



Boletín Médico del Hospital Infantil de México

www.elsevier.es/bmhim



ARTÍCULO DE REVISIÓN

Análisis comparativo de pruebas de tamiz para la detección de problemas en el desarrollo diseñadas y validadas en México



Rodrigo Orcajo-Castelán^{a,d,e,*}, Beatriz Sidonio-Aguayo^b,
Jorge Augusto Alcacio-Mendoza^c y Giovana Lucía López-Díaz^c

^a Unidad Médica de Alta Especialidad Hospital Regional de Psiquiatría Unidad Morelos, Instituto Mexicano del Seguro Social, México D.F., México

^b Departamento de Rehabilitación Pediátrica, Hospital Infantil de México Federico Gómez, México D.F., México

^c Facultad Mexicana de Medicina, Universidad La Salle, México D.F., México

^d Principles and Practice of Clinical Research Course, Spaulding Rehabilitation Center, Massachusetts General Hospital, Harvard Medical School, Cambridge, Massachusetts, U.S.A.

^e Centre For Global Mental Health, King's College London/London School Of Hygiene and Tropical Medicine, Londres, Reino Unido

Recibido el 2 de noviembre de 2015; aceptado el 3 de noviembre de 2015

Disponible en Internet el 30 de diciembre de 2015

PALABRAS CLAVE

Neurodesarrollo;
Pruebas de tamiz;
Riesgo de sesgo;
Validación;
México

Resumen

Introducción: En los últimos años se han desarrollado varias pruebas de tamiz para el desarrollo infantil temprano (DIT) en menores de 5 años en México. El objetivo de esta revisión fue comparar la calidad del reporte de validación publicado y riesgo de sesgo entre diferentes pruebas desarrolladas y validadas en México.

Métodos: Se realizó una búsqueda en bases de datos, literatura gris y referencia cruzada documental. Se efectuó un análisis comparativo de la calidad del reporte (STARD) y el riesgo de sesgo (QUADAS y QUADAS-2).

Resultados: Se incluyeron las siguientes cinco pruebas: Valoración Neuroconductual del Desarrollo del Lactante (VANEDELA), Evaluación del Desarrollo Infantil (EDI), Prueba de Tamiz del Neurodesarrollo Infantil (PTNI), Cartillas de Vigilancia para identificar alteraciones en el Desarrollo del Lactante (CVDL) e Indicadores de riesgo del Perfil de Conductas de Desarrollo (INDIPCD-R). Ninguna cumplió el 100% de los ítems de acuerdo con STARD. Las más completas en su descripción fueron VANEDELA y EDI. El área de procedimientos de muestreo fue en la que hubo menor cumplimiento (VANEDELA, PTNI, CVDL, INDIPCD-R). En QUADAS, todas las pruebas presentaron algún riesgo de sesgo. Las más importantes fueron la selección de la muestra y la elección del estándar de oro, que en dos estudios se identificó que no era el más adecuado (PTNI, INDIPCD-R).

* Autor para correspondencia.

Correo electrónico: rodrigo.aquilino.orcajo.castelan@kcl.ac.uk (R. Orcajo-Castelán).

Conclusiones: Las pruebas de tamiz mexicanas para el DIT varían en la calidad de reporte publicado y riesgo de sesgo. La de mejor calidad de reporte de validación es VANEDELA y la de menor riesgo de sesgo en los datos publicados es la prueba EDI.

© 2015 Hospital Infantil de México Federico Gómez. Publicado por Masson Doyma México S.A. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

KEYWORDS

Child development;
Screening test;
Bias risk;
Validation;
Mexico

Comparative analysis of child development screening tools designed and validated in Mexico

Abstract

Background: In recent years a number of child development screening tools have been developed in Mexico; however, their properties have not been compared. The objective of this review was to compare the report quality and risk bias of the screening tools developed and validated in Mexico in their published versions.

Methods: A search was conducted in databases, gray literature and cross references. The resultant tests were compared and analyzed using STARD, QUADAS and QUADAS-2 criteria.

Results: "Valoración Neuroconductual del Desarrollo del Lactante" (VANEDELA), "Evaluación del Desarrollo Infantil or EDI" (CDE in English), "Prueba de Tamiz del Neurodesarrollo infantil" (PTNI), "Cartillas de Vigilancia para identificar alteraciones en el Desarrollo del Lactante" (CVDL) and "Indicadores de riesgo del Perfil de Conductas de Desarrollo" (INDIPCD-R) were included for the comparison. No test fulfilled all STARD items. The most complete in their methodological description were VANEDELA and EDI. The areas lacking more data on the reports were recruiting and patient selection (VANEDELA, PTNI, CVDL, INDIPCD-R). In QUADAS evaluation, all had some risk bias, but some serious concerns of risk bias were raised by patient sampling and by the choice of gold standard in two tests (PTNI, INDIPCD-R).

Conclusions: Child development screening tests created and validated in Mexico have variable report quality and risk bias. The test with the best validation report quality is VANEDELA and the one with the lowest risk of bias is EDI.

© 2015 Hospital Infantil de México Federico Gómez. Published by Masson Doyma México S.A. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introducción

La identificación temprana de las alteraciones en el desarrollo infantil es esencial para el bienestar de los niños y sus familias, pues permite que se proporcione un diagnóstico adecuado y una pronta intervención en aquellos que presenten alguna alteración¹. Para quien las recibe, dichas intervenciones están asociadas con una mejor funcionalidad en la vida adulta en múltiples esferas², además de que presentan una relación costo-beneficio muy alta³.

La comparación de las características clínicas de cada prueba es importante para auxiliar en la elección del instrumento más adecuado en la evaluación del desarrollo. Además, resulta importante comparar la calidad de reporte y el riesgo de sesgo de las mismas, puesto que los reportes de resultados de exámenes diagnóstico o de tamiz sesgados pueden dar como resultado la adopción generalizada de pruebas que produzcan una clasificación de riesgo imprecisa, y por tanto llevar al personal de salud a tomar decisiones de referencia, diagnóstico o tratamiento, incorrectas⁴.

