

■ Propiedades psicométricas y evaluación del desarrollo infantil con la escala DAUCO Test

Sara María Luque de Dios¹ , Juan Antonio Moriana^{1,2,3} , Carolina Pérez-Dueñas^{1,2,3} , Eduardo Martínez-Gual⁴ , Araceli Camacho Alcaide³ , Ramón González Matilla^{2,3} , Adoración Antolí^{1,2,3} , Araceli Sánchez-Raya^{1,2,3} 

¹Universidad de Córdoba (España)

²Instituto Maimónides de Investigación Biomédica de Córdoba (IMIBIC) - Hospital Universitario Reina Sofía de Córdoba (España)

³Centro de Atención Infantil Temprana UCO (España)

⁴Fundación Termens (España)

Resumen

La evaluación del desarrollo se realiza combinando diferentes procedimientos como entrevistas, observación, cuestionarios y escalas. La mayoría de las escalas de desarrollo, a pesar de su uso generalizado, muestran importantes debilidades métricas y escasos estudios de validación. El objetivo de este estudio fue construir y validar la escala de desarrollo DAUCO destinada a menores de 84 meses a partir del análisis integrado de las escalas más utilizadas en este ámbito. Se realizó un juicio de expertos desde la compilación de todos los ítems extraídos de las escalas Merrill-Palmer-R, Battelle II, Currículo Carolina, IDAT, EOD y Bayley III. Se aplicaron análisis TCT, TRI, AFC y el ajuste de los datos al modelo en una muestra de 638 niños/as españoles. Las subescalas presentan buena fiabilidad, validez y ajuste de los datos al modelo, superando algunas limitaciones de las escalas predecesoras. El análisis integrativo de las escalas más importantes realizado para crear DAUCO supone un avance, mejorando su métrica, aplicabilidad y facilitando el diseño de las posteriores propuestas de intervención.

Palabras clave: Evaluación desarrollo infantil; Validación escalas de desarrollo; Teoría de la Respuesta al ítem; Modelo de Rasch; Teoría Clásica de Test.

Abstract

Psychometric properties and evaluation of child development with the DAUCO Test scale. Development evaluation is carried out by combining different procedures such as interviews, observation, questionnaires and scales. Most development scales, despite their widespread use, show important metric weaknesses and few validation studies. The objective of this study was to build and validate the DAUCO development scale intended for children under 84 months of age based on the integrated analysis of the most used scales in this area. An expert judgment was carried out from the compilation of all the items extracted from the Merrill-Palmer-R, Battelle II, Currículo Carolina, IDAT, EOD and Bayley III scales. TCT, TRI, CFA analysis and data adjustment to the model were applied in a sample of 638 Spanish children. The subscales present good reliability, validity and fit of the data to the model, overcoming some limitations of the predecessor scales. The integrative analysis of the most important scales carried out to create DAUCO represents progress, improving its metrics, applicability and facilitating the design of subsequent intervention proposals.

Keywords: Child development evaluation; Validation of development scales; Item Response Theory; Rasch model; Classic Test Theory.

Autor de correspondencia / Corresponding author: Araceli Sánchez-Raya (ed1saram@uco.es)

Citar como / Cite as: Luque de Dios, S. M., Moriana, J. A., Pérez-Dueñas, C., Martínez-Gual, E., Camacho Alcaide, A., González Matilla, R., Antolí, A., & Sánchez-Raya, A. (2025). Propiedades psicométricas y evaluación del desarrollo infantil con la escala DAUCO Test. *Revista de Psicología Clínica con Niños y Adolescentes*. Advance online publication.

Recibido / Received: 6 de junio de 2024

Aceptado / Accepted: 12 de diciembre de 2024

Puntos clave

1. Este artículo describe la construcción y validación de la escala de desarrollo DAUCO Test para menores de hasta 84 meses.
2. La construcción de esta escala de screening en formato digital parte del análisis bibliográfico, sustantivo y métrico de las escalas más importantes en este ámbito.
3. Las subescalas de DAUCO Test presentan buena fiabilidad, validez y ajuste de los datos al modelo, superando algunas limitaciones de las escalas anteriores.
4. Dauco Test discrimina de forma rápida y eficiente entre niños/as con desarrollo normal o con alteraciones o en riesgo de padecerlos, proporcionando un programa de intervención ajustado a los resultados.

Highlights

1. This article outlines the development and validation of the DAUCO Test, a developmental scale for children up to 84 months old.
2. The creation of this digital screening scale is based on a bibliographic, substantive, and metric analysis of the most significant scales in this field.
3. The DAUCO Test subscales demonstrate good reliability, validity, and data fit to the model, addressing some limitations of previous scales.
4. The DAUCO Test quickly and efficiently distinguishes between children with typical development and those with developmental delays or at risk, offering an intervention program tailored to the results.

Introducción

La evaluación del desarrollo infantil es un proceso dinámico dirigido a la obtención de información del menor y su entorno (Villagra et al., 2023) que resulta fundamental en la valoración, diagnóstico y aplicación de programas de intervención. Se establecen protocolos de valoración utilizando diferentes fuentes (Guralnick, 2023) y técnicas, combinando principalmente: entrevistas semiestructuradas, observación directa e informal en diferentes contextos y la aplicación de escalas de desarrollo, teniendo estas últimas un peso determinante (Sices et al., 2003). Actualmente, las principales escalas son multidimensionales (Figueiras et al., 2011) y estiman principalmente el desarrollo motor, cognitivo, socioemocional, adaptativo y el lenguaje y/o la comunicación. La mayoría de estas escalas están basadas en la evaluación de la consecución de hitos del desarrollo según el área y la edad, es decir, en las conductas observables que un niño/a promedio debería emitir en una determinada edad (Luque de Dios et al., 2023). Y aunque esta forma de medir el desarrollo presenta limitaciones, nos proporciona indicadores fiables de referencia a partir de los cuales tomar decisiones.

