Os quatro hospitais eram responsáveis por 95% dos partos da cidade.

No período de período de 20 de abril a 20 de outubro de 2011, 84 crianças nascidas prematuras (idade gestacional menor que 37 semanas) e com baixo peso (peso de nascimento menor que 2500g) foram localizadas por busca ativa no domicílio ou em seguimento em dois ambulatórios de prematuros do SUS.

Como critérios de inclusão foram considerados: prematuridade e baixo peso ao nascer, nascimento nos hospitais inseridos na pesquisa, apresentar idade cronológica entre dois a três anos incompletos e estar saudável no momento da avaliação. Como critérios de exclusão foram considerados a presença de malformações congênitas e/ou problemas neurológicos que afetassem a expressão da fala, alterações sensoriais auditivas e/ou visuais e sequelas de comprometimento do sistema nervoso central.

Com o objetivo de verificar a aplicabilidade e a viabilidade dos testes, foi realizado um estudo piloto com dez crianças. A coleta de dados foi realizada em sala para este fim, por um único pesquisador e o ambiente de aplicação dos testes foi satisfatório em todas as aplicações. Constatou-se de entrevista inicial com o responsável, aplicação do protocolo de estudo, revisão da caderneta da criança e posteriormente dos prontuários de internação do período neonatal. Foram coletadas informações descritivas da amostra: peso ao nascer, idade gestacional em semanas, condição de nascimento pelo escore de Apgar. Demais características da população estudada estão em Caldas *et al.*¹⁷

Para o Denver II, os critérios de normalidade por idade corresponderam àqueles propostos nos originais do mesmo elaborado a partir de uma população de crianças americanas, segundo adaptação cultural do teste para a população local de Cuiabá-MT.¹⁸ Para a avaliação dos setores do Denver II, na faixa etária de dois a três anos de idade, foram utilizados os seguintes itens interceptados pela linha da idade traçada para cada participante, de acordo com sua

idade cronológica:

- Setor pessoal social: tira vestuário, alimenta boneca, veste roupa, escova os dentes com ajuda, lava e seca as mãos, nomeia amigos, põe camiseta, veste-se sem ajuda, joga cartas, escova os dentes sem ajuda, prepara alimentos;

- Setor motor fino-adaptativo: ergue torre com quatro cubos e com seis cubos, imita linha vertical, ergue torre com oito cubos, sacode polegar, copia círculo, desenha pessoas com três partes, copia “cruz”, pega uma linha comprida;

- Setor linguagem: aponta duas figuras, combina palavras, nomeia uma figura, enuncia seis partes do corpo, aponta quatro figuras, fala metade entendível, nomeia quatro figuras, sabe duas ações, sabe adjetivos, nomeia uma cor, uso de dois objetos, conta um bloco, uso de três objetos, sabe quatro ações, fala tudo entendível, entende quatro preposições, nomeia quatro cores, conhece três adjetivos;

- Setor motor grosso: chuta bola para frente, salta, arremessa bola sobre os ombros, pula largo, balança o pé por um segundo, balança o pé por dois segundos, balança o pé por três segundos.

Cada item foi avaliado pelo histograma cujos limites correspondem às idades de início e extremo para que a criança desempenhe tal habilidade. O limite esquerdo corresponde ao percentil 25 (P25) da idade de realização da habilidade e o limite direito do histograma corresponde ao percentil 90 (P90). Entendia-se como “falha” quando a criança não realizava a prova e sua linha da idade permanecia entre os percentis 25 (P25) e 75 (P75) e como cautela ou atenção quando a criança falhava em uma prova que era interceptada pela linha da idade entre o percentil 75 (P75) e percentil 90 (P90). O P90 corresponde ao ponto de corte utilizado no teste de Denver II para definir atraso, ou seja, quando a criança falhava em um item ou prova que ficava totalmente à esquerda da linha da idade.