En una revisión sistemática y análisis comparativo de la literatura, realizados en 2012 por Romo-Pardo y

colaboradores, se encontraron 13 pruebas de tamiz creadas y validadas en América para la identificación oportuna de problemas en el desarrollo infantil, pero ninguna con datos publicados en revistas científicas para México (excepto Denver-II)⁴. En los últimos años se ha creado un número importante de pruebas de tamiz, algunas de las cuales ya se aplican a la población menor de 5 años en México, y cuya información sobre validación y propiedades no se encuentra publicada en revistas indizadas. Tampoco se ha realizado una comparación entre las mismas.

Con base en lo anterior, el objetivo del presente artículo fue comparar la calidad del reporte de validación publicado y el riesgo de sesgo entre las pruebas de tamiz desarrolladas y validadas en México.

2. Métodos

2.1. Búsqueda y análisis

Dado que parte de la información sobre las pruebas de México no está disponible en motores de búsqueda de revistas científicas, durante octubre de 2015 se realizó una

búsqueda exhaustiva de herramientas de escrutinio para el desarrollo infantil en menores de 5 años, desarrolladas y validadas en México, desde 1980 a la fecha. Se llevó a cabo una estrategia simple utilizando los términos "neurodesarrollo" o "desarrollo infantil" y "tamiz", así como los términos "child development" y "screening", en PubMed/MEDLINE, Scopus, Web Of Science, EMBASE, EBSCO, Google Scholar, LILACS y SciELO limitando los resultados a pruebas realizadas en México.

Mediante las distintas búsquedas, la referencia cruzada documental y de preguntar a expertos en el tema, se identificaron siete pruebas de tamiz elaboradas en México para evaluar el neurodesarrollo. Con el nombre de las pruebas, se realizó una búsqueda exhaustiva de publicaciones que reportaran la validación de las mismas. Además de las fuentes consultadas previamente, se investigó en la literatura gris. Para el análisis, se incluyeron solamente aquellas pruebas en las cuales se encontró validación.

2.2. Instrumentos utilizados para la evaluación de las pruebas

2.2.1. STARD

Los estándares para reportar la precisión diagnóstica (STARD, por sus siglas en inglés)^{5,6} fueron diseñados con el objetivo de mejorar la calidad de reporte en los estudios de precisión diagnóstica. Consisten en una lista de cotejo de 25 ítems y un diagrama de flujo que informa sobre el método de selección de los pacientes, el orden de realización de las pruebas y el número de pacientes que deben evaluarse mediante la prueba índice y la prueba de referencia. Evalúa que las publicaciones proporcionen información suficiente que permita detectar el sesgo potencial en un estudio y juzgar la potencial generalización y aplicabilidad de los resultados.

2.2.2. QUADAS

La herramienta de calidad de la evaluación de los estudios de precisión diagnóstica en revisiones sistematizadas (QUADAS, por sus siglas en inglés)⁷ fue desarrollada para auxiliar en la evaluación del riesgo de sesgo de las publicaciones de precisión diagnóstica. Fue elaborada basándose en tres exámenes de pruebas existentes y un proceso con metodología Delphi con la participación de un grupo de nueve expertos. Consta de una lista de 14 ítems cualitativos que incluye las características de los pacientes, el estándar de referencia, el sesgo de progresión de la enfermedad, el sesgo de verificación, el sesgo de revisión, el sesgo de revisión clínica, el sesgo de incorporación, la prueba de ejecución y los resultados indeterminados. Estos ítems deben ser calificados mediante la respuesta "sí", si se considera que el estudio analizado cuenta con las características descritas en cada ítem; "no", si no cumple con las características; o "poco claro", si el texto no contiene la información suficiente para establecer un juicio.

2.2.3. QUADAS-2

El instrumento QUADAS-2⁸ se diseñó con el objetivo de evaluar el riesgo de sesgo en los estudios de precisión diagnóstica. Consta de un cuestionario estructurado con preguntas abiertas agrupadas en cuatro dominios, que incluyen

la selección de pacientes, la prueba índice, el estándar de referencia y el tiempo transcurrido entre la prueba índice y el estándar de referencia. La herramienta debe ser completada en cuatro fases: la primera pretende establecer la pregunta de revisión; la segunda, desarrollar una guía específica de revisión; posteriormente, evaluar el diagrama de flujo publicado o, en caso de no haber sido publicado, realizar uno con los datos proporcionados; y, finalmente, establecer un juicio de sesgo y aplicabilidad. Cada dominio se evalúa en términos del *riesgo de sesgo* y los tres primeros también son evaluados en términos de *preocupaciones con respecto a la aplicabilidad*. Para ayudar a establecer un juicio sobre el riesgo de sesgo, se incluyen preguntas de señalización. Esta herramienta permite realizar una presentación tabular para cada estudio evaluado, calificando cada ítem como *bajo riesgo*, *alto riesgo* o *riesgo poco claro*.

2.3. Procedimiento de la evaluación

El análisis se efectuó en fases. En una primera fase, cada uno de los autores evaluó, por separado y de forma independiente, la calidad del reporte de validación, mediante el uso de la herramienta STARD⁶, y el riesgo de sesgo en estudios, mediante las herramientas QUADAS⁷ y QUADAS-2⁸. Los resultados de dicha evaluación, incluyendo flujogramas y listas de cotejo, fueron recolectados utilizando los formatos desarrollados para cada instrumento. En una segunda fase, se compararon los formatos de evaluación de cada uno de los autores. Las discrepancias fueron resueltas por consenso. Se produjo una evaluación final que fue vaciada en los formatos antes mencionados y transformada para su presentación gráfica.

3. Resultados

3.1. Identificación de las pruebas de tamiz

Se identificaron siete pruebas de tamiz de alteraciones en el neurodesarrollo creadas en México: la Escala de Desarrollo Integral del Niño⁹, la Evaluación del Neurodesarrollo del Neonato (EVANENE)¹⁰, la Valoración Neuroconductual del Desarrollo del Lactante (VANEDELA)¹¹, Evaluación del Desarrollo Infantil (EDI)¹², Prueba de Tamiz del Neurodesarrollo infantil (PTNI)¹³, Cartillas de Vigilancia para identificar alteraciones en el Desarrollo del Lactante (CVDL)¹⁴ e Indicadores de riesgo del Perfil de Conductas de Desarrollo (INDIPCD-R)¹⁵.