El avance en la creación y validación de estas escalas se ha producido de forma lenta y heterogénea (Silva et al., 2018). Cuentan con escasos estudios independientes, más allá de los realizados por los equipos que las han creado, aplicándose a esta disciplina adaptaciones de la métrica común. El progreso estadístico que ha supuesto la Teoría de Respuesta al Ítem (TRI) frente a la Teoría Clásica de los Test (TCT) (Muñiz, 2010) se ha aplicado escasamente en este ámbito (Luque de Dios et al., 2023). Estas escalas tienen asociados problemas métricos en su construcción propios de la naturaleza y características de su muestra (Sánchez, 1992) como la canalización del desarrollo, múltiples factores externos que pueden influir en el menor, cambios evolutivos entre culturas-subculturas-etnias y entre generaciones, etc. Por ello, es habitual que durante su validación cuenten con muestras insuficientes. Se combinan segmentos de edades amplias que no consideran las características diferenciadoras de cada etapa, estableciendo puntos de corte por edad, en muchas ocasiones, arbitrarios. Esto se da con mayor frecuencia en muestras menores de un año.

En la práctica clínica se valoran niños/as con diferentes alteraciones con instrumentos que no cuentan con las correspondientes adaptaciones físicas, temporales y métricas debido a la inexistencia de éstas y a la escasez de escalas dirigidas a dichas poblaciones (Sobieski et al. 2022). Otra de las debilidades de estas escalas es su antigüedad (García et al., 2009). En los últimos años se ha avanzado en la calidad de vida de los niños/as con déficits (Diez, 2004) y la mayoría de las escalas no reflejan estos progresos. Se deben, por tanto, mejorar las propiedades psicométricas de las escalas, aumentando la muestra en número y diversidad, principalmente en las primeras edades. Además, se debería optimizar también su eficiencia, aplicando para ello los avances estadísticos y tecnológicos disponibles.

Con el fin de mejorar las limitaciones señaladas, el objetivo general de esta investigación es construir y validar una escala de desarrollo de 0 a 84 meses llamada Dauco Test. Para ello, se parte de las hipótesis: si se extraen, analizan y seleccionan los mejores ítems de las principales escalas del desarrollo, Dauco Test tendría que presentar buenas propiedades métricas tanto a nivel de ítem, en primer lugar, como a nivel de subescalas, en segundo lugar, ajustando al modelo teórico. Dauco Test debe permitir discriminar de forma rápida y eficiente entre niños/as con desarrollo normal o con alteraciones o en riesgo de padecerlos, proporcionando un programa de intervención ajustado a los resultados.

Método

Participantes

En la validación participaron 638 niños ($M = 40.3$, $SD = 21.8$) reclutados en escuelas infantiles, colegios y centros de salud de Andalucía (España) (Tabla 1). Los criterios de inclusión han sido niños/as con desarrollo normotípico de las correspondientes franjas de edad y, de exclusión, niños/as con diagnóstico de alteraciones de desarrollo, dificultad o discapacidad con evaluaciones previas o en curso. El reclutamiento se realizó a través del Coordinador de la Estrategia de Atención Temprana de la Junta de Andalucía. Se contó con la colaboración de La Asociación de Pediatras de Atención Primaria de Andalucía y con los Equipos de Orientación Educativa de la Delegación Territorial de Educación de Córdoba.

Las familias fueron informadas del objetivo del estudio y del tratamiento de los datos, firmando un consentimiento. El estudio fue aprobado por el Comité de Ética de la Consejería de Salud de la Junta de Andalucía.

Tabla 1. Estadísticos descriptivos de la muestra ($N = 638$)

Variable	Indicador	N (%)	M	SD
Edad	1 año	70 (11.0)	5.1	3.8
	2 años	92 (14.4)	18.4	4.2
	3 años	123 (19.3)	30.1	3.7
	4 años	86 (13.5)	42.3	3.7
	5 años	109 (17.1)	55.6	3.1
	6 años	115 (18.0)	66.2	3.5
	7 años	43 (6.7)	77.6	3.6
Sexo	Niñas	359 (56.3)		
	Niños	279 (43.7)		
Antecedentes familiares	Sí	52 (8.2)		
	No	561 (87.9)		
	No informado	25 (3.9)		
Estatus socioeconómico	Baja	59 (9.2)		
	Media	318 (49.8)		
	Alta	94 (14.7)		
	No informado	167 (26.2)		

Nota. M = Media; SD = Desviación típica.

Instrumento

DAUCO Test es una escala multidimensional de *screening* computarizada para evaluar el desarrollo de menores de 84 meses. Se accede a través de una aplicación para móviles o tablets con sistema operativo IOS o Android. Cada profesional dispone de un área protegida con sus credenciales. Los profesionales aplican la prueba en forma de observación y/o entrevista con la familia. Se administra en lugares tranquilos, evitando distracciones que puedan influenciar las respuestas a los ítems. Para la valoración de cada niño/a se introducen sus datos personales, sociodemográficos, y sus antecedentes médicos y familiares. La escala evalúa hitos del desarrollo asociados a diferentes segmentos de edad. Consta de seis subescalas: cognitiva, lenguaje, social, adaptativa, y motora fina y gruesa.

