

Construcción y estandarización de un instrumento de evaluación de aprendizajes esperados en educación parvularia basado en tablet

Development and Standardization of a Tablet-based Assessment Tool of Early Childhood Education Learning Goals

¹Doyna Illmer, ²Ricardo Rosas, ²Soledad Véliz, ²María Paz Ramírez,
²Andrés Aparicio, ²Catalina Benavente y ²Carolina Thibaut

¹Facultad de Educación, Pontificia Universidad Católica de Chile

²Centro de Desarrollo de Tecnologías de Inclusión (CEDETi UC), Chile

Resumen

En Chile escasean los instrumentos de evaluación de aprendizajes para párvulos que cuenten con criterios de confiabilidad y validez y que evalúen aprendizajes esenciales definidos de ese modo por el Ministerio de Educación de Chile. El objetivo de esta investigación es validar una prueba de evaluación de aprendizajes a través del juego llamada *dip*, sostenida en *tablet*. La prueba evalúa conceptos básicos relacionados con lenguaje verbal y razonamiento lógico-matemático en niños de entre tres y seis años. La prueba *dip* fue estandarizada con 360 niños entre tres y seis años de instituciones de diversa dependencia administrativa que ofrecen educación parvularia en la Región Metropolitana de Chile. La prueba conlleva evidencia de buena confiabilidad interna (0,7) y confiabilidad consistente por eje de aprendizajes evaluados. Asimismo, evidencia una alta correlación con una prueba de evaluación de competencias lectoras iniciales (0,7) y una correlación moderada con una prueba de aprendizajes en formato tradicional (0,5).

Palabras clave: educación parvularia, evaluación, aprendizajes esperados, construcción de instrumento

Correspondencia a:

Doyna Illmer Craciun

Facultad de Educación, Pontificia Universidad Católica de Chile

Av. Vicuña Mackenna 4860, Macul, Santiago, Chile,

Correo electrónico: dkillmer@uc.cl

Investigación realizada en el contexto del proyecto FONDEF D0911238 "Construcción y estandarización de instrumentos de evaluación psicométrica para el desarrollo de una cultura de evaluación ética y rigurosa". El desarrollo teórico se ha estado apoyado en los desarrollos financiados por el Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt) en el marco del proyecto 1110975.

© 2013 PEL, <http://www.pensamientoeducativo.org> - <http://www.pel.cl>

ISSN: 0719-0409 DDI: 203.262, Santiago, Chile
doi:10.7764/PEL.50.2.2013.9

Abstract

In Chile there are not enough reliable, standardized instruments to assess young children's knowledge of basic concepts defined essential by the Ministry of Education of Chile. The goal of this research is to validate a play-based assessment tool of basic concepts for young children (3 to 6 years old) called *dip*, supported in tablet. The instrument assesses basic concepts related to verbal language and mathematical reasoning. It was standardized with 360 children between 3 and 6 years old attending institutions offering early childhood education in the Metropolitan Region of Chile. The results show good internal reliability (0.7) and consistent reliability for different groups of assessed knowledge. There is evidence of strong correlation with a computer-based assessment instrument of early literacy competencies (0.7) and a moderate correlation with a test of basic concepts in a traditional format (0.5).

Keywords: preschool education, assessment, learning goals, instrument development

En Chile, la evaluación de aprendizajes esperados en educación parvularia está experimentando un creciente interés en respuesta a la necesidad de orientar la toma de decisiones pedagógicas basadas en evidencia para mejorar la calidad de esta oferta educativa. La prueba *dip*, instrumento de evaluación formativa, entrega información a los educadores sobre el nivel de logro de los niños en relación con los aprendizajes esperados en los núcleos de aprendizaje de razonamiento lógico-matemático y cuantificación y de lenguaje verbal. Este instrumento permite al educador en aula tomar decisiones pedagógicas basadas en evidencia para ajustar la propuesta curricular, de modo de responder a la diversidad de aprendizaje (RDA), identificar de manera oportuna los niños en riesgo y derivar cuando corresponda.