Todos las publicaciones que se encontraron en los motores de búsqueda fueron relacionadas con la prueba EDI¹⁶⁻¹⁸, con excepción de una relacionada con NPED (Neuropediatric Development)¹⁹, la cual fue excluida ya que evaluaba un instrumento desarrollado en Cuba y no tenía un diseño de validación concurrente. El resto fue localizado en publicaciones científicas no indizadas ni indexadas (INDIPCD-R¹⁵, CVDL²⁰), en páginas web (PTNI²¹), o como libros, manuales, protocolos institucionales de investigación o tesis de grado (EVANENE²², VANEDELA²³, EDI¹⁶⁻¹⁸, PTNI²¹).

Tres contaban con artículos de validación publicados, una contaba con datos de la validación publicados en línea y otra como reporte de tesis. Finalmente, se incluyeron cinco pruebas para ser evaluadas. Se excluyeron de este estudio

la Escala de Desarrollo Integral del Niño, por no encontrar datos sobre su validación, y EVANENE, por el mismo motivo, ya que se localizó solamente una tesis de validación de la misma pero como instrumento de tamiz de daño cerebral.

3.2. Características generales de las pruebas evaluadas

Las características generales de las pruebas evaluadas se describen en la [tabla 1](#). Se encontró una gran variedad de rango de edades evaluadas. EDI es la prueba que evalúa el rango de edad más amplio (1-59 meses). La mayoría de las pruebas utiliza la observación directa como modalidad de evaluación excepto EDI, que tiene una modalidad de evaluación mixta: observación directa/interrogatorio dirigido. Ninguna utiliza la modalidad exclusiva de interrogatorio a través del llenado de un cuestionario por los padres. Las áreas evaluadas muestran una gran variabilidad, aunque en general se ajustan a las áreas recomendadas por la literatura: motora, lenguaje, adaptativo o cognitivo, personal o social^{24,25}.

Solamente INDIPCD-R, EDI y VANEDELA presentan la evaluación de signos neurológicos. Los sistemas de calificación utilizados difieren ampliamente, pero están adecuadamente descritos. Lo mismo se puede afirmar de los criterios de anormalidad, con excepción de los del INDIPCD-R, cuya definición publicada resulta confusa ya que no es claro cómo se compara el puntaje obtenido con el del estándar de oro. Las propiedades de las pruebas de tamiz reportadas en los estudios de validación revisados se encuentran resumidas en la [tabla 2](#).

Los tamaños de muestra para realizar las validaciones variaron. El más amplio fue el reportado por la PTNI. La selección de la muestra también varió, ya que se encontraron dos pruebas que se validaron en poblaciones provenientes de instituciones de salud (VANEDELA e INDIPCD-R), dos que obtuvieron su muestra en poblaciones específicas (PTNI en rural y CVDL en urbana) y una que presentó una selección de muestra intencionalmente balanceada en cuanto a características demográficas y factores de riesgo biológico (EDI).

El estándar de oro utilizado en la validación también fue diferente para las diversas pruebas: tres pruebas utilizaron una prueba diagnóstica del neurodesarrollo, como el Examen Evolutivo de Gessell (EEG)²⁶ o el Inventario de Desarrollo de Battelle-2 (IDB-2)²⁷ (VANEDELA, EDI, CVDL). Una prueba utilizó la prueba diagnóstica de la cual deriva (PCD-R)^{28,29}. Otra prueba utilizó como indicador *proxy* una serie de medidas de estado nutricional, anemia y crecimiento, a solas y como puntaje compuesto (PTNI).

Todas las pruebas reportaron valores de sensibilidad y especificidad así como valores predictivos positivos (VPP) y negativos (VPN) adecuados con lo recomendado por la literatura²⁴, aunque hubo pruebas que presentaron una amplia variación, dependiendo del grupo de edad evaluado (VANEDELA) o del estándar de oro utilizado (PTNI). En algunos rangos de edad fueron demasiado bajos para ser utilizados. Tres pruebas no describieron intervalos de confianza de sus datos (VANEDELA, PTNI, CVDL).

3.3. Resultados de la evaluación de la calidad del reporte

En la [tabla 3](#) se muestra la lista de cotejo de la herramienta STARD. Ninguna de las pruebas cumplió con la totalidad de los ítems a reportar. Las más completas en su descripción de métodos fueron VANEDELA y EDI. Las áreas en las que se encontraron más pruebas con datos faltantes o incompletos fueron las referentes a la descripción del procedimiento de muestreo y selección de pacientes (VANEDELA, PTNI, CVDL, INDIPCD-R), los métodos mediante los cuales se describieron los datos faltantes y los casos que se perdieron al seguimiento (todas), la descripción de los participantes y el flujograma del estudio (PTNI, CVDL, INDIPCD-R, EDI) y los métodos establecidos para medir la reproducibilidad de la prueba (INDIPCD-R).

3.4. Resultados de la evaluación del riesgo de sesgo

La [tabla 4](#) muestra los resultados de la aplicación de la herramienta de evaluación de riesgo de sesgo QUADAS. Ninguna prueba cumplió al 100% los criterios evaluados. Solo una prueba evaluó un espectro de pacientes representativo de la población (EDI). Ninguna reportó manejo de resultados no interpretados, abandono del estudio u otras pérdidas de casos. Los resultados de la evaluación QUADAS-2 se muestran en la [tabla 5](#) y son una valoración cualitativa final que representa el consenso de la opinión de los autores sobre qué tanto riesgo de sesgo presenta cada una de las pruebas. Debido a la validación mediante el estándar de oro *proxy*, la PTNI presenta un alto riesgo de sesgo con respecto a su estándar de referencia (peso para la edad, talla para la edad, anemia y estimulación temprana que no es independiente a la prueba). La INDIPCD-R presenta un alto riesgo de sesgo de su prueba índice y estándar de oro ya que se valida contra sí misma. Al tratarse de muestras por conveniencia, sin ningún tipo de ajuste, todas las pruebas presentan un alto riesgo de sesgo con respecto a la selección de sus pacientes. La que menos riesgo tuvo fue EDI ya que, intencionalmente, balanceó los grupos evaluados.

