

# Confiabilidade na Aplicação das Escalas Bayley: Avaliando Crianças Acometidas pelo Zika Vírus

Tamiles Cerqueira Lopes da Silva<sup>1</sup>, George Anderson Alves dos Santos<sup>1</sup>, Leticia Marques dos Santos<sup>2</sup>, Vanessa Madaschi<sup>3</sup>, Andrea Perosa Saigh Jurdi<sup>4</sup>, Darci Neves dos Santos<sup>1</sup>

<sup>1</sup> Instituto de Saúde Coletiva, Universidade Federal da Bahia

<sup>2</sup> Instituto de Humanidades, Artes e Ciências, Universidade Federal da Bahia

<sup>3</sup> Universidade Presbiteriana Mackenzie

<sup>4</sup> Universidade Federal de São Paulo

**Submissão:** 2 abr. 2021.

**Aceite:** 20 set. 2023.

**Editor de seção:** Ana Alexandra Caldas Osório.

## Nota dos Autores

Tamiles Cerqueira Lopes da Silva  <https://orcid.org/0000-0002-4769-640X>

George Anderson Alves dos Santos  <https://orcid.org/0000-0002-5600-5287>

Leticia Marques dos Santos Silva  <https://orcid.org/0000-0001-5963-2166>

Vanessa Madaschi  <https://orcid.org/0000-0002-6954-4407>

Andrea Perosa Saigh Jurdi Silva  <http://orcid.org/0000-0002-1111-5562>

Darci Neves dos Santos  <https://orcid.org/0000-0003-1500-8294>

Correspondências referentes a este artigo devem ser enviadas a Tamiles Cerqueira Lopes da Silva, Rua Basílio da Gama, Canela, Salvador – BA, Brasil. CEP: 40110-040. Email: [tamiles.live@hotmail.com](mailto:tamiles.live@hotmail.com)

### Resumo

Entender e mensurar determinantes que influenciam o desenvolvimento infantil, poderá contribuir para um melhor direcionamento dos cuidados de saúde na primeira infância. Utilizaram-se as Escalas Bayley III para avaliar o desenvolvimento de crianças com e sem diagnóstico da síndrome congênita do zika vírus, participantes de um estudo longitudinal avaliados em domicílio. Examinou-se a confiabilidade interobservadores em três pontos de avaliação, descrevendo também o processo de treinamento e supervisão da equipe interdisciplinar, para padronização dos procedimentos de aplicação do instrumento na coleta de dados. Produziram-se medidas de confiabilidade pelo índice kappa e coeficiente de correlação intraclass (CCI). Os valores médios de kappa corresponderam a 0,92, 0,89 e 0,96, respectivamente, para a primeira, segunda e terceira medidas de confiabilidade entre aplicadores. O CCI se manteve acima de 90% para cada subescala avaliada nas três medidas realizadas. Os resultados demonstram excelentes indicadores de confiabilidade na aplicação das escalas, sugerindo a importância do treinamento e supervisão da equipe para conferir um padrão de confiabilidade interobservadores da avaliação de crianças com transtornos do neurodesenvolvimento em estudos populacionais.

**Palavras-chave:** medidas de confiabilidade, transtornos do neurodesenvolvimento, Escalas Bayley III, precisão do teste, avaliação neuropsicológica, desenvolvimento infantil

### RELIABILITY OF THE BAYLEY SCALE: ASSESSING CHILDREN AFFECTED BY ZIKA VIRUS

#### Abstract

Understanding and measuring the determinants that influence child development could help to better target early childhood health care. The Bayley III scales were used to assess the development of children with and without a diagnosis of zika virus congenital syndrome, participating in a longitudinal study and evaluated at home. Inter-observer reliability was examined at three assessment points, and the process of training and supervising the interdisciplinary team was also described, in order to standardize the procedures for applying the instrument in data collection. Reliability measures were produced using the kappa index and the intraclass correlation coefficient (ICC). The average kappa values were 0.92, 0.89 and 0.96, respectively, for the first, second and third measures of inter-rater reliability. The ICC remained above 90% for each sub-scale evaluated in the three measurements carried out. The results show excellent indicators of reliability in the application of the scales, suggesting the importance of training and supervising the team to ensure a standard of inter-observer reliability in the assessment of children with neurodevelopmental disorders in population studies.

**Keywords:** reliability measures, neurodevelopmental disorders, Bayley scale III, development measures, test reliability, neuropsychological assessment, child development

### FIABILIDAD DE LA ESCALA BAYLEY: EVALUACIÓN DE LOS NIÑOS AFECTADOS POR EL VIRUS DEL ZIKA

#### Resumen

Comprender y medir los factores determinantes que influyen en el desarrollo infantil podría contribuir a orientar mejor la atención sanitaria en la primera infancia. Se utilizaron las escalas Bayley III para evaluar el desarrollo de niños con y sin diagnóstico de síndrome congénito por el virus del zika que participaban en un estudio longitudinal y eran evaluados en el hogar. Se examinó la fiabilidad interobservador en tres puntos de evaluación y se describió el proceso de formación y supervisión del equipo interdisciplinar, con el fin de estandarizar los procedimientos de aplicación del instrumento en la recogida de datos. Las medidas de fiabilidad se obtuvieron mediante el índice kappa y el coeficiente de correlación intraclass (CCI). Los valores medios de kappa fueron de 0,92, 0,89 y 0,96, respectivamente, para la primera, segunda y tercera medidas de fiabilidad entre evaluadores. El CCI se mantuvo por encima del 90% para cada subescala evaluada en las tres mediciones realizadas. Los resultados muestran excelentes indicadores de fiabilidad en la aplicación de las escalas, lo que sugiere la importancia de la formación y supervisión del equipo para garantizar un estándar de fiabilidad interobservador en la evaluación de niños con trastornos del neurodesarrollo en estudios poblacionales.