A classificação do desempenho geral no teste de Denver II foi feita de acordo com o número de falhas, considerando os atrasos e cautelas, sendo considerado como:

- Anormal: quando a criança avaliada apresenta dois ou mais atrasos independente da área ou setor;

- Suspeita: quando a criança avaliada apresenta um atraso e/ou duas ou mais cautelas;

- Normal: quando a criança avaliada não apresentava nenhum atraso e no máximo uma cautela.

Para análise deste estudo foram considerados os resultados da análise global do Denver II (suspeitos e alterados). Para análise do setor linguagem isoladamente foram consideradas alterações a existência de pelo menos um atraso ou cautela do setor.¹⁸

Como referência de triagem de linguagem foi utilizada a escala *Early Language Milestone Scale* segundo recomendação dos autores,¹⁹ considerando os padrões de normalidade por idade do teste e os marcos da linguagem agrupados nas funções auditiva receptiva (AR), função auditiva expressiva (AE) e função visual (V).

Na ELM os comportamentos são também dispostos na forma de histogramas, em folha única, e distribuídos em 36 meses, permitindo localizar dentro das funções as idades de início e limite de desempenho de cada item. O gráfico indica os valores de 25%, 50%, 75% e 90% como representativos da porcentagem de crianças de determinada faixa etária que atingiram a habilidade testada durante o processo de validação da escala.

Como no Denver II traçava-se uma linha vertical em toda a escala, correspondendo à idade cronológica da criança na data da avaliação. Em seguida foram avaliados todos os itens que interceptados por esta linha vertical em cada uma das áreas AR, AE e V para se determinar os níveis de teto e base.

Os comportamentos da ELM referentes à idade de 2 a 3 anos são distribuídos de acordo com as seguintes funções:

- Auditiva expressiva: fala de quatro a seis palavras únicas, mais de 50 palavras únicas, fala eu/você, usa preposições, conversação, nome e uso de objetos (copo, bola, colher, lápis);

- Auditiva receptiva: segue ordem de dois comandos sem gesto, aponta objetos nomeados, aponta objetos descritos pelo uso, ordens/ noção espacial;

- Visual: segue ordem de um comando com gesto, inicia jogos gestuais, aponta objetos desejados (três últimos itens até os 18 meses de idade).

O resultado da ELM foi considerado adequado para a área avaliada quando a criança obtinha, de forma sequencial, até três itens de sucesso correspondentes à idade cronológica. Foi considerado como falha ou alterado, quando a criança não realizou uma prova situada totalmente à esquerda da linha da idade, ou seja, acima do P90 e foram em sequência identificados mais três falhas consecutivas na área avaliada.²⁰ Foram considerados também para análise os percentuais de itens falhos ou alterados de acordo com as funções isoladamente: auditiva expressiva, auditiva receptiva e visual.

Quando o desempenho em um dos instrumentos estava alterado, foram realizados re-testes no prazo de até duas semanas. Foram considerados o relato dos pais e a observação direta ou incidental do comportamento avaliado.

Para o cálculo da amostra foram considerados o número e a proporção de crianças nascidas prematuras com baixo peso, no ano de 2009 (ano de nascimento das crianças) em Cuiabá-MT, por residência da mãe, de acordo com os dados do Sistema de Informações sobre Nascidos Vivos (Sinasc), sendo detectadas 411 crianças nessas condições. Com base na média dos resultados obtidos em estudos,^{21,22} o percentual de crianças nascidas prematuras e com baixo peso que apresentaram alterações no desenvolvimento da fala e linguagem foi de, aproximadamente 21%, percentual este considerado para fins de cálculo da amostra. Considerando um erro amostral de 7%, a amostra calculada resultou em 75 crianças.