4. Discusión

En previas revisiones⁴ y reportes^{14,15} es repetitivo el comentario sobre las escasas o nulas menciones de pruebas mexicanas en la literatura, así como la dificultad de encontrar las validaciones de pruebas latinoamericanas. El escaso número de resultados obtenido al realizar búsquedas sencillas limitando los resultados a México podría explicarse por dos factores: el primero sería el bajo nivel de visibilidad de las revistas latinoamericanas, ya que muchas no se encuentran indexadas e indizadas, y por tanto no aparecen en los motores de búsqueda³⁰. El segundo, por una mezcla entre un probable "temor" a las publicaciones internacionales y un efecto de "malinchismo" hacia las nacionales³¹, que dificulta la decisión de iniciar el laborioso proceso de transformación de textos de tesis de grado en artículos científicos³².

Como no existe una prueba de tamiz del desarrollo ideal, se puede decir que las características generales de las prue-

Tabla 1 Descripción general de las pruebas de tamiz comparadas en el estudio

Prueba	Rango de edad	Modalidad de evaluación	Aspectos evaluados	Sistema de calificación	Criterios de normalidad y anormalidad
Valoración Neuroconductual del desarrollo del Lactante (VANEDELA) ²³	1-24 m (evaluados en 6 grupos)	Observación directa	Conductas del desarrollo (alimentación, reflejos, postura antigravitatoria, motor grueso, motor fino, coordinación ojo-mano, percepción visual, percepción auditiva, lenguaje receptivo, lenguaje expresivo, cognoscitivo, emocional, social) Reacciones del desarrollo (reflejos) Signos de alteración o alarma	Cada pregunta en conductas del desarrollo y reacciones del desarrollo tiene tres posibles respuestas: sí, no o no aplica Los signos de alteración o alarma tienen tres posibles respuestas: ausente, presente o no registrado El resultado se expresa en normal, dudoso, anormal y no aplicable (criterio de validez si dos ítems son marcados como no aplicables)	Se obtiene un puntaje a través de las 10 conductas solicitadas en los ítems por grupo de edad, estableciendo como normal si los hace todos, dudosos si hace 8-9 y anormal si hace menos de 7
Prueba Evaluación del Desarrollo Infantil (EDI) ¹⁶	1-59 m (evaluados en 14 grupos)	Interrogado y observado	Áreas del desarrollo (motor fino, motor, grueso, lenguaje, social y conocimiento), factores de riesgo biológico, señales de alerta, señales de alarma y exploración neurológica	Cada pregunta tiene dos posibles respuestas: sí o no El resultado está basado en un sistema de semáforo: desarrollo normal (verde), rezago en el desarrollo (amarillo) y riesgo de retraso (rojo)	Cuenta con criterios de calificación específicos para grupos de edad (1 m, 2-4 m y 5-59 m)
Prueba de Tamiz para Evaluar en Neurodesarrollo Infantil (PTNI) ²¹	11 a 49 m (evaluados en 5 grupos)	Observación directa	Áreas del desarrollo (motor gruesa, motor fina, lenguaje, cognoscitiva, socio-afectiva, independencia)	Cada ítem tiene dos posibles respuestas: sí o no (se otorga un punto por cada respuesta afirmativa). El resultado se expresa en cuatro categorías: neurodesarrollo normal, retraso leve (riesgo leve), retraso moderado (riesgo moderado) y retraso grave (alto riesgo)	Se obtiene un puntaje a través de las 12 conductas solicitadas en los ítems por grupo de edad, estableciendo puntos de corte específicos para niños de 12 a 24 m y otros para los de 36 a 48 m
Cartillas de Vigilancia para identificar alteraciones en el Desarrollo del Lactante (CVDL) ²⁰	1 a 24 m (evaluados en 13 grupos)	Interrogado	No se agrupan por áreas ni dominios	Los ítems están ordenados por complejidad del 1 al 64. La cartilla establece qué número de ítem debe realizar a cada edad determinada. El resultado se expresa de forma dicotómica en riesgo de retraso o desarrollo normal	Se evalúa que el niño realice las conductas indicadas por la cartilla y las del grupo etario anterior, para clasificarlo como desarrollo normal
Indicadores de conductas de riesgo del Perfil de Conductas del Desarrollo (INDIPCD-R) ¹⁵	0 a 4 a (evaluados en 6 grupos)	Interrogado	Áreas del desarrollo (sedestación, gateo, bipedestación, marcha, lenguaje expresivo, lenguaje receptivo, emocional/social, habilidad manual, cognición y praxis), antecedentes	Cada pregunta se califica usando una escala ordinal, dependiendo si una conducta se observa frecuentemente (75-100%), a veces (15-75%) o nunca (0-15%). El resultado se expresa en alteración del desarrollo leve, moderada o severa	Se obtiene un puntaje al sumar los indicadores de riesgo a la edad cronológica del niño. Este puntaje se compara con una escala normativa obtenida por DE: <1DE, leve; 1-2DE, alteración moderada; >2DE, alteración severa

m: meses; a: años; DE: desviación estándar.

Tabla 2 Comparación de las propiedades de las pruebas de tamiz reportadas en los estudios de validación