En Chile, los educadores de párvulos cuentan con escasas herramientas que permitan respaldar las decisiones pedagógicas tomadas en el aula. Esto es importante, ya que esta es una de las áreas de bajo desempeño de los educadores, como lo demuestran los resultados del Sistema de Evaluación del Desempeño Profesional Docente que comenzó a implementarse en el año 2003 y alcanzó una cobertura nacional el 2005. Este sistema evalúa principalmente la calidad de la enseñanza, según criterios establecidos en el Marco para la Buena Enseñanza (CPEIP y Mineduc, 2008). Según datos recientes, en esta evaluación las educadoras de párvulo reciben los puntajes más bajos en comparación con los profesores de educación básica y media (Manzi, González y Sun, 2011). Dentro de las áreas evaluadas, se califican los siguientes aspectos como básicos o insatisfactorios: (a) contribución de las actividades a los objetivos de aprendizaje de la clase (78%); (b) atribución de los resultados de aprendizaje (77%); (c) capacidad de reformular estrategias didácticas (72%); y (d) análisis de las decisiones pedagógicas (68%).

Estas cuatro áreas evaluadas en nivel básico o insatisfactorio son fundamentales para lograr los aprendizajes esperados propuestos para párvulos. El test *dip* arroja informes de resultados de fácil interpretación, divididos en dimensiones: bajo, en proceso, esperado o sobre lo esperado, identificando claramente los aspectos que se deben reforzar o fortalecer, así como las áreas donde los niños ya han alcanzado los logros de aprendizaje esperados y deben ser expuestos a nuevos aprendizajes, desafiantes pero alcanzables. El informe de resultados de *dip* incluye el detalle sobre los conceptos que el niño respondió incorrectamente, lo que permite al educador ser preciso en las modificaciones que haga a la cobertura curricular para asegurar el logro de los aprendizajes esperados según las Bases Curriculares de la Educación Parvularia (BCEP). De este modo, el test *dip* entrega información concreta y detallada en un formato de fácil uso e intuitivo que debiera permitir al profesional a cargo del aprendizaje de los niños tomar decisiones pedagógicas basada en evidencia para reforzar, fortalecer, desafiar, intervenir o derivar oportunamente.

¿Qué es el test *dip*?

El instrumento *dip* está diseñado para evaluar, de manera formativa y complementaria, los aprendizajes esperados de niños y niñas de entre tres y seis años en educación parvularia. La prueba *dip* evalúa dos núcleos de aprendizaje: relaciones lógico-matemáticas, y cuantificación, y lenguaje verbal. Para operacionalizar los aprendizajes esperados descritos en las BCEP, el test *dip* se basa en los logros de aprendizaje descritos en los Mapas de Progreso de la Educación Parvularia (MPÉP), una sistematización operacional de las BCEP.

Bases teóricas y referentes para la construcción del test *dip*

Aprendizaje y desarrollo infantil

El desarrollo y el aprendizaje se encuentran estrechamente relacionados. Esto tiene importantes implicancias para la enseñanza, ya que identificar el nivel de desarrollo en el que se encuentra un niño, ya sea en relación con sí mismo o con otros, permite apoyarlo en su aprendizaje (Bransford, Derry, Berliner, Hammerness, & Beckett, 2005). Para poder llevar a cabo esta tarea los educadores necesitan comprender cómo aprenden los niños (Bransford et al., 2005) y conocer las progresiones generales del desarrollo y las diferencias individuales en el mismo. Saber cómo la instrucción puede apoyar el desarrollo es una de las características centrales de una enseñanza efectiva (Bransford et al., 2005).

Educación parvularia en Chile

En los primeros seis años de vida, los niños son especialmente sensibles al desarrollo de destrezas básicas, particularmente aquellas vinculadas con un desempeño óptimo en la educación formal (Leseman, 2002). Existe consenso en cuanto a la importancia de la educación durante la infancia temprana. Cuando asisten a una educación parvularia de calidad, los niños desarrollan conocimientos, habilidades y disposiciones necesarias para cursar kínder y primero básico con éxito. Esto aumenta las probabilidades de seguir patrones de logro académico positivos, y se reducen las posibilidades de fracaso académico y deserción escolar. Invertir en educación parvularia resulta eficiente y eficaz para el desarrollo económico y de fuerza laboral. Cuanto antes se haga la inversión, mayor será el retorno (Heckman & Masterov, 2004).