Procedimiento

En primer lugar, se valoraron mediante juicio de expertos todos los ítems por hitos de edad de las escalas de desarrollo más utilizadas: Escala de Desarrollo Merrill-Palmer Revisada (Merrill Palmer R; Roid, 2004), Inventario de Desarrollo Battelle Segunda Edición (Battelle-2; Newborg, 2005), Currículo Carolina: evaluación y ejercicios para bebés y niños pequeños con necesidades especiales (Currículo Carolina; Johnson-Martin et al., 1994), Inventario de Desarrollo en Atención Temprana (IDAT; Pérez y Lorenzo, 2002), Escala Observacional del Desarrollo (EOD; Secadas, 2009) y Escala del desarrollo infantil Bayley III (Bayley-III; Bayley, 2006). El equipo de profesionales expertos estaba formado por psicólogos, pediatras, logopedas y fisioterapeutas infantiles. Este juicio de expertos se realizó en tres fases. En la primera fase, se hizo un inventario con todos los ítems y se eliminaron las coincidencias, que-

dando un total de 4551 ítems. En la segunda fase, se realizaron dos juicios de expertos, en dos momentos temporales diferentes. En el primer juicio de expertos se valoraron todos los ítems por seis profesionales con experiencia en diferentes áreas de conocimiento. Se revisaron los ítems teniendo en cuenta los siguientes criterios: relevancia de los ítems (pertinencia al constructo e importancia teórica), claridad y que fueran comprensibles (lenguaje claro y preciso) y variedad en la cobertura de ítems (que cubrieran todas las dimensiones del constructo). Se aceptaron para la escala todos aquellos ítems que los evaluadores habían puntuado 3 o más en el ítem (de una escala de 5). Una vez realizado el análisis por los expertos se redujeron a 2287. Por áreas: lenguaje: 444, cognitiva: 570, adaptativa: 320; social: 382; motricidad fina: 307 y motricidad gruesa: 264. Entre el primer y el segundo juicio de expertos se le pasaron los ítems seleccionados a una pequeña muestra de niños/as. En el segundo juicio de expertos, se volvieron a revisar todos los ítems y, una vez revisados, se reevaluaron hasta en dos ocasiones más, quedando un total de 1786 ítems.

En la tercera fase, se hizo otro juicio de expertos con otros seis profesionales independientes diferentes a los de la segunda fase. Se ajustaron, modificaron o redefinieron ítems muy parecidos, haciendo una valoración sustantiva de cada uno y de éstos en su conjunto. Al finalizar esta fase se seleccionaron 1767 ítems.

Los ítems seleccionados se adaptaron a una misma unidad de medida creando una escala de tipo dicotómico ("Sí" / "No"), contemplando las opciones "No se puede evaluar" y "Observaciones".

Por último, se realizó una validación estadística. La escala se administró en centros de salud y escuelas infantiles por profesionales psicólogos, logopedas, fisioterapeutas y pediatras en forma de observación y/o en entrevista con padres y/o madres. El tiempo de aplicación de toda la escala, para un intervalo concreto de edad, fue de 30-40 minutos aproximadamente, dado que el número de ítems antes del análisis estadístico aún era elevado.

Por las características propias del desarrollo, los menores de un año se valoraron mes a mes. Entre los 12-24 meses, se agruparon en franjas de edad de tres meses. Entre los 24-36 meses se agruparon en dos franjas de seis meses. De los cuatro años en adelante se valoraron año a año. Se evaluaron los siguientes ítems por área (número): lenguaje (285), cognitiva (485), adaptativa (243), social (290), motricidad gruesa (213) y fina (251), todos ellos distribuidos en las franjas de edad indicadas. De estos ítems se seleccionaron aquellos que poseían mejor calidad métrica y ajuste al modelo. Las subescalas presentan por cada segmento de edad un total de entre 5 y 10 ítems.

Análisis estadísticos

En todas las franjas de edad se realizaron análisis descriptivos de porcentajes y frecuencias con el programa Excel. En aquellas franjas de edad donde el tamaño de la muestra lo permitía, se realizaron estimaciones de fiabilidad, Chi Cuadrado de Homogeneidad, Análisis Factoriales Confirmatorios (AFC) y análisis de la Teoría de Respuesta al Ítem (TRI) a nivel de ítems y subescala. Los análisis de fiabilidad se calcularon mediante el coeficiente Alfa de Cronbach (Zumbo & Rupp, 2004) y Omega de Mc'Donald (Ventura-León & Caycho-Rodríguez, 2017) que trabaja con cargas factoriales. Se realizaron

Tabla 2. Índices de las subescalas para la franja de edad de 12-15 meses

Áreas N (29)	Media	Desviación típica	Alfa Cronbach	Omega Mc' Donald	CMIN/DF	CFI	RMSEA
Cognitiva	0.847	0.507	0.895	0.897	1.108	0.992	0.062
Lenguaje	0.698	0.382	0.824	0.832	0.897	1	0.00
Adaptativa	0.768	0.244	0.703	0.727	1.339	0.955	*0.110
Social	0.803	0.340	0.715	0.717	0.909	1	0.00
M. Gruesa	0.819	0.261	0.735	0.786	1.048	0.989	0.041
M. Fina	0.875	0.561	0.906	0.909	0.821	1	0.00

Nota. M = Motricidad.