Prueba (año de publicación)	Población estudiada	Estándar de oro y punto de corte para anormalidad	Sensibilidad	Especificidad	VPP	VPN
VANEDELA (2001) ²³	184 niños de una población urbana que acuden al laboratorio de seguimiento del neurodesarrollo en INP/UAM-X	CTD < 90 en el EEG o al menos un área del desarrollo evaluada en el EEG con un CDD < 76	1 m: 1.00 4 m: 0.88 8 m: 0.88 12 m: 0.96 18 m: 0.88 24 m: 0.88	1 m: 0.93 4 m: 0.75 8 m: 0.67 12 m: 0.45 18 m: 0.67 24 m: 0.78	1 m: 0.83 4 m: 0.91 8 m: 0.81 12 m: 0.69 18 m: 0.70 24 m: 0.78	1 m: 1.00 4 m: 0.69 8 m: 0.78 12 m: 0.90 18 m: 0.86 24 m: 0.88
Versión modificada de EDI (2013) ^{16,18}	438 niños en tres estados del país (202 niños de 16-60 m)	CTD < 90 en el IDB-2 ¹⁶	1-15 m: 0.74 IC (0.65-0.82)	1-15 m: 0.60 IC (0.51-0.68)	1-15 m: 0.61 IC (0.53-0.70)	1-15 m: 0.72 IC (0.63-0.81)
			16-60 m: 0.89 IC (0.82-0.95) Total 0.81 IC (0.64-0.86)	16-60 m: 0.62 IC (0.53-0.71) Total 0.61 IC (0.54-0.67)	16-60 m: 0.68 IC (0.60-0.76) Total 0.65 IC (0.59-0.70)	16-60 m: 0.86 IC (0.78-0.94) Total 0.78 IC (0.72-0.84)
		CDD < 80 en IDB-2 ¹⁸	Motor 0.84 IC(0.74-0.90) Comunicación 0.79 IC (0.71-0.85) Personal social 0.86 IC(0.77-0.92) Adaptativo 0.91 IC(0.76-0.96) Cognitivo 0.83 IC(0.76-0.88)	Motor 0.83 IC(0.87-0.90) Comunicación 0.79 IC(0.74-0.83) Personal social 0.85 IC(0.80-0.88) Adaptativo 0.85 IC(0.81-0.88) Cognitivo 0.88 IC(0.80-0.93)		

Tabla 2 (continuación)

PTNI (2013) ²¹	27,059 niños en comunidades rurales de México	Peso para la edad (puntaje Z ≤ 2)	12 m: 0.78	24 m: 0.73	12 m: 0.41	24 m: 0.52	12 m: 0.18	24 m: 0.22	12 m: 0.92	24 m: 0.91
		Talla para la edad (puntaje Z ≤ 2)	12 m: 0.71	24 m: 0.62	12 m: 0.44	24 m: 0.58	12 m: 0.60	24 m: 0.71	12 m: 0.56	24 m: 0.48
		Anemia (no específica punto de corte)	12 m: 0.65	24 m: 0.60	12 m: 0.65	24 m: 0.60	12 m: 0.44	24 m: 0.28	12 m: 0.60	24 m: 0.78
		Estimulación (no específica punto de corte)	12 m: 0.92	24 m: 0.89	12 m: 0.60	24 m: 0.64	12 m: 0.50	24 m: 0.36	12 m: 0.94	24 m: 0.97
		Todas las variables (no específica punto de corte ni criterios de combinación)	12 m: 0.88	24 m: 0.86	12 m: 0.77	24 m: 0.79	12 m: 0.85	24 m: 0.74	12 m: 0.83	24 m: 0.91
CVDL (2014) ²⁰	2,702 niños procedentes de zonas urbanas de tres estados del país	Resultado menor a -2 DE en cualquiera de las escalas de dos pruebas diagnósticas: Escala de Desarrollo de Bayley II o en la EEG, "diagnóstico normal o anormal del niño"	0.84		0.75		0.83		0.76	
INDIPCD-R (2015) ¹⁵	145 niños provenientes de una clínica y dos CENDI	Perfil de coeficientes del desarrollo o PCD-R (no menciona el punto de corte)	Clínica 1.00 IC (0.96-1.00) CENDI 0.94 IC (0.82-1.00)		Clínica 0.69 IC (0.40-98.17) CENDI 0.84 IC (0.61-1.00)		Clínica 0.90 IC (0.80-1.00) CENDI 0.90 IC (0.74-1.00)		Clínica 1.00 IC (0.94-1.00) CENDI 0.91 IC (0.71-1.00)	

VANEDELA: Valoración Neuroconductual del Desarrollo del Lactante; EDI: Evaluación del Desarrollo Infantil; PTNI: Prueba de Tamiz del Neurodesarrollo Infantil; CVDL: Cartillas de Vigilancia para identificar alteraciones en el Desarrollo del Lactante; INDIPCD-R: Indicadores de riesgo del Perfil de Conductas de Desarrollo Revisado; VPP: valor predictivo positivo; VPN: valor predictivo negativo; m: meses; INP: Instituto Nacional de Pediatría; UAM-X: Universidad Autónoma Metropolitana plantel Xochimilco; DE: desviaciones estándar; IDB-2: Inventario de Desarrollo de Battelle 2^a. ed. en español; CTD: cociente toral de desarrollo; EEG: Examen evolutivo de Gessell; CDD: cociente de desarrollo del dominio; PCD-R: perfil de conductas del desarrollo; CENDI: centro de desarrollo infantil.

Tabla 3 Lista de cotejo STARD para evaluar calidad de reporte de estudios de precisión diagnóstica

Lista de cotejo del instrumento STARD			VANEDELA ²³	EDI ¹⁶	PTNI ²¹	CVDL ²⁰	INDIPCD-R ¹⁵
Sección y tema	Ítem	Descripción del ítem	Sí/incompleto/no; número de página (p)				
Título/resumen/ palabra clave Introducción	1	Identificar el artículo como estudio de precisión diagnóstico	Sí; p. 1	Sí; p. 196	No	Incompleto	Sí; p. 135
	2	Mencionar explícitamente las preguntas de investigación o los objetivos del estudio	Sí; p. 91	Sí; p. 197	No	Sí, p. 25	Sí; p. 138
Métodos Participantes	3	Población del estudio, criterios de inclusión y exclusión	Sí; pp. 91, 92	Sí; p. 197	Incompleto; p. 44	Incompleto; p. 26	Incompleto; p. 139
	4	Reclutamiento de pacientes: ¿qué criterio se utilizó para su selección?	Sí; pp. 92, 97	Incompleto; p. 197	Sí; p. 44	No; p. 26	Sí; p. 138
	5	Muestreo: procedimientos	Sí; pp. 92, 97	Incompleto; p. 197	Incompleto; p. 44	Incompleto; p. 26	Sí; p. 138
	6	Momento de recolección de datos: ¿fue planeado antes o después del estudio?	Sí; pp. 92, 97	Incompleto; p. 197	No	Incompleto; p. 26	No se menciona
Pruebas	7	El estándar de la prueba y su racional	Sí; pp. 29-93	Sí; p. 197	Sí; p. 40	Sí; p. 26	Incompleto; p. 138
	8	Especificaciones técnicas de material y métodos involucrados	Sí; pp. 29-97	Sí; p. 197	Sí; pp. 46-48	Sí; p. 26	Sí; p. 139
	9	Definición y explicación para las unidades utilizadas, puntos de corte, resultados de test índice y estándares de referencia	Sí; pp. 29-95	Sí; pp. 198, 201	Incompleto; pp. 41, 42	Sí; p. 26	Sí; pp. 138, 139
	10	Número, entrenamiento y pericia de las personas ejecutando e interpretando las pruebas de referencia y normales	Incompleto; p. 95	Sí; p. 197	Incompleto; p. 43	Incompleto; p. 26	Incompleto; p. 139
	11	Mencionar si hubo un procedimiento ciego para evaluar las pruebas	No	Sí; p. 197	No	No hubo; p. 26	No hubo; p. 139
Estadísticos	12	Métodos estadísticos para calcular o comparar medidas de precisión	Incompleto; pp. 95-101	Sí; p. 197	No	No; p. 26	Sí; pp. 139-140
	13	Métodos para calcular la reproducibilidad de las pruebas	No	Incompleto; p. 197	Incompleto; p. 59	Incompleto; p. 26	No