Los niños no tienen una preparación innata para enfrentar con éxito las demandas de la educación formal. En este contexto, la iniciación escolar¹ puede entenderse como:

Procesos que cambian a medida que el niño adquiere habilidades académicas importantes en sus primeros años de escolaridad. Esto incluye no solo el nivel de habilidad al entrar a la escuela (...) sino también la curva de crecimiento de esas habilidades en el tiempo, como función tanto de las características del niño y de su familia como de las características del aula y del contexto escolar. (Vernon-Feagans, Odom, Pancsofar, & Kainz, 2007, p. 63).

Esta definición de iniciación escolar enfatiza la necesidad de incorporar tanto el contexto como las relaciones que influyen en el desarrollo del niño (Knoche, Sheridan, Edwards, & Osborn, 2010), integrando las numerosas variables que influyen en las habilidades con las que un niño ingresa a la educación formal. De este modo, el ajuste entre el niño, la familia, la escuela y la comunidad resulta en procesos de desarrollo que cambian en el tiempo (Vernon-Feagans et al., 2007).

Si bien se tiende a dar más importancia a las habilidades académicas en el desarrollo infantil temprano para la predicción del desempeño académico posterior (Kolb & Fantie, 2009), las últimas investigaciones en esta materia han asignado una gran importancia también al desarrollo de las funciones ejecutivas (Welsh, Nix, Blair, Bierman, & Nelson, 2010). Estas funciones muestran un desarrollo sustancial entre los tres y cinco años (Welsh et al., 2010) y, conceptualmente, son las funciones necesarias para que los niños organicen su pensamiento y conducta con “flexibilidad creciente (...) así como para involucrarse en una conducta autorregulada y gobernada por reglas” (Welsh et al., 2010, p. 2). Algunas de estas funciones, como memoria de trabajo y autocontrol, serían facilitadoras de la iniciación escolar y del aprendizaje temprano (Welsh et al., 2010).

Conceptos relacionales básicos

Los conceptos relacionales básicos son herramientas esenciales de pensamiento utilizadas en todas las culturas (Siegler, 1998) y son esenciales para la iniciación a la escuela.

¹ Esta descripción responde al concepto *school readiness*, utilizado internacionalmente.

El instrumento *dip* evalúa conceptos en las áreas de razonamiento lógico-matemático y cuantificación y de lenguaje verbal. Según Boehm (2001), conocer, comprender y saber utilizar conceptos como estos permite comprender y describir relaciones entre objetos y la ubicación y las características de personas, lugares y cosas; comprender el orden de eventos; seguir instrucciones de la educadora/profesora y en tests estandarizados; involucrarse en actividades de literacidad emergente; y participar en actividades de resolución de problemas que comprendan tareas de clasificación, comparación, secuenciación e identificación de distintos atributos.

A medida que los niños transitan hacia la educación básica, el conocimiento de estos conceptos básicos se hace cada vez más crítico para la comprensión de lo que se está comunicando en la sala de clases (Boehm, 2013), lo que justifica que se considere un elemento importante de evaluar en detalle durante la educación parvularia.

Bases Curriculares de la Educación Parvularia (BCEP)

Las BCEP fueron elaboradas por la Unidad de Curriculum y Evaluación del Ministerio de Educación de Chile. Comenzaron a implementarse en 2002 y fueron publicadas en 2005. Fueron elaboradas para actualizar, reorientar y enriquecer los contextos y oportunidades de aprendizaje que se ofrecen a niños y niñas y responden a la necesidad de integrar y articular en un mismo instrumento curricular, con criterios comunes, un solo cuerpo de aprendizajes esperados y orientaciones educativas (Mineduc, 2005).

Las bases curriculares se organizan en ámbitos de experiencias para el aprendizaje que se encuentran vinculados entre sí y organizan aprendizajes esperados para todos los niños y niñas desde los primeros meses de vida hasta la educación básica. Cada ámbito de experiencia para el aprendizaje está compuesto por núcleos de aprendizaje junto a su respectivo objetivo general. Asimismo, los aprendizajes esperados se organizan en dos ciclos de aprendizaje. El primer ciclo va de los primeros meses a los tres años y el segundo ciclo, de los tres a los seis años o al ingresar a la educación básica.