**Palabras-clave:** medidas de fiabilidad, trastornos del neurodesarrollo, Escala Bayley III, precisión de test, evaluación neuropsicológica, desarrollo infantil

Sabe-se que diversos determinantes podem influenciar o desenvolvimento infantil. Entendê-los e mensurá-los de maneira sistemática e quantitativa poderão contribuir para um melhor direcionamento das ações de saúde. Nesse sentido, a Academia Americana de Pediatria (Sandler et al., 2001) recomenda a avaliação e acompanhamento de bebês e crianças pequenas de risco ao longo da primeira infância, permitindo intervenções precoces e específicas em possíveis alterações do neurodesenvolvimento. Estudos populacionais que se propõem a avaliar crianças com transtornos do desenvolvimento infantil desafiam padrões de pontuação da aquisição de habilidades dos resultados obtidos, em decorrência da natureza dos instrumentos de abordagem quantitativa, que frequentemente medem aquisições de habilidades por faixas etárias e utilizam amostras normativas de populações com desenvolvimento típico (Rodrigues, 2012).

O atraso do desenvolvimento pode ter diferentes causas. Destaca-se que a síndrome congênita do zika vírus (SCZV) engloba um conjunto de anomalias congênicas que podem incluir alterações visuais, auditivas e neuropsicomotoras que ocorrem em sujeitos expostos a esta infecção durante a gestação (Ministério da Saúde, 2016). Tais alterações podem variar quanto à severidade, provocando desafios aos profissionais de saúde que se voltam para a vigilância e a avaliação do desenvolvimento infantil.

O primeiro passo para garantir a atenção e o acesso a programas específicos se dá pelo diagnóstico, sendo fundamental o rastreamento, a identificação e o encaminhamento precoce adequado, já que fortes evidências apontam para a importância da intervenção precoce (Zwaigenbaum & Penner, 2018), destacando o uso de medidas confiáveis com valores elevados de sensibilidade, especificidade e confiabilidade (Campos et al., 2006; Santos & Ravanini, 2006; Blair & Hall, 2006).

Para crianças com alterações complexas em seu desenvolvimento, procedimentos que envolvam desde a vigilância até a triagem para fatores de risco, assim como a avaliação para determinar o diagnóstico funcional, são pilares para garantir as premissas recomendadas pelo Ministério da Saúde e intervir precocemente sempre que necessário. A partir do processo avaliativo, instaura-se o raciocínio clínico para decidir o melhor plano de intervenção (Gourladin & Sá, 2022).

Contudo, a disponibilidade de escalas padronizadas e validadas em nosso idioma, com garantia dos aspectos culturais, coloca-se como desafio para avaliação de crianças com transtornos do desenvolvimento infantil (Visser et al., 2014; Madaschi et al., 2016). A avaliação de indivíduos com deficiências que impactam todos os domínios do desenvolvimento infantil requer acomodações do teste para permitir participação plena do sujeito no processo (Bayley, 2006). Por acomodações, entendem-se as alterações nos procedimentos-padrão de administração dos testes para superar deficiências funcionais do participante, visando aumentar a validade das inferências a partir dos escores obtidos. Considera-se relevante o comprometimento funcional que o sujeito pode experimentar ao tentar demonstrar proficiência em uma avaliação (Kettler, 2012).

Originalmente publicadas em 1969, as Escalas Bayley, versão original americana, são consideradas padrão-ouro por atender a todas as propriedades psicométricas (Diamond, 2000). Após quase 50 anos de pesquisa, atualmente está disponível a versão 4 do instrumento, mantendo seus excelentes níveis de qualidade, respondendo a rigorosas propriedades psicométricas.

As escalas Bayley versão III (BSID III) utilizada nesta pesquisa é a mais recente disponível no Brasil (Madaschi & Paula, 2011), é também um dos instrumentos recomendados pelas diretrizes de estimulação precoce decorrente de microcefalia (Ministério da Saúde, 2016), para identificar atrasos desenvolvimentais, planejar intervenção e documentar progresso e evolução (Bayley, 2006). Embora seja padronizada, possui flexibilidade de aplicação ao considerar a dinamicidade inerente às diversas situações de avaliação de bebês (Bayley, 2006), tornando-se apropriada para avaliar habilidades cognitivas, linguísticas e motoras de crianças acometidas por anomalias congênitas.

A versão III da escala foi adotada como instrumento de avaliação no Projeto de Pesquisa “Efeitos das manifestações neurológicas congênitas associadas ao zika vírus sobre o desenvolvimento infantil: um estudo de coorte prospectiva no contexto da Atenção Básica, em Salvador-BA” para avaliar consequências desenvolvimentais em crianças nascidas durante a epidemia, tendo o domicílio como lócus primário para avaliação (Santos et al., 2022).

Tendo em vista os prováveis efeitos do entrevistador sobre a confiabilidade das respostas obtidas, o desenho de pesquisa incluiu procedimentos para medir a concordância entre os avaliadores ao longo do estudo. A confiança nos resultados é parcialmente uma função da quantidade de desacordo ou erro introduzido no estudo, pela inconsistência entre os aplicadores do instrumento. A confiabilidade ocorre de forma dinâmica, dependendo da função do instrumento, população na qual é administrado, circunstâncias e contexto. Tais aspectos colocam a importância do treinamento e supervisão contínuos para alcançar adequada confiabilidade interobservadores (Souza et al., 2017). A confiabilidade de equivalência permite identificar até que ponto os avaliadores estariam aptos para observar e mensurar o fenômeno ou variável, de maneira adequada e prevista pelo instrumento conforme sua validade.