*Valor fuera de parámetro. RMSEA superior a 0.08.

análisis Chi Cuadrado de Homogeneidad de la muestra (Quevedo, 2011) para las variables sexo y estatus socioeconómico (bajo- medio- alto). Estos cálculos se realizaron con el programa SPSS y Jamovi 2.2.5. Se analizó el ajuste de los datos al modelo teórico mediante AFC con el programa SPSS extensión AMOS (Bollen, 1989). Los índices de referencia fueron: índice de ajuste absoluto RMSEA, razón de Chi Cuadrado sobre los grados de libertad CMIN/DF y el índice de ajuste comparativo CFI. Se aplicó la TRI modelo de Rasch utilizando el método máxima verosimilitud (ML) (Reise et al., 2005) con el programa IRTPRO 4.2. Se asumieron los supuestos de unidimensionalidad e independencia local tras el AFC.

Resultados

Los resultados de la validación se presentan a continuación resumidos por franjas de edad. Los resultados detallados para cada uno de los ítems y las subescalas se encuentran en el repositorio abierto OSF (<https://osf.io/dashboard>).

La franja de edad 12-15 meses (Tabla 2) presenta alta fiabilidad y buen ajuste de los datos al modelo teórico. Se destaca en este segmento la importancia del área del lenguaje, junto con el séptimo, octavo y noveno mes. El intervalo de edad 15-18 meses presenta una validación óptima, mejorando la fiabilidad al aplicar el Omega de Mc' Donald (Tabla 3).

Tabla 3. Índices de las subescalas para la franja de edad de 15-18 meses

Áreas N (15)	Media	Desviación típica	Alfa Cronbach	Omega Mc' Donald
Cognitiva	0.792	0.354	0.932	0.941
Lenguaje	0.841	0.346	*0.962	0.967
Adaptativa	0.875	0.212	*0.556	0.782
Social	0.880	0.253	0.799	0.890
M. Gruesa	0.800	0.327	*0.670	0.827
M. Fina	0.792	0.375	0.820	0.865

Nota. M = Motricidad.

*Valores fuera de parámetro. Alfa de Cronbach inferiores a 0.7 y superiores a 0.9

La subescala 18-21 meses presenta buen ajuste de los datos y fiabilidad. La subescala social tiene baja potencia para discriminar niños/as con dificultades sociales (Tabla 4). Sin embargo, el área social es más representativa en el tercer mes, y los rangos: 18-21 y 59-71 meses. El segmento de edad 21-24 meses (Tabla 5) tiene buena validez, mostrando baja fiabilidad en el área motora. El área cognitiva y motora presentan baja discriminación entre menores con dificultades de razo-

namiento. Sin embargo, y aunque el área cognitiva destaca en todas las edades, es especialmente llamativo en este segmento y en los 36-48 meses.

Tabla 4. Índices de las subescalas para la franja de edad de 18-21

Áreas N (24)	Media	Desviación típica	Alfa Cronbach	Omega Mc' Donald
Cognitiva	0.800	0.445	0.876	0.890
Lenguaje	0.850	0.259	0.848	0.881
Adaptativa	0.847	0.329	0.921	0.934
Social	0.918	0.267	*0.980	0.988
M. Gruesa	0.857	0.228	*0.508	*0.651
M. Fina	0.821	0.356	0.904	0.920

Nota. M = Motricidad.

*Valores fuera de parámetro. Alfa de Cronbach inferior a 0.7 y superior a 0.9. Omega Mc' Donald inferior a 0.7

Tabla 5. Índices de las subescalas para la franja de edad de 21-24 meses

Áreas N (24)	Media	Desviación típica	Alfa Cronbach	Omega Mc' Donald
Cognitiva	0.927	0.251	*0.972	0.984
Lenguaje	0.803	0.359	0.766	0.782
Adaptativa	0.844	0.239	0.722	0.768
Social	0.802	0.299	0.808	0.821
M. Gruesa	0.909	0.319	*0.636	*0.676
M. Fina	0.789	0.335	0.823	0.856

Nota. M = Motricidad.

*Valores fuera de parámetro. Alfa de Cronbach inferior a 0.7 y superior a 0.9. Omega Mc' Donald inferior a 0.7

La fiabilidad y el ajuste del modelo teórico a los datos en la franja de edad 24-30 meses es bueno, siendo más débil la fiabilidad en el área social y el ajuste en el área de motricidad fina (Tabla 6). Los ítems del área cognitiva y motricidad fina muestran heterogeneidad de la muestra en la variable estatus socioeconómico.

En cuanto a las Curva Características de los Test (CCT), esta franja de edad discrimina mejor a los menores con habilidades medias. Los menores con rasgos medios-altos serán mejor discriminados en las áreas cognitiva y de lenguaje, y los menores con rasgos medios-bajos en las áreas social y de motricidad. Las gráficas analizadas se pueden encontrar en el repositorio OSF. La franja 30 a 36 meses presenta una fiabilidad aceptable. Las áreas cognitiva y adaptativa son débiles para discriminar a los menores con dificultades. El área motora no se ajusta a los parámetros evaluados (Tabla 7).