Mapas de Progreso de Educación Parvularia (MPEP)

Los MPEP fueron desarrollados por el Ministerio de Educación del Gobierno de Chile en 2004 junto con la Junta Nacional de Jardines Infantiles (Junji) y la Fundación Integra. En su elaboración también participó la Unicef. Los Mapas de Progreso pueden entenderse como la operacionalización de las Bases Curriculares. Hay diferencias entre las BCEP y los MPEP, en particular en cuanto a que los mapas se orientan al nivel o tramo en el que se encuentra el niño en relación con sus logros de aprendizaje.

Perfiles de Logro de Aprendizaje en la Educación Parvularia (PLAEP-R)

Un test que se utilizó como referencia y es anterior al *dip* es PLAEP-R (2009), desarrollado por la Fundación Integra y la Pontificia Universidad Católica de Chile. Esta prueba, de formato tradicional, fue piloteada a nivel nacional entre los años 2008 y 2009. Es una prueba estandarizada para la comunidad educativa que pretende mejorar el proceso educativo desarrollado en jardines infantiles y salas cuna a través de la medición del nivel de logro de aprendizajes de niños y niñas de entre 1 y 5 años de edad. Al ser construida de acuerdo con las Bases Curriculares de la Educación Parvularia, entrega información comparable con respecto a todos los ámbitos de aprendizaje definidos en ellas.

La muestra total (PLAEP-R) para la estandarización estuvo compuesta por 1192 niños asistentes a los jardines de Fundación Integra, Junji, centros privados y colegios municipales, particulares subvencionados y pagados de todo el país. PLAEP-R y *dip* son pruebas que, a pesar de basarse en las bases curriculares para determinar los aprendizajes a evaluar, difieren en cuanto a formato de presentación de los ítems, portabilidad, estandarización de los ítems, tiempo de aplicación, automatización del proceso de corrección y tiempo de corrección. La prueba *dip*, además, es un tipo de evaluación formativa, lo que permite a la educadora en aula tomar decisiones de manera oportuna y así realizar mejoras antes de que termine el año escolar.

Descripción del instrumento

La prueba *dip* ha sido desarrollada bajo un modelo de evaluación a través del juego. El juego no solo es una experiencia privilegiada de aprendizaje (Rosas et al., 2003), sino que representa “una ventana al nivel de desarrollo de los niños” (Kelly-Vance & Ryalls, 2005, p. 549). La evaluación basada en juegos conlleva una alta validez ecológica, y la Asociación Nacional de Psicólogos Colegiados (NASP, en su sigla en inglés) de Estados Unidos la señala como una forma apropiada de evaluar las necesidades de los niños pequeños (Kelly-Vance & Ryalls, 2005). La prueba *dip* se encuentra sostenida en tablet, dispositivo en formato digital. La evaluación fue diseñada como un juego digital con una narrativa donde el personaje principal capta la atención y estimula la motivación de los niños a lo largo de las tareas que se les solicitan. El personaje principal de *dip*, Caja, es de género neutro, lo que facilita la identificación por parte tanto de niños como de niñas. Caja transita por varias etapas con el objetivo de encontrar una mascota que ha perdido. La narrativa del juego pretende mantener el *flujo*, un estado de experiencia óptima donde la persona se encuentra tan involucrada en la actividad que desaparece la autoconsciencia y el sentido del tiempo (Shute & Kim, en prensa). Este estado contribuiría a que la evaluación no sea percibida como tal por los evaluados.

¿Qué evalúa la prueba *dip*?

La prueba *dip* evalúa una selección de aprendizajes esperados que dan cuenta de dos ámbitos de las BCEP: la *relación con el medio natural y cultural* y la *comunicación*, así como de los núcleos de aprendizaje *relaciones lógico-matemáticas y cuantificación y lenguaje verbal* asociados a cada ámbito. Hay al menos un ítem por núcleo de aprendizaje.

Además de la evaluación de aprendizajes esperados de acuerdo con los lineamientos curriculares del Mineduc, la prueba *dip* evalúa precursores de la lectura asociados a la conciencia fonológica como reconocimiento de sonido y sílabas inicial y final. Por último, el test *dip* evalúa conceptos relacionales básicos esenciales para que los niños puedan seguir instrucciones, comprender el contenido del material de instrucción, involucrarse en actividades dentro y fuera del aula y comunicarse efectivamente con otros.