As análises foram realizadas com o pacote estatístico SPSS, versão 20. Para comparação das frequências dos resultados dos testes estudados foi utilizado o teste Qui-Quadrado de Pearson. A intensidade da concordância entre os resultados do Denver II da ELM, assim como a análise específica do setor de linguagem do teste de Denver II e a ELM foi avaliada pelo coeficiente *Kappa*.²³ Considerou-se estatisticamente significativa as concordâncias com $p < 0,05$.

Este trabalho foi aprovado pelo Comitê de Ética em Pesquisa do Hospital Universitário Júlio Müller da Universidade Federal de Mato Grosso, protocolo no 967/2010 e parecer 1.141.638 de 30/07/2015. As crianças com desempenho alterado foram encaminhadas para atendimento especializado.

RESULTADOS

No período do estudo foram avaliadas 84 crianças nascidas prematuras e com baixo peso. Destas foram excluídas quatro crianças: uma com diagnóstico de hidrocefalia e três com paralisia cerebral. Outras três crianças não compareceram para realizar os re-testes, sendo consideradas como perda de seguimento, permanecendo 77 crianças na amostra final.

Das crianças estudadas 58% eram do sexo masculino, 50% eram prematuros com idade gestacional maior ou igual a 34 semanas, 21% tinham peso ao nascer menor que 1 500 g e 7% escore de Apgar menor que 7 no quinto minuto.

As Tabelas 1, 2 e 3 mostram as frequências de alterações do Denver II e da ELM, assim com das alterações dos setores e das funções dos dois testes respectivamente, segundo presença ou não de alterações. Não houve diferença entre o percentual de crianças com resultados alterados do Denver II comparado com os resultados da ELM (Tabela 1). No Denver II observa-se que a maior frequência de atrasos foi no setor de linguagem, comparado aos

setores pessoal-social, motor fino-adaptativo e motor grosso (Tabela 2). Em relação às funções da ELM, todas as crianças que mostraram alteração na linguagem pela ELM apresentavam alteração na área auditiva expressiva (32,5%), e esta alteração foi mais frequente que as observadas na área na área auditiva receptiva (18,2%). A função visual, que apresenta itens para avaliação até os 18 meses, foi

normal em todas as crianças (Tabela 3).

A Tabela 4 mostra as concordâncias obtidas nas crianças estudadas, a partir da comparação dos resultados do Denver II com a ELM e do setor de linguagem do Denver II com a ELM. Ambos os coeficientes *Kappa* foram significantes e se encontram na faixa considerada ótima para este escore (0,81 a 0,99).

Tabela 1

Frequência das alterações de desempenho no Denver II, no setor de linguagem do Denver II e na ELM nas 77 crianças avaliadas.

	Normal		Alterado		Total	
	n	%	n	%	n	%
Denver II (1)	49	63,6	28	36,3	77	100,0
Denver II Setor linguagem (2)	48	62,3	29	37,7	77	100,0
ELM (3)	52	67,5	25	32,5	77	100,0

Comparações: χ^2 : 1x3 (ns); 2x3 (ns).

Tabela 2

Frequências das alterações do desempenho nos setores do Denver II nas 77 crianças avaliadas.

	Normal		Alterado		Total	
	n	%	n	%	n	%
Linguagem (1)	48	62,3	29	37,7	77	100,0
Pessoal-social (2)	68	88,3	9	11,7	77	100,0
Motor fino (3)	73	94,8	4	5,2	77	100,0
Motor grosso (4)	72	93,5	5	6,5	77	100,0

Comparações: χ^2 : 1x2 $p < 0,001$; 1x3 $p < 0,001$; 1x4 $p < 0,001$
 χ^2 : 2 x 3 (ns); 2x4 (ns); 3x4 (ns).

Tabela 3

Frequência das alterações do desempenho nos setores da ELM nas 77 crianças avaliadas.

	Normal		Alterado		Total	
	n	%	n	%	n	%
Auditiva expressiva (1)	52	67,5	25	32,5	77	100,0
Auditiva receptiva (2)	63	81,8	14	18,2	77	100,0
Visual (3)	77	100,0	0	-	77	100,0

Comparações: χ^2 : 1x2 ($p < 0,05$).