Tabla 3 (continuación)

Lista de cotejo del instrumento STARD			VANEDELA ²³	EDI ¹⁶	PTNI ²¹	CVDL ²⁰	INDIPCD-R ¹⁵
<i>Resultados</i>							
Participantes	14	¿Cuándo fue realizado el estudio? ¿Se mencionan las fechas de inicio del reclutamiento?	Sí; p. 97	No; p. 197	Incompleto; p. 44	No; pp. 26, 27	No
	15	Características demográficas de la población: edad, sexo, espectro de síntomas presentados	Sí; pp. 98, 99	Sí; pp. 198, 200	Incompleto; pp. 45, 46	No; pp. 26, 27	Sí; p. 140
	16	Número de pacientes que cumplían con criterios de inclusión, que fueron evaluados por ambos test, referencia e índice, con flujograma	Incompleto; p. 100	No	No se refiere	No; p. 26	Incompleto; pp. 138, 139
	17	Intervalo de tiempo entre la aplicación del test índice y el estándar de referencia	Sí; p. 92	Sí; p. 197	No se refiere	Sí; p. 26	No
	18	Distribución de severidad del padecimiento	Incompleto; pp. 98, 99	Sí; pp. 197, 198	No se refiere	No; p. 26	Sí; p. 141
	19	Una tabulación cruzada entre los resultados del test índice y el estándar	Sí; pp. 100-133	Sí; p. 202	Sí; pp. 50, 51	Sí; pp. 26, 27	Sí; p. 142
	20	Eventos adversos de aplicación del test de referencia o índice	No aplica	No aplica	No aplica	No	No aplica
	21	Estimados de precisión diagnóstica y medidas de incertidumbre	Incompleto	Sí	Incompleto; pp. 51, 52	Incompleto	Sí; p. 142
	22	Forma de manejar/clasificar resultados indeterminados, datos faltantes, distribución del test índice	Incompleto; p. 98	Incompleto	No	No	Incompleto; p. 139
	23	Estimados de la variabilidad de la precisión diagnóstica entre subgrupos de participantes, centros de evaluación y evaluadores	No aplica	Sí; p. 203	No	No	Sí; p. 142
	24	Estimados de la reproducibilidad del test	No	Incompleto; p. 197	Sí	Incompleto	No
Discusión	25	Discusión sobre los resultados del estudio	Sí; p. 143	Sí; p. 207	No	Sí	Sí; pp. 142-144

STARD: Standards for Reporting of Diagnostic Accuracy; VANEDELA: Valoración Neuroconductual del Desarrollo del Lactante; EDI: Evaluación del Desarrollo Infantil; PTNI: Prueba de Tamiz del Neurodesarrollo Infantil; CVDL: Cartillas de Vigilancia para identificar alteraciones en el Desarrollo del Lactante; INDIPCD-R: Indicadores de riesgo del Perfil de Conductas de Desarrollo Revisado.

Tabla 4 Cuestionario para evaluar riesgo de sesgos en estudios de precisión diagnóstica QUADAS

Ítem	VANEDELA ²³	EDI ¹⁶	PTNI ²¹	CVDL ²⁰	INDIPCD-R ¹⁵
¿El espectro de pacientes evaluados fue representativo de aquellos que recibirán la prueba en la práctica?	No	Sí	No	No	No
¿Los criterios de selección están claramente definidos?	Sí	Sí	Sí	Incierto	No
¿Es muy probable que el estándar de referencia clasifique la condición blanco adecuadamente?	Sí	Sí	No	Sí	No
¿El intervalo de tiempo entre la aplicación de las dos pruebas es suficientemente corto como para afirmar que la condición no cambió?	Sí	Sí	Incierto	Sí	No
¿Toda la muestra recibió evaluación confirmatoria o solamente una submuestra de la misma?	Sí	Sí	Sí	Sí	No
¿Fue la muestra evaluada con la misma prueba diagnóstica, independientemente del resultado en la prueba índice?	Sí	Sí	Sí	Sí	Sí
¿La prueba índice es independiente de la prueba de referencia?	Sí	Sí	Sí	Sí	No
¿La ejecución de la prueba índice fue descrita con suficiente detalle como para ser replicada?	Sí	Sí	Sí	Sí	Sí
¿La ejecución de la prueba de referencia fue descrita con suficiente detalle como para ser replicada?	Sí	Sí	Incierto	Incierto	Sí
¿Fueron los resultados del test de referencia interpretados sin conocimiento del resultado del test índice?	Sí	Sí	No	Incierto	Incierto
¿Fueron los resultados del test índice interpretados sin conocimiento del resultado del test de referencia?	Sí	Sí	No	Incierto	Incierto
¿Existía la misma información clínica disponible para interpretar el test índice que en la práctica clínica común?	Sí	Sí	Sí	Sí	Sí
¿Se reportaron resultados intermedios/no interpretados?	No	No	No	No	No
¿Se explican las pérdidas de casos?	No	No	No	No	No

VANEDELA: Valoración Neuroconductual del Desarrollo del Lactante; EDI: Evaluación del Desarrollo Infantil; PTNI: Prueba de Tamiz del Neurodesarrollo Infantil; CVDL: Cartillas de Vigilancia para identificar alteraciones en el Desarrollo del Lactante; INDIPCD-R: Indicadores de riesgo del Perfil de Conductas de Desarrollo Revisado.

bas las hacen adecuadas para su uso en México. Antes de considerar los aspectos de la calidad de validación, la decisión de utilizar una sobre otra debe considerar su flexibilidad de uso²⁴. En este sentido, el rango de edades evaluadas, la variedad en modalidades de evaluación y la sencillez del

sistema visual de calificación por "semáforo" favorecen a la prueba EDI.