Tabla 6. Índices de las subescalas para la franja de edad de 24-30 meses

Áreas N (81)	Media	Desviación típica	Alfa Cronbach	Omega Mc' Donald	CMIN/DF	CFI	RMSEA
Cognitiva	0.795	0.323	0.757	0.763	1.487	0.897	0.078
Lenguaje	0.748	0.325	0.781	0.804	1.302	0.964	0.061
Adaptativa	0.744	0.281	0.758	0.778	1.455	0.941	0.075
Social	0.782	0.238	*0.603	*0.633	1.084	0.957	0.032
M. Gruesa	0.740	0.323	0.711	0.737	*0.633	1	0.000
M. Fina	0.782	0.345	0.789	0.794	2.185	*0.884	*0.122

Nota. M = Motricidad.

*Valores fuera de parámetro. Alfa de Cronbach inferior a 0.7. Omega Mc' Donald inferior a 0.7. CMIN/CF inferior a 1. CFI en valor límite entre 0.8-0.90. RMSEA superior a 0.08

Tabla 7. Índices de las subescalas para la franja de edad de 30-36 meses

Áreas N (42)	Media	Desviación típica	Alfa Cronbach	Omega Mc' Donald	CMIN/DF	CFI	RMSEA
Cognitiva	0.903	0.250	0.719	0.749	*0.093	1	0.000
Lenguaje	0.861	0.257	*0.675	0.737	*0.739	0.964	0.061
Adaptativa	0.929	0.182	0.743	0.785	e	e	e
Social	0.897	0.209	*0.602	0.801	e	e	e
M. Gruesa	0.923	0.295	*0.666	*0.693	*0.069	1	0.000
M. Fina	0.896	0.322	*0.687	0.715	*0.814	1	0.000

Nota. M = Motricidad.

*Valores fuera de parámetro. Alfa de Cronbach inferiores a 0.7. Omega Mc' Donald inferior a 0.7. CMIN/CF inferior a 1. "e" error de cálculo por datos extremos.

Tabla 8. Índices de las subescalas para la franja de edad de 36-48 meses

Áreas N (86)	Media	Desviación típica	Alfa Cronbach	Omega Mc' Donald	CMIN/DF	CFI	RMSEA
Cognitiva	0.745	0.339	0.799	0.811	1.136	0.979	0.040
Lenguaje	0.869	0.225	0.762	0.772	0.999	1	0.000
Adaptativa	0.841	0.236	*0.673	0.721	1.781	*0.886	*0.096
Social	0.915	0.193	*0.644	0.722	*0.102	1	0.000
M. Gruesa	0.800	0.411	*0.691	0.707	*0.869	1	0.000
M. Fina	0.896	0.341	*0.699	0.720	1.029	0.997	0.018

Nota. M = Motricidad.

*Valores fuera de parámetro. Alfa de Cronbach inferiores a 0.7. CMIN/CF valores inferiores a 1. CFI en valor límite entre 0.8-0.9. RMSEA superior a 0.08.

Tabla 9. Índices de las subescalas para la franja de edad de 48-59 meses

Áreas N (109)	Media	Desviación típica	Alfa Cronbach	Omega Mc' Donald	CMIN/DF	CFI	RMSEA
Cognitiva	0.954	0.129	0.734	0.826	1.046	0.998	0.021
Lenguaje	0.760	0.285	0.719	0.735	1.300	0.987	0.053
Adaptativa	0.905	0.182	*0.610	*0.635	1.816	*0.859	*0.087
Social	0.849	0.206	*0.513	*0.638	1.474	0.977	0.066
M. Gruesa	0.956	0.114	*0.455	*0.526	e	e	e
M. Fina	0.983	0.192	*0.634	0.747	*0.760	1	0.000

Nota.

*Valores fuera de parámetro. Alfa de Cronbach inferiores a 0.7. Omega Mc' Donald inferiores a 0.7. CMIN/CF valor inferior a 1. CFI en valor límite inferior entre 0.8-0.9. RMSEA superior a 0.08. "e" error de cálculo por datos extremos.

La subescala 36-48 meses presenta buena fiabilidad (Tabla 8) y discriminación como se puede apreciar en la media y en las CCT, a excepción del área social que es algo más débil. El área cognitiva, de lenguaje y motricidad fina discriminan bien a niños/as con rasgos medio-medio bajos. La franja de edad 48-59 meses discrimina mejor a sujetos

con rasgos medio-bajos en las habilidades medidas según las CCT. La subescala presenta una fiabilidad baja y la necesidad de realizar adaptaciones en algunas áreas para que ajusten al modelo teórico propuesto (Tabla 9). Los ítems de lenguaje muestran heterogeneidad de la muestra en la variable estatus socioeconómico.

Tabla 10. Índices de las subescalas para la franja de edad de 59-71 meses

Áreas N (115)	Media	Desviación típica	Alfa Cronbach	Omega Mc' Donald	CMIN/DF	CFI	RMSEA
Cognitiva	0.960	0.201	*0.381	*0.575	1.646	0.900	0.075
Lenguaje	0.949	0.142	*0.261	*0.452	*0.418	1	0.000
Adaptativa	0.910	0.162	*0.449	*0.518	1.834	*0.742	*0.086
Social	0.832	0.221	*0.556	*0.565	1.389	*0.879	0.058
M. Gruesa	0.995	0.198	*0.582	*0.600	e	e	e
M. Fina	e	e	e	e	e	e	e

Nota.

*Valores fuera de parámetro. Alfa de Cronbach inferiores a 0.7. Omega Mc' Donald inferiores a 0.7. CMIN/CF valor inferior a 1. CFI valor inferior a 0.8 y valor límite entre 0.8-0.9. RMSEA superior a 0.08. "e" error de cálculo por datos extremos.