Tabela 4

Concordância entre o desempenho no Denver II e na ELM das 77 crianças estudadas.

Desempenho do Denver II	Desempenho na ELM				Kappa (p)
	Alterado		Adequado		
	n	%	n	%	
Alterado	24	85,7	4	14,3	0,856 (<0,001)
Adequado	1	2,0	48	98,0	

Concordância entre o desempenho no setor de linguagem do Denver II e na ELM das 77 crianças estudadas.

Desempenho no setor de Linguagem do Denver II	Desempenho na ELM				Kappa (p)
	Alterado		Adequado		
	n	%	n	%	
Alterado	25	86,2	4	13,8	0,886 (<0,001)
Adequado	0	0,0	48	100,0	

Discussão

A prevalência de alterações em ambos os testes utilizados foi semelhante nas crianças estudadas. Da mesma forma predominaram as alterações de linguagem no teste de Denver II, em percentual semelhante ao de crianças triadas positivamente para linguagem pela ELM. Já as alterações de linguagem expressiva pela ELM foram mais frequentes. E valorizando a presença de cautelas e atrasos no Denver II e no seu setor de linguagem, ao se fazer a comparação com a ELM, a concordância global entre os testes aplicados demonstrou valores significativos do coeficiente *Kappa*.²³

A triagem dos atrasos de linguagem é o método mais efetivo para identificar distúrbios de linguagem. Estudos iniciais^{16,22} já haviam discutido as semelhanças e as características dos dois métodos de triagem aqui abordados e observaram diferentemente, que a utilização da ELM era um instrumento de maior sensibilidade e especificidade para avaliar o desenvolvimento da linguagem de crianças consideradas de risco.

Poucos estudos fazem comparação entre o Denver II e a ELM. O'Hara *et al.*²⁴ avaliaram crianças em sistema de adoção temporária na faixa de idade de zero a 18 meses pelos dois métodos. Estes autores²⁴ observaram que 35% de suas crianças falharam no Denver II, sendo que o maior percentual delas (35%) no componente da linguagem. Usando o modo de pontuação por escore da ELM, o percentual de falha, por este método, foi

de apenas 8%. As divergências observadas, em relação ao aqui utilizado, poderiam ser atribuídas à faixa etária diferente e à variação na aplicação da ELM feita por escore.

Revisões sobre alterações de linguagem em crianças nascidas prematuras até dois anos de idade ressaltaram a frequência de alterações com predomínio significativo de alterações da linguagem expressiva nestas crianças. Embora não detectáveis aos 12 meses, estas alterações podem surgir aos dois anos, idades próximas àquelas avaliadas neste estudo.^{6,25}

O encontro de alterações no desenvolvimento pelo Denver II em crianças que foram prematuras ao nascer, aqui observado, é coincidente com o achado em outras publicações, assim como o comprometimento da linguagem na área auditiva expressiva e em menor escala na área auditiva receptiva através da triagem pela ELM.^{1,4,9}

Distúrbios de comunicação nas linguagens receptiva e expressiva em pré-termos extremos, não têm sido atribuídos a déficits neurológicos ou sensoriais, mas a diminuição geral da função mental nestes pacientes.⁴ De forma semelhante, se justificam achados na linguagem receptiva à imaturidade nas habilidades de atenção e em tarefas que envolvam duração e direcionamento ao foco atencional. Além disso, as alterações em linguagem expressiva poderiam estar associadas a fatores biológicos dessas crianças ou a uma estimulação ambiental inadequada.^{2,26}

O Denver II difere do seu predecessor pela

adição de 20 itens, a maioria dos quais englobam linguagem expressiva e destreza.¹¹ Nos itens analisados na população estudada, ou seja, para a faixa etária dos participantes, 18 dos 45 itens do Denver II avaliados são de linguagem dos quais 16 são de linguagem expressiva. Esta ênfase pela linguagem expressiva neste período de utilização do Denver II provavelmente favoreceu a concordância observada, considerando que também o maior percentual de alteração observado com a aplicação da ELM se deu neste mesmo setor da linguagem.