A mensuração da concordância entre os coletadores refere-se à estabilidade, consistência interna e equivalência da medida, embora a confiabilidade não seja propriedade fixa do instrumento (Souza et al., 2017).

Portanto, cabe submeter uma equipe de avaliadores ao processo de treinamento e teste-reteste para mensurar o nível de equivalência entre eles, visando minimizar os erros de medição. A equivalência refere-se ao grau de concordância entre dois ou mais observadores quanto aos escores de um instrumento (Heale & Twycross, 2015). A consistência interna entre aplicadores se expressa pelo coeficiente kappa (K) que mede o grau de concordância entre proporções derivadas de amostras dependentes (Cohen, 1968). Acrescente-se o coeficiente de correlação intraclassa (ICC), medida que avalia a confiabilidade ou consistência entre múltiplas medições feitas por diferentes aplicadores.

Dada a importância da avaliação desenvolvimental para o processo de intervenção terapêutica com crianças e familiares, este artigo tem como objetivo examinar a confiabilidade obtida em três pontos de avaliação de um estudo longitudinal com as Escalas Bayley III do desenvolvimento infantil e descrever o processo de treinamento e supervisão permanente da equipe interdisciplinar, visando à padronização dos procedimentos de aplicação na coleta de dados.

## Método

Estudo longitudinal quantitativo, observacional e descritivo para avaliar concordância entre aplicadores das Escalas Bayley III do desenvolvimento infantil, na cidade de Salvador, na Bahia, a quarta capital mais populosa do Brasil.

### Participantes

O modelo de desenho de confiabilidade escolhido para este estudo foi constituído pelos blocos incompletos equilibrados, descrito por Fleiss (1981), no qual um dos examinadores conduz a entrevista, enquanto o outro assiste ao exame como espectador neutro. Blocos incompletos equilibrados, segundo o autor, referem-se a um tipo específico de *design* experimental usado em estudos que envolvem a avaliação de múltiplos tratamentos. Nesse delineamento, cada participante, ou unidade experimental, não é exposto a todas as combinações possíveis de tratamentos, mas sim a um subconjunto delas. A abordagem é particularmente útil diante de elevado número total de combinações, tornando inviável o teste de todas elas.

O método de blocos incompletos equilibrados permite reduzir o tamanho do experimento e economizar recursos, mantendo, ao mesmo tempo, o equilíbrio entre as condições testadas. Para garantir a validade dos resultados, o planejamento da alocação dos tratamentos aos blocos de participantes deve ser aleatório ou sistemático, dependendo da estratégia adotada. Assim, controla-se a variabilidade externa e melhora-se a precisão das conclusões, tornando o estudo mais robusto e confiável.

As duplas constituídas por dois aplicadores após o treinamento teórico registram escores a partir da mesma entrevista, mas fazem uma avaliação independente. Utilizou-se o método de arranjo simples, por meio de análise combinatória para formação dessas duplas. Em outro momento, os papéis da dupla eram invertidos na avaliação de outra criança. As medidas da confiabilidade dos três pontos de avaliação do estudo longitudinal foram obtidas de maneira independente. Mensurou-se o nível de concordância entre avaliadores, para cada ponto de avaliação do estudo longitudinal, realizando-se três medidas independentes do desempenho cognitivo, motor e linguístico de crianças com e sem exposição à SCZV.

### Composição da equipe

Constituiu-se uma equipe com estudantes de Bacharelado Interdisciplinar (BI) em Saúde, de Psicologia, Saúde Coletiva, Fisioterapia e com profissionais de Psicologia. A primeira medida ocorreu na linha de base entre abril de 2017 e março de 2018, em que seis entrevistadores realizaram 29 aplicações do instrumento seguindo o desenho acima descrito para avaliar confiabilidade. Uma segunda medida, realizou-se entre maio de 2018 e março de 2019, quando oito novos entrevistadores se juntaram a um membro antigo e esses nove membros realizaram 67 aplicações para o estudo de confiabilidade. A última medida ocorreu entre fevereiro e agosto de 2019, na qual quatro aplicadores realizaram 11 avaliações para a amostra de confiabilidade.

## **Instrumento**

As Escalas Bayley III visam medir o desempenho de bebês e crianças pequenas, com o propósito de identificar competências e pontos críticos, contribuindo para um planejamento adequado da intervenção terapêutica. Cinco deles são investigados a partir de avaliações diretas com a criança, abordando domínios cognitivo, linguagem expressiva e receptiva, motricidade global e fina, além das escalas socioemocional e de comportamento adaptativo, aplicadas em entrevistas com os pais (Bayley, 2006).

Conforme o Manual Técnico do instrumento, o tempo de aplicação pode variar entre 50 minutos para crianças e meses, e 90 minutos para maiores de 13 meses.

A estrutura do teste fornece cinco tipos de pontuações: (i) pontuações totais brutas; (ii) pontuações escalonadas; (iii) pontuações compostas; (iv) classificações com base em percentis; e (v) pontuação de desenvolvimento. Para cada domínio, define-se pontuação bruta como o número total de itens pelos quais a criança recebe crédito, somado com o número de itens anteriores ao ponto de início da criança. A pontuação escalonada deriva da pontuação bruta e varia de 1-19 com média de 10 e desvio-padrão de 3, enquanto a pontuação composta é calculada com base na pontuação escalonada e varia de 40-160 com média de 100 e desvio-padrão de 15.