Por tanto, *dip* ofrece una alternativa de evaluación formativa y complementaria a la observación natural, única en Chile, para evaluar el logro de aprendizajes esperados en educación parvularia. Estos datos permiten tomar de decisiones pedagógicas basadas en evidencia, lo cual contribuye positivamente al mejoramiento de la calidad de esta oferta educativa.

Características y estructura del instrumento

El test *dip* se juega en dos tipos de formatos, ambos en el tablet: Mundos (60 ítems) y Entremundos (32 ítems), con un total de 92 ítems. La aplicación toma aproximadamente 20-30 minutos. La corrección es automática, a excepción de dos ítems de corrección manual (iniciación a la escritura y comunicación oral).

El formato de juego de los Mundos es un fondo o escenario que se mueve continuamente (como una cinta transportadora) de modo que los ítems se exponen al niño por un corto período. El tiempo de respuesta de los Mundos es el mismo para todos los ítems, con excepción de algunos que requieren que el ítem se detenga para que el evaluado observe en detalle las imágenes y pueda dar una respuesta. El movimiento se reanuda una vez que el evaluado responde. Los ítems que no han sido respondidos en los Mundos vuelven a presentarse al final de cada escenario, de modo que el evaluado tiene otra oportunidad para dar una respuesta. Por su parte, los ítems pertenecientes a Entremundos son problemas presentados en una pantalla estática y evalúan aprendizajes en un formato donde el ítem puede ser respondido en un tiempo prácticamente ilimitado. Para ver un ejemplo de ítem, véase la Figura 1.

En cada ítem de *dip* la respuesta del sujeto evaluado es seguida por la ocurrencia de un estímulo (bajo la forma de una animación y/o sonido). Este estímulo no entrega indica al niño si su respuesta fue correcta o incorrecta, sino que ayuda a mantener la dinámica del juego.

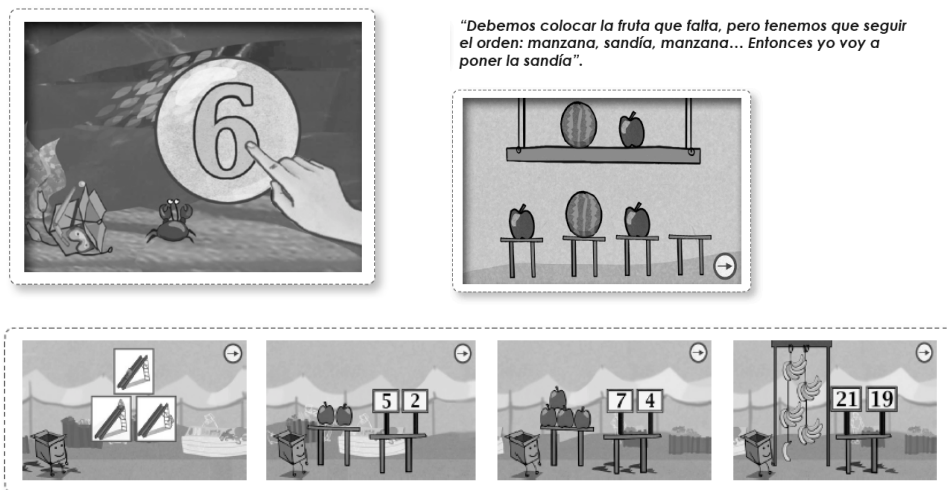


Figura 1. Ejemplos de ítems.

Método

Participantes

El universo de la muestra fueron los niños y niñas asistentes a centros de educación parvularia de la Región Metropolitana de Chile. Se realizó un muestreo por conveniencia a través del contacto con la Escuela de Educación UC e intencionado, donde se buscó un número similar de instituciones con varios tipos de dependencia administrativa: Junji, Fundación Integra y organismos dependientes del Ministerio de Educación, así como instituciones de dependencia privada. Se concluyó con una lista de 30 centros con educación de párvulos para niños y niñas de entre 3 y 6 años. Los criterios centrales para segmentar la muestra fueron la dependencia administrativa (municipal, particular subvencionada y particular pagada) y las edades de los niños (tres rangos etarios). Se cuidó de que hubiera un equilibrio en términos de género. Solo se incluyeron en el estudio los niños que entregaron consentimiento firmado por los cuidadores responsables. Se excluyó de la muestra a los niños con discapacidad sensorial y/o motora grave.