Os pacientes estudados tinham, ao exame, entre 2 e 3 anos de idade cronológica. Isto não permitiu que fossem avaliados quanto ao elemento visual da linguagem que está incorporado na escala ELM até os 18 meses de idade. Isto provavelmente não interferiu na aplicação do teste, pois todas as crianças estudadas preenchiam o requisito de estar na faixa de aplicação do mesmo e cumpriram os três últimos elementos de avaliação visual. A forma da ELM utilizada neste estudo foi por aprovação nos itens. Esta forma tem aplicação mais rápida, minimiza os falsos negativos, mas é menos específica e pode levar a um aumento discreto nos falso positivos.¹⁵

Testes de triagem alertam para um risco potencial que pode ser confirmado através de um acompanhamento sistemático e mediante testes diagnósticos. Além disso, a avaliação do desenvolvimento em um único momento não permite que se determine de forma definitiva um atraso no desenvolvimento da criança, mas indica a necessidade de investigação mais cuidadosa e profunda.^{8,27} Considerando a ênfase da detecção de atrasos do desenvolvimento incluindo a linguagem, através de rastreamento até os dois anos de idade,^{1,3} o encontro de alterações em dois métodos de triagem para alterações do desenvolvimento em crianças que foram prematuras enfatiza a importância de manter o seguimento dessas crianças de risco.

Ambos os testes aqui utilizados não estão validados para a língua portuguesa. Desta forma, não se tem conhecimento sobre as medidas de reprodutibilidade ou padrões de referência para nossas crianças. A ampla utilização destes testes em nosso meio, incluindo seu desmembramento por setores,⁹ justifica a análise mesmo que parcial da sua utilização como a efetuada neste estudo e ressaltam a necessidade de uma maior normatização para fundamentar investigações posteriores.

Estes achados em uma população de risco não podem em princípio ser generalizados para a população geral.²⁸ Embora a ênfase recaia sobre o setor de linguagem, o mais acometido neste grupo de população,¹⁷ limitações do estudo podem ter

influência na concordância, como a homogeneidade da amostra, sono e fadiga da criança e o fato de ambos os testes terem sido aplicados em sequência pela mesma pessoa.

O conhecimento dos resultados dos resultados do teste de referência (ELM) e do teste índice (setor de linguagem do teste Denver II) pela aplicadora dos mesmos pode ter influenciado a interpretação dos dois testes, provavelmente aumentando a concordância entre os resultados e provocando distorção na medida de acurácia. Isto pode ter resultado na ocorrência de um viés de revisão – que consiste em não avaliar de forma mascarada os testes realizados para outros testes e para desfechos.²⁸

Contudo, a aplicação de forma uniforme dos dois testes e a existência da predominância de elementos de linguagem expressiva de formulações diferentes permite inferir que a concordância observada na amostra reflete, provavelmente, muito mais a concordância entre os testes, do que propriamente o efeito do viés decorrente da aplicação por um único examinador.

O coeficiente *Kappa*, usado para dados categóricos, constitui um avanço em relação à taxa geral de concordância, por ser um indicador ajustado que leva em consideração a concordância devida à chance. Sua significância permite a realização de inferência estatística, mas não traz informação sobre a qualidade da medida realizada pelos observadores ao estimar apenas a concordância entre as observações efetuadas.²³

Já a avaliação de acurácia de um teste baseia-se na sua relação com algum meio de saber se a doença está ou não realmente presente - um indicador mais fiel da verdade geralmente referido com “padrão ouro”.²⁹ Esta avaliação não foi realizada no presente estudo, porque os dois testes comparados, apesar de sua ampla utilização em nosso meio, não foram validados para a população brasileira.