Dentre os efeitos decorrentes da SCZV, destaca-se o desenvolvimento inadequado da motricidade ampla, o qual compromete a capacidade da criança em rolar, sentar-se e, em muitos casos, em manter o controle cervical. Quanto ao desenvolvimento da motricidade fina, relatam-se dificuldades na realização de atividades manuais; quanto ao sistema sensorial, observa-se comprometimento da capacidade visual e auditiva, implicando severa dificuldade de compreender e produzir linguagem (Weeler, 2018). Todavia a BSID III permite ajustes na sua versão-padrão, sendo possível propor facilitações nas provas da avaliação desde que não ocorram modificações no conteúdo e objetivo das referidas provas (Visser et al., 2013; Visser et al., 2014).

## **Seleção da equipe de coleta**

Ao início do projeto lançou-se edital para oferta de curso introdutório sobre as Escalas Bayley III do desenvolvimento infantil, com subsequente seleção de aplicadores conforme desempenho demonstrado e nível de envolvimento do participante com o projeto. Considerando os pontos de avaliação do estudo longitudinal, idade limite de 42 meses para participação da criança e oscilação de entrevistadores ao longo da coleta, novos editais foram abertos, transformando a oferta do treinamento em curso de extensão certificado pela Universidade Federal da Bahia (UFBA).

## **Capacitação dos aplicadores**

O conteúdo do treinamento em suas três etapas abordou tópicos relativos ao desenvolvimento infantil típico e atípico, entendimento e manuseio das Escalas Bayley III.

O processo de capacitação e supervisão foi conduzido por duas docentes de psicologia com experiência em desenvolvimento infantil, aspectos epidemiológicos do desenvolvimento e

instrumentos de avaliação quantitativa, incluindo uma terceira psicóloga responsável pela coordenação da equipe na condução de todas as atividades envolvidas na visita domiciliar ou das unidades de saúde, para coleta de dados durante todo o período estudado.

A primeira atividade prática da capacitação consistiu na avaliação de crianças conhecidas da equipe realizadas no próprio domicílio, seguindo-se com a experiência piloto, contemplando crianças com desenvolvimento típico e atípico. Cumpridas as exigências teórico-práticas, os aplicadores iniciavam as avaliações com os participantes do estudo no cenário da investigação. Inicialmente, todos os integrantes da equipe tiveram um parceiro mais experiente na dupla constituída, na qual um dos membros abordava a criança e registrava a pontuação, enquanto o segundo apenas observava a abordagem, registrando a pontuação de forma independente e silenciosa. Posteriormente, tais aplicações eram discutidas em supervisão semanal, para avaliação de desempenho dos membros em treinamento. Com o avanço das aplicações durante a segunda e terceiras medidas realizadas na coorte, reuniões regulares para supervisões cederam lugar a encontros agendados mediante demandas dos aplicadores.

### **Coleta de dados**

O percurso do trabalho iniciava-se com a saída da equipe do Instituto de Saúde Coletiva, chegada ao local de avaliação – o domicílio ou excepcionalmente a unidade de saúde por risco de segurança para equipe –, tempo de aplicação do instrumento e retorno ao instituto. O tempo de cada avaliação com o instrumento foi de aproximadamente duas horas, oscilando conforme o perfil de desenvolvimento da criança, variações do estado de saúde e bem-estar e necessidade de pausas eventuais, conforme características individuais. A equipe observou que a avaliação domiciliar tornou o processo mais confortável para a criança, permitindo familiaridade do avaliador com a limitação existente, melhoria na comunicação e flexibilidade necessária para adotar as necessárias acomodações do teste.

Desde as primeiras avaliações das crianças com diagnóstico confirmado da SCZV, observaram-se dificuldades na aplicação das Escalas Bayley III, com eventual desvantagem para a criança pela própria administração do teste. Reconheceu-se que as acomodações seriam possibilidades necessárias de alterações nos procedimentos-padrão, para superar deficiências funcionais dos indivíduos e aumentar a validade das respostas (Kettler, 2012). Constatou-se também que o próprio manual (Bayley, 2006) estabelece critérios e rigor necessários para adaptações. Apesar de escassos, estudos que aplicaram a BSID III com facilitações, para avaliar crianças com múltiplos comprometimentos, verificaram que o uso das facilitações corrigiu diferenças na pontuação bruta das provas, principalmente da escala cognitiva, aumentando a validade e uso do instrumento perante essas condições (Ruiter et al., 2010). Com base nessas evidências, elaborou-se um material para determinar o desempenho das habilidades cognitivas do sujeito, com a melhor expressão e precisão possível do seu desenvolvimento. Padronizou-se o uso da escala mediante treinamentos e supervisões ao longo do estudo.

### Análise dos dados

Realizou-se análise de equivalência para identificar o grau de concordância entre duplas de observadores, em relação aos escores do instrumento Escalas Bayley (Heale & Twycross, 2015). Para calcular a concordância, quando os resultados foram variáveis contínuas, utilizou-se o coeficiente de correlação intraclass. Esse coeficiente é amplamente empregado para avaliar concordância entre medidas repetidas, especialmente quando há múltiplos observadores. Consideraram-se resultados excelentes quando a concordância entre valores é superior a 0,75.

Para analisar a concordância entre variáveis categóricas, optou-se pelo índice kappa (K) para medir o grau de concordância entre proporções derivadas de amostras dependentes (Cohen, 1968). Os valores obtidos são assim classificados:  $\leq 0$  nenhuma concordância, 0.01-0.20 ligeira concordância, 0.21-0.40 concordância considerável, 0.41-0.60 concordância moderada, 0.61-0.80 concordância substancial e 0.81-1.00 concordância excelente (McHugh, 2012; Souza et al., 2017). Utilizaram-se os programas Statistical Packages for the Social Sciences (SPSS versão 20) e \*R (R Statistical Language versão 3.6.1).