Tabla 11. Índices de las subescalas para la franja de edad de 71-84 meses

Áreas N (43)	Media	Desviación típica	Alfa Cronbach	Omega Mc' Donald	CMIN/DF	CFI	RMSEA
Cognitiva	0.760	0.271	*0.615	*0.637	*0.690	1	0.000
Lenguaje	0.899	0.206	0.739	0.746	1.145	0.983	0.059
Adaptativa	0.894	0.189	0.708	0.775	e	e	e
Social	0.886	0.231	*0.572	*0.601	*0.715	1	0.000

*Nota. Valores fuera de parámetro. Alfa de Cronbach inferiores a 0.7. Omega Mc' Donald inferiores a 0.7. CMIN/CF valores inferiores a 1. "e" error de cálculo por datos extremos.

La fiabilidad de la escala 59-71 meses es baja (Tabla 10). Discrimina de forma adecuada entre menores con rasgos bajos. En la franja de edad 71-84 meses la motricidad ya está adquirida (Pica, 2011), por lo que no se presentan análisis. La subescala tiene adecuada fiabilidad y ajuste al modelo (Tabla 11).

Respecto a la dificultad de la escala en su conjunto, los ítems seleccionados en el análisis TRI están entre 0 y -2 (ver repositorio). Hay algunas excepciones en las que se pueden exceder hasta -3.

Discusión

El proceso de evaluación en las etapas de infantil utilizando escalas de *screening* nos permite una detección temprana de retrasos o alteraciones del desarrollo (Schonhaut et al., 2023). El objetivo de este estudio fue construir y validar la escala DAUCO, partiendo de sus escalas predecesoras, intentando superar algunas de las limitaciones encontradas en ellas. Los resultados muestran que las subescalas de DAUCO presentan buena fiabilidad y validez, aunque se han realizado ajustes de redacción y métricos.

La primera hipótesis sostenía que, si los ítems analizados y seleccionados para la construcción de este instrumento habían sido extraídos de las escalas más representativas, tendrían que presentar buenas propiedades. Sin embargo, durante la validación se detectaron algunos errores. En primer lugar, respecto al análisis sustantivo de los ítems, para que éstos discriminen bien a los niños/as con alteraciones, su redacción tiene que ser clara y directa, especificando lo que se requiere del evaluado/a. Los ítems tienen que proponer acciones y conceptos contextualizados y cercanos al niño/a. Las herramientas que se usan para probar estos ítems, generalmente juguetes, tienen que ser de fácil acceso y conocidos. En el análisis realizado en la primera fase del estudio se detectaron ítems que no cum-

plen con estos requisitos. Esto podría deberse principalmente a dos motivos: una incorrecta formulación en el origen o, con más probabilidad, a una mala adaptación cultural. En segundo lugar, el análisis del número de aciertos por ítem. Cuando los ítems se aplican a la población general, entre el 80 y el 90 por ciento de los niños/as deberían poder superarlo. Sin embargo, se hallaron ítems que son superados por el 100 % de los menores y de ítems extremos en el parámetro b (muy fácil) tras la aplicación de la TRI. En menor medida, se localizaron otros que fueron superados por entre el 40 y el 60 por ciento de los niños/as. Esto significa que las escalas presentan una dificultad baja, aunque aceptable, para la población. Esto disminuye la fiabilidad y provoca en los agentes implicados en la evaluación expectativas desajustadas. En tercer lugar, se detectaron ítems que se sitúan en diferentes segmentos de edad o áreas del desarrollo, dependiendo de la escala analizada. Estos ítems, aunque correspondan a importantes hitos evolutivos, no son buenos discriminando y pueden generar confusión. En cuarto lugar, se encontraron ítems que valoran de forma diferente a niños/as según su estatus socioeconómico. No ocurre lo mismo con la variable sexo. Es importante destacar que los ítems mal formulados, extremos o con baja discriminación o fiabilidad están siendo utilizados en la actualidad como parte de estas escalas. En consecuencia, se considera importante realizar una revisión de las escalas analizadas. También habría que valorar si esto sucede con los ítems de las escalas no contempladas en este estudio. Por último, como consecuencia de la dificultad baja de los ítems, podemos afirmar que estas escalas discriminan mejor entre valores bajos del rasgo (valores negativos).

En cuanto a la segunda hipótesis, planteaba que, tras la extracción y el análisis de los ítems de las escalas más representativas, las subescalas resultantes tendrían que tener óptimos parámetros estadísticos, ajustando los datos al modelo teórico planteado. Esto tampoco se cumple. En el análisis compara-

tivo de las áreas del desarrollo, se observaron diferencias en la importancia que se da a unas áreas respecto a otras, en conjunto y por cada segmento de edad en particular (Luque de Dios et al., 2023). Esto se hace evidente por la cantidad y calidad de ítems en cada una de las subescalas analizadas. A diferencia de las áreas de lenguaje, social y cognitiva que están más vinculadas a periodos de edad concretos, el área adaptativa aumenta el número de ítems por edad de forma ascendente y la cantidad de ítems es menor. En las áreas de motricidad fina y gruesa el número de ítems está compensado, siendo en la motricidad fina (21-24 meses) algo superior. Sin embargo, se debe considerar que las franjas de 15-18, 18-21 y 21-24 meses presentan una fiabilidad muy baja. Esto confirma errores en su construcción, principalmente en el segundo año de vida. Las escalas de motricidad gruesa son las menos fiables, con menos cantidad de ítems y un gran número de ellos mal formulados. Esto podría deberse a que hasta el momento los fisioterapeutas pediátricos no han participado del proceso de elaboración de escalas, necesitando éstas una revisión exhaustiva por estos profesionales. Se deben valorar con cautela los resultados que arrojan estas escalas.