Neste estudo é reforçada a ênfase ao elemento linguagem dentre os elementos triados pelos métodos utilizados. Embora usando instrumentos de triagem, que possuem vantagens e desvantagens e sabendo que fatores relacionados ao ponto de corte de métodos de triagem podem contribuir para dificultar a diferença entre doentes e não doentes,³⁰ este estudo procura contribuir para verificar a validade de utilização desses instrumentos em nosso país.

Parece reforçar, em particular, o uso do teste Denver II como método de triagem em nosso meio, e o valor de seu setor de linguagem, ao mostrar sua ampla concordância com o outro método de triagem específico para linguagem.

De acordo com os dados deste trabalho, pode-se

afirmar que há uma forte concordância entre dois métodos de triagem do desenvolvimento - Denver II e ELM, em particular no item linguagem, e sugere que o Denver II seja capaz de triar desvio de desenvolvimento em uma população com predomínio de

alterações de linguagem na faixa etária de dois a três anos. Somente estudos com maior aprofundamento metodológico permitiriam concluir sobre a acurácia das sugestões aqui propostas.

Referências

1. Isotani SM, Azevedo MF, Chiari BM, Perissinoto J. Linguagem expressiva de crianças nascidas pré-termo e termo aos dois anos de idade. *Pró-Fono R Atual Cient.* 2009; 21 (2): 155-60.
2. Ribeiro LA, Zachrisson HD, Schjolberg S, Aase H, Rohrer-Baumgartner N, Magnus P. Attention problems and language development in preterm low-birth-weight children: Crosslagged relations from 18 to 36 months. *BMC Pediatrics.* 2011; 11: 59.
3. Santos RS, Araújo APQC, Porto MAS. Diagnóstico precoce de anormalidades no desenvolvimento em prematuros: instrumentos de avaliação. *J Pediatr (Rio J).* 2008; 84 (4): 289-99.
4. Lamônica DAC, Carlino FC, Alvarenga KF. Avaliação da função auditiva receptiva, expressiva e visual em crianças prematuras. *Pró-Fono R Atual Cient.* 2010; 22 (1): 19-24.
5. Pinto FCA, Isotani SM, Sabatés AL, Perissinoto J. Denver II: comportamentos propostos comparados aos de crianças paulistanas. *Rev. CEFAC.* 2015; 17 (4): 1262-9.
6. Zerbeto AB, Cortelo FM, Filho EB. Association between gestational age and birth weight on the language development of Brazilian children: a systematic review. *J Pediatr (Rio J).* 2015; 91: 326-32.
7. American Academy of Pediatrics. Committee on Children with Disabilities. Developmental Surveillance and Screening of Infants and Young Children. *Pediatrics.* 2001; 108 (1): 192-6.
8. Sigolo ARL, Aiello ALR. Análise de instrumentos para triagem do desenvolvimento infantil. *Paidéia.* 2011; 21(48): 51-60.
9. Magalhães LC, Fonseca KL, Martins LDTB, Dornelas LF. Desempenho de crianças pré-termo com muito baixo peso e extremo baixo peso segundo o teste Denver-II. *Rev Bras Saúde Matern Infant.* 2011; 11 (4): 445-53.
10. Borowitz KC, Glascoe FP. Sensitivity of Denver Developmental Screening Test in speech and language screening. *Pediatrics.* 1986; 78: 1075-8.
11. Glascoe FP, Byrne KE, Ashford LG, Hohnson MA, Chang B, Strickland B. Accuracy of the Denver-II in developmental screening. *Pediatrics.* 1992; 89 (6): 1221-5.