Ante la imposibilidad de acceder a información relativa al ingreso familiar para determinar el nivel socioeconómico (tradicionalmente asociado a resultados de aprendizaje y desarrollo infantil), se utilizó la dependencia administrativa del centro como aproximación al nivel socioeconómico de las familias de los niños evaluados².

La muestra para la estandarización del instrumento estuvo conformada por 360 niños y niñas pertenecientes a instituciones de diversa dependencia administrativa (véase la Tabla 1). Se perdió menos del 10% de la muestra durante las aplicaciones y análisis posteriores, de modo que los análisis se realizaron con 333 niños y niñas.

² "Chile es el país con los niveles más bajos de inclusión social en sus escuelas, donde menos del 50% de la varianza en nivel socioeconómico radica en estas instituciones (OCDE, 2012). Por lo mismo, es poco probable que estudiantes de distintos niveles socioeconómicos estén inscritos o asistan a los mismos colegios"

Tabla 1
Características de la muestra experimental

Dependencia	Sexo	3 a. – 3 a. y 11 m.	4 a. – 4 a. y 11 m.	5 a. – 5 a. y 11 m.	6 a. – 6 a. y 11 m.	Total
Municipal	Niño	18	17	16	18	69
	Niña	18	16	15	15	64
Subvencionado	Niño	10	17	15	13	55
	Niña	8	8	18	16	50
Particular	Niño	13	16	17	13	59
	Niña	12	17	16	18	63
	Total	79	91	97	93	360

Procedimiento de construcción de ítems

El instrumento *dip* fue desarrollado en 3 fases: prepiloto, piloto y aplicación experimental. Con base en los lineamientos del Ministerio de Educación de Chile acerca de los aprendizajes esperados para niños y niñas chilenos, y junto a un grupo de expertos de las áreas de educación y psicología del desarrollo, se definieron los conceptos que podían ser evaluados con la tecnología táctil en un formato de tablet, así como su pertinencia en relación con las edades evaluadas. Las primeras versiones del test contaban con 4 tramos de edad: de 3 años 0 meses a 3 años y 6 meses, de 3 años y 6 meses a 3 años y 11 meses, de 4 años a 5 años y 11 meses y de 6 a 6 años y 11 meses. La fase prepiloto se llevó a cabo en agosto de 2011. Se aplicaron los ítems a una muestra de 6 niños de modo de probar la funcionalidad de la plataforma, los tiempos de reacción a los ítems, la jugabilidad, etc. Después de esta aplicación se solucionaron errores de programación y diseño. El piloto se llevó a cabo en el mes de octubre de 2011. Se aplicó una versión del test que contaba con 28 ítems por tramo de edad. Se evaluó a 60 niños pertenecientes a instituciones con 3 tipos de dependencia administrativa: privada, subvencionada y municipal. Cada aprendizaje esperado contaba con 3 ítems al momento de la evaluación piloto, cada uno de dificultad creciente. Todos los niños evaluados, independientemente de su edad, fueron expuestos a la totalidad de ítems de modo de generar los niveles de dificultad de cada ítem. La duración total de la prueba fue de cerca de 60 minutos.

Los ítems a incluir en la fase experimental fueron revisados por un panel de expertos y, de acuerdo con los cambios realizados en las fases prepiloto y piloto, se seleccionaron aquellos ítems que discriminaban el desempeño por nivel de dificultad. Es decir, se eliminaron aquellos ítems que, en la fase de pilotaje, fueron respondidos igual de bien o mal por todos los grupos de edad. Se eliminó alrededor de un cuarto de los ítems. Dado que cada aprendizaje esperado estaba representado por 3 ítems, ningún aprendizaje esperado quedó sin ser representado, al menos, por un ítem. Para la fase experimental se utilizaron solo dos ítems por aprendizaje esperado, de diferentes niveles de dificultad. La duración total de la prueba fue aproximadamente de 20-30 minutos. Para un detalle del número de ítems en relación a los núcleos de aprendizaje evaluados, véase la Tabla 2. Para un detalle del procedimiento en relación a las fases del estudio, véase la Tabla 3.