Las áreas que peor adaptan los datos al modelo teórico son la cognitiva y las de motricidad gruesa y fina. Si consideramos las diferentes edades, las subescalas de los niños/as de 4 años son las que presentan mejores índices estadísticos. Por el contrario, las subescalas destinadas a niños/as de 5, 6 y 7 años muestran los peores indicadores.

Aunque este estudio hace contribuciones importantes en el campo de las escalas de desarrollo infantil, no está exento de limitaciones. En primer lugar, el uso de una muestra incidental y reducida, sobre todo en los niños/as de menor edad. La investigación en intervención infantil temprana presenta de forma generalizada el problema de acceso a la muestra que restringe los estudios métricos (Luque de Dios et al., 2023). Esto se acentúa en las edades más tempranas y en niños/as con diversidad funcional (Visser et al., 2012; Silva et al., 2018; Luque de Dios et al., 2023). En nuestro estudio, la franja de edad de 0 a 12 meses cuenta con 70 sujetos por lo que se han realizado análisis descriptivos y sustantivos (ver repositorio). En cuanto al segundo año de vida, no se aplica TRI porque, aunque en conjunto tengamos 92 sujetos, la muestra no es suficiente para cada ítem por franja de edad y los análisis no serían pertinentes. En segundo lugar, esta escala presenta limitaciones métricas provenientes de los errores de las escalas utilizadas para la selección de ítems. Aunque en su mayoría han sido depurados, nos encontramos con subescalas y segmentos de edad que deben volver a ser revisados, reformulados y/o reevaluados por especialistas, como el área motora y adaptativa, y el intervalo de edad de cinco y seis años. Se propone completar el estudio con el análisis diferencial de los ítems (DIF) con menores con diversas alteraciones y de diferentes regiones, y comparar estos análisis con las pruebas de homogeneidad con las variables sexo y condición sociodemográfica ya realizados. También realizar análisis de validez convergente y discriminante con las escalas más utilizadas en este ámbito como Merrill Palmer- R, Battelle II y Bayley III.

Las investigaciones futuras de las escalas de desarrollo deberían centrarse en mejorar el tamaño y la heterogeneidad de la muestra, principalmente en las primeras franjas

de edad. Además, se debería trabajar en acortar las pruebas, mejorar el formato y aumentar su calidad métrica. Para ello se deben aplicar avances estadísticos y tecnológicos, y crear *softwares* adaptados. Estos avances se están dando de forma limitada. Además, es fundamental instruir de forma continuada a los examinadores.

En conclusión, las escalas del desarrollo analizadas incluyen ítems y subescalas que no son suficientemente fiables y no ajustan al modelo teórico. Además, la valoración de las distintas áreas y de los rangos de edad no es equilibrada, viéndose mermada la posibilidad de valorar ciertos intervalos de desarrollo en la consecución de sus hitos. Así, se destaca la idea de que, aunque contar con una escala de desarrollo estandarizada ayuda a evaluar a los menores, es importante seguir considerando diversos instrumentos. En este contexto, la escala DAUCO supone un avance respecto a las pruebas de screening anteriores. Es una escala actualizada que mejora sus características psicométricas y su aplicabilidad mediante un formato digital. Se han seleccionado ítems que discriminan bien en valores medio-bajo de rasgo, valorando todos los niveles de dificultad en los diferentes intervalos de edad, establecidos estos intervalos según la lógica del desarrollo y no de forma arbitraria, por lo que su capacidad para discriminar niños/as es buena. Además, es una escala eficiente ya que el tiempo total de aplicación de todas las escalas, para un intervalo concreto de edad, es de 20-30 minutos. Es flexible y práctica, pues la aplicación de la prueba permite evaluar las distintas áreas del desarrollo de forma independiente. Durante la valoración se guardan de forma automática las respuestas y se muestra el progreso de la prueba indicando las subescalas e ítems completados, pudiendo acceder de forma directa a cada uno de ellos y permitiendo hacer un seguimiento acorde con las necesidades. También, cuando introducimos la fecha de nacimiento del menor, sale automáticamente la edad cronológica en las escalas que queremos pasar, pudiendo modificar de edad cronológica a edad evolutiva si lo consideramos oportuno. Al finalizar la valoración, la aplicación facilita un informe del niño/a indicando el percentil de desarrollo en cada una de las áreas, y un programa de intervención con propuestas de actividades adaptadas a los resultados que pueden ser modificadas por los profesionales. Así, el sistema informatizado y la propuesta de actividades proporciona el diseño de planes de intervención individualizados, la evaluación y el seguimiento a lo largo del tiempo.

Financiación

Esta investigación ha sido financiada por la Convocatoria de subvenciones para la Investigación, Desarrollo e Innovación Biomédica y en Ciencias de la Salud en Andalucía (PIN-0072-2016). Consejería de Salud y Familias. Junta de Andalucía (España).

Conflicto de intereses

Los autores de este estudio declaran que no tienen ningún conflicto de intereses.