### Resultados

A primeira medida de confiabilidade realizada na linha de base envolveu seis aplicadores e 29 crianças avaliadas por duplas de aplicadores, obtendo-se os seguintes resultados: os valores de kappa variaram entre 0.47 e 1 com média igual a 0.92, e pelo menos 70% das questões apresentaram valor próximo de 1, obtendo concordância excelente na classificação entre aplicadores. A média dos 58 escores das 29 avaliações nesse ponto foi 84,2 (DP = 24,8) para escala cognitiva, 81,3 (DP = 25,4) para linguagem e 80.7 (DP = 28.8) para motricidade.

A segunda medida contou com 67 crianças avaliadas e nove aplicadores para exame de equivalência. Os valores de kappa variaram de 0.43 a 1 com média igual a 0.89, com aproximadamente 70% das questões apresentando valor de kappa próximo de 1, mantendo a classificação excelente entre aplicadores. Para as 67 avaliações realizadas nesse ponto, a média dos 134 escores obtidos foi 73,3 (DP = 22,7) para escala cognitiva, 72,2 (DP = 24,7) para escala de linguagem e 70,0 (DP = 27,5) para escala motora.

Na terceira medida, 11 crianças compuseram a amostra de equivalência com a participação de quatro aplicadores. Os valores de kappa variaram de 0.21 a 1 com média igual a 0.96, mantendo-se 70% das questões com valores de kappa próximo de 1, e classificação excelente entre aplicadores. Nesse último ponto de medida, realizaram-se 11 avaliações (22 escores) com médias iguais a 63,9 (DP = 17), 57,9 (DP = 19) e 58,1 (DP = 26), para a escala cognitiva, de linguagem e motora, respectivamente. A Tabela 1 apresenta a distribuição da classificação do coeficiente kappa para os três pontos de avaliação do estudo.

A análise das medidas de confiabilidade na aplicação das Escalas Bayley III do desenvolvimento infantil realizada neste estudo longitudinal demonstrou valores satisfatórios do coeficiente kappa, indicando concordância excelente entre aplicadores, com discreta elevação para valores médios na primeira e terceira medidas (Tabela 1).



Cabe referir ainda que os itens com baixa concordância se distribuíram da seguinte forma neste estudo longitudinal de confiabilidade. Na linha de base, registrou-se apenas um item do domínio cognitivo (Item 8). A segunda medida de avaliação entre aplicadores registrou seis itens, sendo dois de linguagem expressiva (1;43), dois de linguagem receptiva (2;6), um item cognitivo (4) e outro de motricidade fina (1). Na terceira medida, registaram-se apenas dois itens relativos à linguagem expressiva (2;6).

### Tabela 1

*Concordância obtida pelo coeficiente kappa para três medidas de avaliação de confiabilidade entre aplicadores do estudo longitudinal de abril de 2017 a agosto de 2019, Salvador (BA).*

Concordância	Nº de questões	%	Valor de kappa
<b>Primeira medida:</b> linha de base			
Moderada	1	0,31	0,41 a 0,60
Substancial	36	11,04	0,61 a 0,80
Excelente	289	88,65	0,81 a 1,00
<b>Segunda medida</b>			
Moderada	10	3,37	0,41 a 0,60
Substancial	47	15,82	0,61 a 0,80
Excelente	238	80,13	0,81 a 1,00
<b>Terceira medida</b>			
Considerável	2	0,70	0,21 a 0,40
Substancial	24	8,42	0,61 a 0,80
Excelente	259	90,88	0,81 a 1,00

Fonte: Elaboração dos autores.

Nota. Categorias de concordância sem frequência foram omitidas.

O coeficiente de correlação intraclasses para cada um dos três escores de desempenho em cada ponto da avaliação se manteve acima de 90% revelando elevado nível de concordância entre aplicadores também classificado como excelente (Tabela 2).

**Tabela 2**

*Coefficientes de correlação intraclasse (CCI) para as medidas de confiabilidade conforme as subescalas cognitiva, linguagem e motricidade realizadas em três pontos de avaliação do estudo longitudinal de abril de 2017 a agosto de 2019, Salvador (BA)*

Avaliações conforme escalas	CCI	CCI 95%
<b>Primeira medida: linha de base</b>		
Cognitiva	0,925	0,865
Linguagem	0,951	0,907
Motricidade	0,939	0,889
<b>Segunda medida</b>		
Cognitiva	0,963	0,942
Linguagem	0,994	0,991
Motricidade	0,980	0,968
<b>Terceira medida</b>		
Cognitiva	0,998	0,995
Linguagem	0,997	0,992
Motricidade	1,000	0,999

Fonte: Elaboração dos autores.

### Discussão

Em se tratando de um estudo longitudinal com renovação da equipe de aplicadores, foi possível manter a qualidade dos dados, reproduzindo resultado consistente no tempo e espaço, conforme as medidas de confiabilidade obtidas. O formato de treinamento inicial se completou com a sequência de supervisões regulares, qualificando a equipe ao identificar dúvidas e discordâncias, estabelecendo, portanto, um padrão de elevada confiabilidade interobservadores nas três medidas de avaliação realizadas (Souza et al., 2017).

Analisando-se os resultados gerais dos três pontos avaliados, observaram-se algumas variações dos níveis de confiabilidade, indicando a importância de que para cada novo ponto de avaliação do estudo se realize a confiabilidade entre aplicadores, ainda que o instrumento em questão tenha demonstrado alta confiabilidade anteriormente.

O número crescente de estudos nacionais recentes que utilizaram as escalas Bayley III indica a importância e utilidade desse instrumento no diagnóstico de atrasos motores, cognitivos e de linguagem em crianças brasileiras (Ferreira et al., 2014; Hentges et al., 2014). Consideram-se cientificamente relevantes investigações que examinam os achados para populações com alterações complexas do desenvolvimento infantil, enfatizando a necessidade de testar as acomodações necessárias, para manter evidências de validade e confiabilidade de determinados instrumentos de avaliação. Enfrentou-se neste estudo desafio importante, para avaliar crianças

com alto risco de desenvolvimento decorrente de múltiplas deficiências e frequente comprometimento do desenvolvimento global.