Ningún estudio se encuentra libre de fallas en la calidad de su reporte. El más completo fue el reporte de VANEDELA, muy amplio y completo al ser una tesis de 180 páginas.

Tabla 5 Consenso de resultados de la aplicación de la herramienta de evaluación de riesgo de sesgo QUADAS-2

Estudio	Riesgo de sesgo				Preocupaciones sobre aplicabilidad		
	Selección de pacientes	Prueba índice	Estándar de referencia	Diagrama de flujo y secuencia del estudio	Selección de pacientes	Prueba índice	Estándar de referencia
VANEDELA ²³							
EDI ¹⁶							
PTNI ²¹							
CVDL ²⁰							
INDIPCD-R ¹⁵							

Bajo Riesgo Alto Riesgo Riesgo poco claro

VANEDELA: Valoración Neuroconductual del Desarrollo del Lactante; EDI: Evaluación del Desarrollo Infantil; PTNI: Prueba de Tamiz del Neurodesarrollo Infantil; CVDL: Cartillas de Vigilancia para identificar alteraciones en el Desarrollo del Lactante; INDIPCD-R: Indicadores de riesgo del Perfil de Conductas de Desarrollo Revisado.

Si bien se encuentran casi todos los datos requeridos por STARD para valorar la confiabilidad de los datos reportados, hay aspectos metodológicos que limitan su validez externa, como el tamaño de muestra pequeño de cada uno de los grupos etarios evaluados y el proceso de reclutamiento de la muestra, que fue por conveniencia en una población urbana proveniente de ambientes clínico-hospitalarios. Otros reportes de validación, como el del INDIPCD-R, omiten datos importantes para poder evaluar su validez, lo que dificulta su evaluación objetiva. Una posible solución a este fenómeno podría ser el extender el uso de las herramientas de evaluación utilizadas en este artículo como una lista de cotejo guía para garantizar que los escritos científicos estén completos antes de publicarse³³.

No existe publicación científica que esté libre de sesgo; sin embargo, existen procedimientos para reducirlos. Los resultados de las dos evaluaciones finales (QUADAS y QUADAS-2) muestran que, si bien los datos de sensibilidad y especificidad son, *a grosso modo*, similares entre las pruebas comparadas, la validez de estos datos está comprometida en distintos grados. Algunas de las publicaciones evaluadas omiten datos básicos, como las medidas de dispersión de datos o precisión de la medición, y otras comprometen el procedimiento de validación concurrente al comparar la prueba contra sí misma o una medición *proxy*. Otras pruebas, como VANEDELA, también presentan alto riesgo de sesgo por lo pequeño de la muestra.

De manera similar a lo concluido en revisiones comparativas de pruebas de tamiz del desarrollo realizadas en otras partes del mundo³⁴, se encontró que de entre las pruebas para tamiz del neurodesarrollo creadas en México no hay ninguna que sea perfecta. La más flexible en su aplicación y con menor riesgo de sesgo en sus resultados de validación fue la prueba EDI.

Las pruebas de tamiz del neurodesarrollo, creadas y validadas en México, son, en general, adecuadas para su uso, aunque presentan calidades de reporte de publicación y riesgo de sesgo variables, y ninguna es perfecta. La prueba con mejor calidad de reporte de validación es VANEDELA, seguida de EDI, y la de menor riesgo de sesgo en los datos publicados es la prueba EDI. Se requerirá de un estudio comparativo entre las pruebas de tamiz contra el estándar de oro para poder establecer cuál de ellas es la que tiene las mejores propiedades.

Financiamiento

Este trabajo no contó con financiamiento alguno.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Agradecimientos

Se agradece al Lic. Josué Laguna Hernández de la Sección de Archivo Histórico, Biblioteca Dr. Ramón Villarreal Pérez, Universidad Autónoma Metropolitana-plantel Xochimilco, por el apoyo brindado en la localización de una de las tesis evaluadas en este estudio.

Referencias

1. Council on Children with Disabilities; Section on Developmental Behavioral Pediatrics; Bright Futures Steering Committee; Medical Home Initiatives for Children with Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. *Pediatrics*. 2006;118:405–20.
2. Hamilton S. Screening for developmental delay: reliable, easy-to-use tools. *J Fam Pract*. 2006;55:415–22.
3. Glascoe FP, Foster EM, Wolraich ML. An economic analysis of developmental detection methods. *Pediatrics*. 1997;99:830–7.
4. Romo-Pardo B, Liendo-Vallejos S, Rizzoli-Córdoba G, Vargas-López A, Buenrostro-Márquez G. Global neurodevelopmental screening tests for children under 5 years of age in the United States and Latin America: systematic review and comparative analysis. *Bol Med Hosp Infant Mex*. 2012;69:450–62.
5. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ*. 2003;326:41–4.
6. Bossuyt P, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem*. 2003;49:7–18.
7. Whiting PF, Rutjes AWS, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:25.
8. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529–36.
9. Atkin LC, Supervielle T, Sawyer R, Cantón P. Paso a paso: cómo evaluar el crecimiento y desarrollo de los niños. México D.F.: UNICEF, Editorial Pax; 1987.
10. Alvarado-Ruiz GA, Sánchez-Pérez MC, Mandujano-Valdés MA. EVANENE. Evaluación del Neurodesarrollo del Neonato. México D.F.: Universidad Autónoma Metropolitana Unidad Xochimilco; 2013.
11. Sánchez-Pérez MC, Benavides-González H, Mandujano-Valdés M, Rivera-González IR, Martínez-Vázquez RI, Alvarado-Ruiz GA. Valoración Neuroconductual del Desarrollo del Lactante (VANEDELA). México D.F.: Universidad Autónoma Metropolitana; 2007.
12. Comisión Nacional de Protección Social en Salud. Manual para la Aplicación de la Prueba Evaluación del Desarrollo Infantil EDI. México, D.F.: Secretaría de Salud; 2013. 100p. Disponible en: <http://www.himfg.edu.mx/descargas/documentos/EDI/ManualparaPruebadeEvaluaciondelDesarrolloInfantil-EDI.pdf>
13. Ávila-Curiel CA, Álvarez-Izazaga MA, Lopez-Arce Coria AM, Medina-Carranza BE. Prueba de Tamiz del Neurodesarrollo Infantil. México, 2008. Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán. No. de registro 03-2008-120312060800.
14. Rivera R. Vigilancia del desarrollo en el primer nivel de atención. En: Rivera R, Sánchez C, editores. Vigilancia del desarrollo integral del niño. México: Editores de Textos Mexicanos; 2009. p. 9–15.
15. Bolaños C, Mina-Lara MM, Ramos-Ortega G. Indicadores de riesgo del Perfil de Conductas de Desarrollo como instrumento para la detección temprana de retrasos y alteraciones en el desarrollo. *Acta Pediatr Mex*. 2015;36:135–46.
16. Rizzoli-Córdoba A, Schnass-Arrieta L, Liendo-Vallejos S, Buenrostro-Márquez G, Romo-Pardo B, Carreón-García J, et al. Validación de un instrumento para la detección oportuna de problemas del desarrollo en menores de 5 años en México. *Bol Med Hosp Inf Mex*. 2013;70:195–208.