Referencias

- Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development* (3rd ed.). EE. UU: The Psychological Corporation.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- Bollen, K. A. (1989). A new incremental adjustment index for general structural equation models. *Sociological Methods & Research*, 17, 303-316. <https://doi.org/10.1177/0049124189017003004>
- Consejería de Educación, Formación y Empleo (2010). *Educational care for students with autism and other pervasive developmental disorders*. Consejería de Educación, Formación y Empleo.
- Diez, A. C. (2004). Special educational needs. Educational policies around alterity. *Cuadernos de Antropología Social*, 19, 157-171. <https://doi.org/10.34096/cas.i19.4574>
- Figueiras, A. C., Neves de Souza, I. C., Ríos, V. G. & Benguigui, Y. (2011). *Manual for monitoring child development (0-6 years) in the context of IMCI* (2ª Ed.). EE. UU: OPS.
- García-Pérez, E.M. & Magaz, A. (2009). *How to assess psychometric tests? Conceptual and Methodological Errors in the Psycho-Educational Evaluation*. Spain: Group Albor-Cohs.
- Guralnick, M. J. (2023). A framework for designing inclusive community based early childhood intervention programs. *Infants y Young Children*, 36(4), 270-284.
- Johnson-Martin, N., Attermeier, S. & Hacker, B. (1994) *Carolina Curriculum: Assessment and Exercises for Infants and Toddlers with Special Needs*. Spain: TEA Editions.
- Luque de Dios, S.M., Sánchez-Raya, A., & Moriana-Elvira, J. A. (2023). Quality of child development scales. A systematic review. *International Journal of Educational Psychology*, 12(2), 119-148. <http://doi.org/10.17583/ijep.10733>
- Muñiz, J. (2010). Test of theories of the Classical Theory and Theory of Responses to Items. *Papeles del Psicólogo*, 31(1), 57-66. <http://doi.org/2010-04075-005>
- Newborg, J. (2005). *Battelle Developmental Inventory -second edition*. EE. UU: Riverside Pub.
- Pérez, I. & Lorenzo, J. (2002). *Inventory of development in early care IDAT*. Spain: Amaru Editions.
- Pica, R. (2011). The role of motor skill development in early childhood education. *Early Childhood Education Journal*, 39(6), 421-428. <https://doi.org/10.1007/s10643-011-0462-0>
- Quevedo, F. (2011). The chi-square test. *Statistics Applied to Health Research*, 11(12). <https://doi.org/10.5867/medwave.2011.12.5266>
- Reise, S. P., Ainsworth, A. T. & Haviland, M. G. (2005). Item Response Theory: Foundations, applications, and promise in psychological research. *Current Directions in Psychological Science*, 14 (2), 95-101. <https://doi.org/10.1111/j.0963-7214.2005.00342.x>
- Roid, G., Samplers, J., Anderson, G., Erickson, J. & Post, P. (2004). *Merrill-Palmer-Revised. Scales of Development*. EE. UU: Stoelting Company.
- Sánchez Turrión, G. (1992). *Assistance to nursery and child development* (Doctoral Thesis). Pontifical University of Salamanca, Spain.
- Schonhaut, B. L., Buron, K. V., Aguilera, E. R. & Vargas, B. L. (2023). Early detection of autism spectrum disorder: Review of screening tools validated in Chile. *Andes pediatric*, 94(4), 425-435. <https://doi.org/10.32641/andespediatr.v94i4.4901>
- Secadas, F. (2009). *Observational Scale of Development EOD*. Spain: TEA Editions.
- Sices, L., Feudtner, C., McLaughlin, J., Drotar, D. & Williams, M. (2003). How do primary care physicians identify young children with developmental delays? A national survey. *Journal of Developmental Behavioral Pediatrics*, 24, 409-417. <https://doi.org/10.1097/00004703-200312000-00002>
- Silva, M., Mendonça, F., Euclides J., Mõnego, B. & Bandeira, D. (2018). Instruments for multidimensional assessment of child development: A systematic review. *Early Child Development and Care*, (), 1-15. <https://doi.org/10.1080/03004430.2018.1528243>
- Sobieski, M., Sobieska, A., Sekulowicz, M. & Bujnowska-Fedak, M. M. (2022). Tools for early screening of autism spectrum disorders in primary health care - a scoping review. *BMC Primary Care*, 23(1), 46. <https://doi.org/10.1186/s12875-022-01645-7>
- Ventura-León, J. L. & Caycho, T. (2017). The Omega coefficient: An alternative method for estimating reliability. *Latin American Journal of Social Sciences, Children and Youth*, 15(1), 625-627.
- Villagra, M. R., Jara, R. L. R., Cosp, S., Sardi, J. M. G., Encina, A., Jara, L. R. & Barreto, N. V. (2023). Implementation of the early care service in neurodevelopment in Coronel Oviedo, Paraguay: Case study. *Scientific Journal Studies and Research*, 12(1), 107-124.
- Visser, L., Ruiter, S., van der Meulen, B., Ruijsenaars, W. & Timmerman, M. (2012). A review of standardized developmental assessment instruments for young children and their applicability for children with special needs. *Journal of Cognitive Education and Psychology*, 11(2), 102-127. <https://doi.org/10.1891/1945-8959.11.2.102>
- www.osf.io
- Zumbo, B.D. & Rupp, A.A. (2004). Responsible modelling of measurement data for appropriate inferences: Important advances in reliability and validity theory. En D. Kaplan (Ed.): *The SAGE Handbook of Quantitative Methodology for the Social Sciences* (pp. 73-92). EE. UU: Sage Press.