As acomodações são requeridas em tais situações, permitindo alterações nos procedimentos-padrão de administração dos testes, para superar deficiências funcionais do participante, aumentando a validade das inferências a partir dos escores obtidos. Considera-se relevante o comprometimento funcional que o sujeito pode experimentar ao tentar demonstrar proficiência em uma avaliação (Kettler, 2012). Empreendeu-se um esforço para viabilizar avaliação psicológica em uma população com tantas limitações, as quais poderiam colocar os sujeitos em condição de não testáveis perante outros instrumentos, contribuindo, portanto, para avanços na avaliação do desenvolvimento infantil no Brasil, um país sem nenhuma ferramenta similar validada (Madaschi et al., 2016).

A BSID-III foi administrada por meio de procedimentos-padrão, com adaptações para deficiências visuais ou motoras da criança, conforme sugerido pelo manual (Bayley, 2006). Dentre as acomodações utilizadas em nosso estudo, referimos alguns exemplos como uso de luz e brilho, aumento do tamanho dos objetos manipulativos e ampliação do tempo para resposta da criança.

Outras acomodações do teste na aplicação da BSID-III têm sido utilizadas. Por exemplo, para crianças apenas com visão clara, as luzes do teto foram desligadas e uma lanterna foi usada para fornecer contraste (Wheeler et al., 2020). Embora a abrangência dos itens alcance o espectro de desenvolvimento desde o nascimento até a primeira infância, a dependência da produção visual e motora pode penalizar as crianças com SCZ em demonstrar sua capacidade de realização. Por outro lado, os escores brutos do BSID-III e os escores equivalentes à idade fornecem uma medida sensível da mudança potencial ao longo do tempo, permitindo monitoramento de detecção de mudanças no desenvolvimento (ganho ou perda de habilidades), acompanhamento dos impactos para identificar forças e fraquezas da criança como direcionadores em resposta ao tempo, tratamento, da intervenção, podendo também ser um auxiliar determinante dos impactos dos agravos das intercorrências clínicas que afetam diretamente o desenvolvimento, como ou convulsões ou outros sintomas clínicos (Wheeler et al., 2020).

Embora os valores de kappa em pelo menos 70% das questões analisadas tivessem sido próximos de 1, houve questões distribuídas ao longo das três medidas com baixa concordância entre aplicadores. Na linha de base, isso ocorreu com questões do domínio cognitivo e nos demais pontos com os domínios da linguagem e motricidade. Uma hipótese possível que poderia explicar a baixa concordância seria a probabilidade de prejuízo do domínio cognitivo pela infecção viral e da alta correlação entre os três domínios. Sabendo-se que o prejuízo motor decorrente de hipertonia axial e apendicular pode comprometer a magnitude da resposta ou do comportamento da criança e, por conseguinte, dificultar o julgamento do avaliador na aplicação dessas provas que se afastaram da alta concordância. Nesse sentido, recomenda-se mais atenção ao treinamento dos itens que requerem maior sensibilidade do avaliador para reconhecer o desempenho da criança.

Ainda sobre itens da linguagem expressiva com baixa concordância, as provas que envolvem sons guturais indiferenciados requerem familiaridade do avaliador para interpretá-los e pontuar. E, por fim, a baixa concordância neste estudo atingiu itens constituídos por elementos da dimensão subjetiva como esboço do sorriso social ao falar com a criança. Apesar das várias fontes potenciais de discordância na aplicação de um teste psicométrico, os escores obtidos neste estudo foram bastante concordantes, refletindo tanto uniformidade na abordagem das crianças quanto na interpretação das respostas dadas por elas. Considera-se ter sido possível lidar com fatores relativos ao instrumento, população e contexto, atingindo um nível de confiabilidade elevado (Souza et al., 2017). Ao utilizar a confiabilidade do tipo de equivalência, identificou-se o grau de aptidão dos aplicadores para observar e mensurar o fenômeno de maneira adequada, conforme preconizado pelo Manual das Escalas Bayley III de Desenvolvimento Infantil.

Avaliar a criança no seu domicílio ou eventualmente na unidade de saúde de sua referência facilitou a abordagem. O acesso do avaliador a esse espaço favoreceu alguma familiaridade com a limitação da criança, assim como a melhoria na comunicação e flexibilidade necessária para adotar acomodações. A validade ecológica enfatiza uma nova compreensão da relação entre resultados da avaliação de um teste com o desempenho das tarefas cotidianas. Considera ainda o desenvolvimento de testes compostos por tarefas cognitivas cotidianas, de modo que as inferências possam ser facilmente extraídas dos resultados e da provável capacidade do indivíduo de realizar aquelas tarefas na vida diária (Spoonner & Pachana, 2006). Para Pasquali (2017), a validade ecológica se refere à forma como as evidências devem ser buscadas, aproximando métodos, materiais e situações da avaliação do mundo real que está sendo examinado.

Para a criança com desenvolvimento atípico altamente provável pelo diagnóstico da SCZV, a avaliação na residência ampliou a possibilidade de expressão do seu desenvolvimento. Os obstáculos encontrados desde as primeiras aplicações decorrentes de atrasos consideráveis nos diferentes domínios desenvolvimentais nos remeteu ao Manual da Bayley, para entender as adaptações e definir condutas e procedimentos para o trabalho de campo com essa população. Realizaram-se oficinas, estruturaram-se rotinas de monitoramento e supervisão permanente para garantia dos procedimentos de adaptação do instrumento sem afetar a integridade do teste.