17. Rizzoli-Córdoba A, Ortega-Ríosvelasco F, Villasís-Keever MA, Pizarro-Castellanos M, Buenrosto-Márquez G, Aceves-Villagrán D, et al. Reliability for detection of developmental problems using the semaphore from the Child Development Evaluation test: Is a yellow result different from a red result. *Bol Med Hosp Infant Mex.* 2014;71:277-85.
18. Rizzoli-Córdoba A, Schnaas-Arrieta L, Ortega-Ríosvelasco F, Rodríguez-Ortega E, Villasís-Keever MA, Aceves-Villagrán D, et al. Child Development Evaluation Test analysis by field improves detection of developmental problems in children. *Bol Med Hosp Infant Mex.* 2014;71:154-62.
19. Guadarrama-Celaya F, Otero-Ojeda GA, Pliego-Rivero FB, Porcayo-Mercado MR, Ricardo-Garcell J, Perez-Ábalo MC. Screening of neurodevelopmental delays in four communities of Mexico and Cuba. *Public Health Nurs.* 2012;29:105-15.
20. Rivera-González R, Villanueva-Romero Y, Amaro-López L, Sánchez-Pérez C, Figueroa-Olea M, Soler-Limón KM. Validez concurrente de las cartillas vigilancia para identificar alteraciones en el desarrollo del lactante. *Ciencias Clínicas.* 2014;15:22-9.
21. Ávila-Curiel A, Álvarez-Izazaga MA, Reidl-Martínez L, López-Arce AM. Vigilancia Epidemiológica del Neurodesarrollo Infantil en Comunidades Marginales de México. Capítulo III Elaboración de la prueba de tamiz del neurodesarrollo infantil (PTNI). Disponible en: <http://www.nutricionemexico.org.mx/encuestas/neuro%20web.pdf>
22. Palma-Reyes CR. Valor diagnóstico de dos instrumentos clínicos en la detección de daño cerebral perinatal en prematuros. Informe preliminar (Tesis de Licenciatura). Universidad Autónoma de Querétaro. Querétaro, Querétaro, México; 2014.
23. Martínez-Vázquez RI. Evaluación de la sensibilidad y especificidad del instrumento de detección: Valoración Neuroconductual del Desarrollo del Lactante. (VANEDELA) (Tesis de Maestría). Universidad Autónoma Metropolitana-Xochimilco. México D.F., México; 2001.
24. Rydz D, Shevell MI, Majnemer A, Oskoui M. Developmental screening. *J Child Neurol.* 2005;20:4-21.
25. Shevell M, Majnemer A, Platt RW, Webster R, Birnbaum R. Developmental and functional outcomes in children with global developmental delay or developmental language impairment. *Dev Med Child Neurol.* 2005;47:678-83.
26. Ball RS. The Gesell developmental schedules: Arnold Gesell (1880-1961). *J Abnorm Child Psych.* 1977;5:233-9.
27. Newborg J. Development and standardization. En: Newborg J, editor. *Battelle Developmental Inventory.* Itaska, IL: Riverside Publishing; 2004. p. 95-148.
28. Bolaños C. Manual del Perfil de Conductas de Desarrollo Revisado. México D.F.: Instituto de Terapia Ocupacional; 2005.
29. Bolaños C. Estudio de validación del perfil de conductas de desarrollo (PCD) como instrumento para evaluar el nivel de desarrollo de los niños de 0-4 años. *Bol Med Hosp Infant Mex.* 2003;60:79-89.
30. Romero-Torres M, Acosta-Moreno LA, Tejada-Gómez MA. Ranking de revistas científicas en Latinoamérica mediante el índice *h*: estudio de caso Colombia. *Rev Esp Document Cient.* 2013;36:e003, doi: <http://dx.doi.org/10.3989/redc.2013.1.876>
31. Cicero-Sabido R. ¿Por qué los autores mexicanos no envían trabajos importantes a las revistas médicas mexicanas? Un comentario breve. *Gac Med Mex.* 2006;142:128-9.
32. Gross Portney L, Watkins MP. Reporting the results of clinical research. En: *Foundations of Clinical Research. Applications to Practice.* New York: Pearson Prentice Hall; 2008. p. 879-900.
33. Cook C, Cleland J, Huijbregts P. Creation and critique of studies of diagnostic accuracy: use of the STARD and QUADAS methodological quality assessment tools. *J Man Manip Ther.* 2007;15:93-102.
34. Bedford H, Walton S, Ahn J; Policy Research Unit in the Health of Children, Young People and Families; Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health. *Measures of Child Development: A Review.* London: University College London (UCL); 2013. 103p. Disponible en: https://www.ucl.ac.uk/cpru/documents/review_of_measures_of_child_development