12. Custódio ZAO, Crepaldi MA, Cruz RM. Desenvolvimento de Crianças Nascidas Pré-Termo Avaliado pelo Teste de Denver-II: Revisão da Produção Científica Brasileira. *Psicol Reflex Crít.* 2011; 25(2), 400-6.
13. Mendes JCP, MM2, Carabetta Júnior V, Ferreira Novo NF, Colombo-Souza P. Fatores associados a alteração da linguagem em crianças pré-escolares. *Rev Soc Bras Fonoaudiol.* 2012; 17 (2): 177-81.
14. Frankenburg WK, Dodds JB, editors. Denver II training manual. Denver: Denver Developmental Materials Inc.; 1990.
15. Coplan J, Gleason JR. Test-retest and interobserver reliability of the Early Language Milestone Scale, second edition. *J Pediatr Health Care.* 1993; 7: 212-9.
16. Walker D, Gugenheim S, Downs MP, Northern JL. Early Language Milestone Scale and language screening of young children. *Pediatrics.* 1989; (83): 284-8.
17. Caldas CO, Takano OA, Mello PRB, Souza SC, Zavala AA. Language abilities performance of children born preterm and low birth weight and associated factors. *Audiol Commun Res.* 2014; 19 (2): 158-66.
18. Souza SC, Leone C, Takano OA, Moratelli HB. Desenvolvimento de pré-escolares na educação infantil em Cuiabá, Mato Grosso, Brasil. *Cad Saúde Pública.* 2008; 24(8): 1917-26.
19. Coplan J. The Early Language Milestone Scale. 2 ed. Texas: Pro-Ed; 1993.
20. Gejão MG, Lamônica DAC. Habilidades do desenvolvimento em crianças com hipotireoidismo congênito: enfoque na comunicação. *Pró-Fono R Atual Cient.* 2008; 20 (1): 25-30.
21. Brito CML, Vieira GO, Costa COM, Oliveira NF. Desenvolvimento neuropsicomotor: o teste de Denver na triagem dos atrasos cognitivos e neuromotores de pré-escolares. *Cad Saúde Pública.* 2011; 27 (7): 1403-14.
22. Glascoe FP, Martin ED, Humphrey S. A comparative review of developmental screening tests. *Pediatrics.* 1990; 86 (4): 547-54.
23. Lorca JC, Pino LAV. Evaluación de la concordancia inter-observador em investigación pediátrica: Coeficiente de Kappa. *Rev Chil Pediatr.* 2008; 79 (1): 54-8.
24. O'Hara MT, Church CC, Blatt SD. Home base developmental screening of children in foster care. *Pediatric Nursing.* 1998; 24 (2): 113-7.
25. Tanner AM. "The Effects of Premature Birth on Language Development" (2012). Research Papers. Paper 250. [acesso em 18 mar 2016]. Disponível em http://opensiuc.lib.siu.edu/gp_rp/250.
26. Van Lierde KM, Roeyers H, Boerjan S, De Grootte I. Expressive and receptive language characteristics in three-year-old preterm children with extremely low birth weight. *Folia Phoniatr Logop.* 2009; 61: 296-9.
27. Zeppone SC, Volpon LC, Del Ciampo LA. Monitoramento do desenvolvimento infantil realizado no Brasil. *Rev Paul Pediatr.* 2012; 30 (4): 594-9.
28. Souza MR, Ribeiro ALP. Revisão sistemática e meta-análise de estudos de diagnóstico e prognóstico: um tutorial. *Arq Bras Cardiol.* 2009; 92 (3): 241-51.
29. Fletcher RH, Fletcher SW, Wagner EH. *Epidemiologia Clínica: Elementos essenciais.* 3 ed. Porto Alegre: Artes Médicas; 1996.

30. Moura DR, Costa JC, Santos IS, Barros AJD, Matijasecich A, Halpern R, Dumith S, Karam S, Barros FC. Risk factors for suspected developmental delay at age 2 years in a Brazilian birth cohort. *Paediatr Perinat Epidemiol.* 2010; 24: 211-21.

Recebido em 19 de novembro de 2015

Versão final apresentada em 17 de junho de 2016

Aprovado em 1 de setembro de 2016