Enfim, a adaptação do BSID-III para acomodar dificuldades visuais e motoras tornou possível avaliar com mais precisão a função de desenvolvimento de crianças com SCZ. Tratou-se de 16 adaptações, organizadas conforme o tipo de facilitação empregada (visual, motora ou geral), utilizadas em todas as áreas de avaliação deste instrumento. Os dados observados sugeriram a importância da construção de novas perspectivas na avaliação de crianças com desenvolvimento atípico, minimizando a interferência de déficits (Araújo et al., 2017).

### **Considerações Finais**

Demonstrou-se ser possível realizar avaliação domiciliar do desenvolvimento de crianças com múltiplas deficiências em contexto populacional, utilizando adaptações previstas pelas Escalas Bayley III, com níveis satisfatórios de confiabilidade. O treinamento, supervisão e

monitoramento da equipe calibraram os avaliadores conforme as medidas de confiabilidade demonstradas pelo kappa e ICC. Evidências sobre bom desempenho de instrumentos neuropsicológicos em estudos populacionais investigando grupos ainda pouco explorados em saúde coletiva favorecem o andamento da pesquisa pela credibilidade dos resultados obtidos e auxiliam futuros pesquisadores na escolha da ferramenta.

Conclui-se que o estudo contribui para avançar o conhecimento sobre crianças com deficiências múltiplas na perspectiva da saúde coletiva, conferindo confiabilidade ao processo de avaliação psicológica em população com limitações múltiplas do desenvolvimento infantil no contexto comunitário. Considera-se cientificamente relevantes investigações que examinem os achados para populações com alterações complexas do desenvolvimento infantil, enfatizando a necessidade de testar as acomodações necessárias, para manter evidências de validade e confiabilidade de determinados instrumentos de avaliação.

### Referências

- Araújo, C. F., Cabral, C. B., Dantas, J., Oliveira, K. N. R. de F., Flores, M. C. M., Almeida, T. M., Silva, T. C. L. & Santos, D. N. (2017). *Manual do protocolo de facilitações sensoriais e motoras para uso da escala Bayley de desenvolvimento infantil – BSIDIII em crianças com síndrome congênita do zika vírus*. <https://repositoriohml.ufba.br/handle/ri/31933>
- Bayley, N. (2006). *Manual of Bayley Scales of Infant Development TM*. 3rd ed. The Psychological Corporation.
- Blair, M.; & Hall, D. (2006). *From health surveillance to health promotion: The changing focus in preventive children's services* (pp. 730–735). <http://www.archdischild.com>
- Campos, D., Santos, D. C. C., Gonçalves, V. M. G., Goto, M. M. F., Arias, A. V., Brianeze, A. C. G. S., Campos, T. M., & Mello, B. B. A. (2006). Agreement between scales for screening and diagnosis of motor development at 6 months. *Jornal de Pediatria*, 82(6), 470–474. <https://doi.org/10.2223/JPED.1567>
- Cohen, J. (1968). Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220. doi: 10.1037/h0026256
- Diamond, A. (2000). Close interrelation of motor development and cognitive development and of the cerebellum and prefrontal cortex. *Child Development*, 71, 44–56.
- Ferreira, R. C., Mello, R. R., & Silva, K. S. (2014). Neonatal sepsis as a risk factor for neurodevelopmental changes in preterm infants with very low birth weight. *Jornal de Pediatria*, 90, 293–299. doi: <https://doi.org/10.1016/j.jpmed.2013.09.006>
- Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement*, 5(1), 105–112. doi: 10.1177/014662168100500115.
- Gourladins, J. B., & Sá, C. dos S. C. (2022). *Desenvolvimento e saúde mental na infância*. Editora Ampla.
- Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence Based Nursing*, 18(3), 66–67. doi.org/10.1136/eb-2015-102129.
- Hentges, C. R., Silveira, R. C., Procianny, R. S., Carvalho, C. G., Filipouski, G. R., Fuentefria, R. N., Marquezotti, F., Terrazan, A. C. (2014). Association of late-onset neonatal sepsis with late neurodevelopment in the first two years of life of preterm infants with very low birth weight. *Jornal de Pediatria*, 90, 50–57. <https://doi.org/10.1016/j.jpmed.2013.10.002>
- Kettler, R. J. (2012). Testing accommodations: Theory and research to inform practice. *International Journal of Disability, Development and Education*, 59(1), 53–66. <https://doi.org/10.1080/1034912X.2012.654952>
- Madaschi, V., & Paula, C. S. (2011). Medidas de avaliação do desenvolvimento infantil: uma revisão de literatura nos últimos cinco anos. *Cadernos de Pós-Graduação em Distúrbios do Desenvolvimento*, São Paulo, 11(1), 52–56.
- Madaschi, V., Mecca, T. P., Macedo, E. C., & Paula, C. S. (2016). Baykey III Scales of Infant and Toddler Development: Transcultural Adaptation and Psychometric Properties. *Paidéia*, 26(64), 189–197. <https://doi.org/10.1590/1982-43272664201606>
- McHugh M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. doi: 10.11613/BM.2012.031
- Ministério da Saúde. Secretaria de Atenção à Saúde. (2016). *Diretrizes de estimulação precoce: Crianças de zero a 3 anos com atraso no desenvolvimento neuropsicomotor decorrente de microcefalia [Early stimulation guidelines: children aged 0 to 3 years with delayed neuropsychomotor development due to microcephaly]*. [https://www.sbp.com.br/fileadmin/user\\_upload/2016/01/Diretrizes-de-Estimulacao-Precoce\\_Microcefalia.pdf](https://www.sbp.com.br/fileadmin/user_upload/2016/01/Diretrizes-de-Estimulacao-Precoce_Microcefalia.pdf).
- Pasquali, L. (2017). Validade dos testes. *Revista Examen*, 1(1), 14–48.
- Rodrigues, O. M. P. R. (2012). Escalas de desenvolvimento infantil e o uso com bebês. *Educar em Revista*, (43), 81–100. <https://doi.org/10.1590/S0104-40602012000100007>
- Ruiter, S. A., Nakken, H., van der Meulen, B. F., & Lunenborg, C. B. (2010). Low motor assessment: A comparative pilot study with young children with and without motor impairment. *Journal of developmental and physical disabilities*, 22(1), 33–46. <https://doi.org/10.1007/s10882-009-9165-5>

- Sandler, A. D., Brazdziunas, D., Cooley, C. W., De Pijem, L. G., Hirsch, D., Kastner, T. A., Kummer, M. E., Quint, R. D., Ruppert, E. S., Anderson, W. C., Crider, B., Burgan, P., Garner, C., McPherson, M., Michaud, L., Yeargin-Allsp, M., Cartwright, D., Johnson, C. P., & Smith, K. (2001). Developmental surveillance and screening of infants and young children. *Pediatrics*, *108*, 192–195. <https://doi.org/10.1542/peds.108.1.192>
- Santos, D. N., Araujo, T. M., Santos, L. M., Kuper, H., Aquino, R., Silveira, I. H., Miranda, S. S., Pereira, M., & Werneck, G. L. (2022). The Salvador Primary Care Longitudinal Study of child development (CohortDI-Ca) following the zika epidemic study protocol. *International Journal of Environmental Research and Public Health*, *19*(5), 2514. <https://doi.org/10.3390/ijerph19052514>
- Santos, D. C. C., & Ravanini, S. G. (2006). Aspectos do diagnóstico do desenvolvimento motor. In Ribeiro, M. V.; Gonçalves, V. M. Neurologia do desenvolvimento da criança (pp. 258–269). Revinter.
- Souza, A. C., Alexandre, N. M. C., & Guirardello, E. B. (2017). Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade [Psychometric properties in instruments evaluation of reliability and validity]. *Epidemiologia e Serviços de Saúde* Brasília, *26*(3), 649–659. doi: 10.5123/S1679-49742017000300022
- Spooner, D. M., & Pachana, N. A. (2006). Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. *Archives of Clinical Neuropsychology*, *21*(4), 327.
- Visser, L., Ruiter, S. A., van der Meulen, B. F., Ruijsenaars, W. A., & Timmerman, M. E. (2013). Validity and suitability of the Bayley-III low motor/vision version: A comparative study among young children with and without motor and/or visual impairments. *Research in developmental disabilities*, *34*(11), 3736–3745. <https://doi.org/10.1016/j.ridd.2013.07.027>
- Visser, L., Ruiter, S. A., van der Meulen, B. F., Ruijsenaars, W. A., & Timmerman, M. E. (2014). Accommodating the Bayley-III for motor and/or visual impairment: A comparative pilot study. *Pediatric Physical Therapy: the Official Publication of the Section on Pediatrics of the American Physical Therapy Association*, *26*(1), 57–67. <https://doi.org/10.1097/PEP.000000000000004>
- Zwaigenbaum, L., & Penner, M. (2018). Autism spectrum disorder: advances in diagnosis and evaluation. *BMJ*, *361*:k1674. Doi: 10.1136/bmj.k1674.
- Wheeler A. C. (2018). Development of infants with congenital zika syndrome: What do we know and what can we expect? *Pediatrics*, *141*(Suppl 2), S154–S160. <https://doi.org/10.1542/peds.2017-2038D>.
- Wheeler, A. C., Toth, D., Ridenour, T., Lima Nóbrega, L., Borba Firmino, R., Marques da Silva, C., Carvalho, P., Marques, D. Okoniewski, K., Ventura, L. O., Bailey, D. B., Jr, & Ventura, C. V. (2020). Developmental outcomes among young children with congenital Zika syndrome in Brazil. *JAMA Network Open*, *3*(5), e204096. <https://doi.org/10.1001/jamanetworkopen.2020.409>

**EQUIPE EDITORIAL****Editores-chefe**

Cristiane Silvestre de Paula

**Editores associados**

Alessandra Gotuzo Seabra

Ana Alexandra Caldas Osório

Luiz Renato Rodrigues Carreiro

Maria Cristina Triguero Veloz Teixeira

**Editores de seção****"Avaliação Psicológica"**

Alexandre Luiz de Oliveira Serpa

André Luiz de Carvalho Braule Pinto

Juliana Borges Sbicigo

Natália Becker

Lisandra Borges

**"Psicologia e Educação"**

Alessandra Gotuzo Seabra

Carlo Schmidt

Regina Basso Zanon

**"Psicologia Social e Saúde das Populações"**

Enzo Banti Bissoli

Marina Xavier Carpena

Daniel Kveller

**"Psicologia Clínica"**

Ana Alexandra Caldas Osório

Carolina Andrea Ziebold Jorquera

Julia Garcia Durand

**"Desenvolvimento Humano"**

Maria Cristina Triguero Veloz Teixeira

Rosane Lowenthal

**Artigos de Revisão**

Jessica Mayumi Maruyama

**Suporte técnico**

Fernanda Antônia Bernardes

Giovana Gatto Nogueira

**PRODUÇÃO EDITORIAL****Coordenação editorial**

Surane Chilian Vellenich

**Editorial intern**

Isabelle Callegari Lopes

**Preparação de originais**

Mônica de Aguiar Rocha

**Revisão**

Alessandra Biral

**Diagramação**

Acqua Estúdio Gráfico