Tabla 2
Ítems de *dip* y su relación con los ámbitos y núcleos de aprendizajes

Núcleos de aprendizaje	Ámbitos de experiencias para el aprendizaje											TOTAL		
	Formación personal y social			Comunicación				Relación con el medio natural y cultural						
	Autonomía	Identidad	Convivencia	Lenguaje verbal			Lenguajes artísticos	Seres vivos y su entorno	Grupos humanos, sus formas de vida y acontecimientos relevantes	Relaciones lógico-matemáticas y cuantificación				
Mapas de progreso (3 a 4 años)				CO	IL	IE						RLM	C	
<i>dip</i>	0	0	0	4	22	3	0	0	0	0	42	21	92	

Tabla 3
Detalle del procedimiento en relación con las fases del estudio

	Marzo de 2011	Octubre de 2011	Noviembre-enero de 2012
Fase	Aplicación Prepiloto	Aplicación piloto	Aplicación experimental
Criterios a evaluar	1. Evaluación de la exigencia por tramo de edad 2. Evaluación agrupación de ítems en categorías (cuantificación, comprensión, tiempo y espacio)	Aplicación a muestra de 60 niños de 3 tipos de dependencia administrativa	Aplicación a muestra de 360 niños de 3 tipos de dependencia administrativa
Cambios	1. Reducción de ítems 2. Asegurar progresión de dificultad	1. Corrección de errores en diseño y programación 2. Se detecta "efecto techo" en la evaluación en todas las edades 3. Se modifican ítems 4. Se modifica la pauta de evaluación 5. Se asegura que cada evaluadora cuente con un tablet	
Características evaluadores		1. Ocho tituladas en Educación de Párvulos	1. Diez tituladas en Educación de Párvulos y Psicología
Cantidas ítems	39 por tramo de edad	28 ítems por tramo de edad más 12 Entremundos	92 ítems en total, para todas las edades
Tiempo aplicación	60 minutos por tramo edad	45 minutos en total	20-30 minutos en total
Criterios de calidad	1. Se aplican todos los ítems a todos los niños de todas las edades	1. Se implementa una plataforma de corrección en línea 2. Una sesión de entrenamiento a evaluadoras; se les entrega manual de aplicación 3. Se entrega consentimiento informado a los apoderados	1. Se capacita a las evaluadoras en dos sesiones 2. Se realiza un <i>focus group</i> con las evaluadoras a la mitad de la aplicación 3. La plataforma de corrección entrega casos al azar a las evaluadoras para la asignación de puntajes 4. Se utiliza corrección cruzada: si dos evaluadoras no coinciden en la asignación de puntajes para el mismo caso, este caso pasa a un tercer evaluador que actúa como juez 5. El establecimiento educacional entrega un espacio para evaluar durante la jornada de clases

Procedimiento de aplicación

Para la muestra experimental, la primera aproximación a los centros educativos fue telefónica. A continuación se concertó una reunión con la directiva del establecimiento. Se consideraron aptos para el estudio aquellos centros que accedieran formalmente, a través de su directiva, al estudio. A cada una de las instituciones elegidas se les entregó, a través de los alumnos, consentimiento informado a los apoderados de los niños. Solo los niños que retornaron al establecimiento el consentimiento firmado por el cuidador responsable fueron incluidos en la evaluación. Por otra parte, los criterios de selección para los evaluadores que aplicaron el test *dip* en jardines infantiles y escuelas fueron: ser educador de párvulos egresado o titulado con todas las prácticas al día y tener experiencia en aula en el trabajo con niños y, dentro de lo posible, tener recomendaciones. También se consideraron psicólogos de 4° año o más con experiencia en trabajo de campo o práctica clínica, egresados o titulados. Todos los evaluadores presentaron sus currículos académicos y profesionales y participaron en un proceso de capacitación que duró tres sesiones, de dos horas cada una. En estas instancias se capacitó en el adecuado uso del tablet y en la entrega de instrucciones para los diferentes ítems y subpruebas, se explicitaron acuerdos relativos a la adecuada presentación y comportamiento en los establecimientos educacionales, se formalizaron las responsabilidades y compromisos, se entregaron conocimientos básicos sobre evaluación y contexto de evaluación y se resolvieron dudas. Posteriormente a las capacitaciones, se seleccionó a los evaluadores más competentes según su desempeño en terreno, disponibilidad de tiempo, compromiso y responsabilidad.

Una de las limitantes de este estudio, y que tiene implicancias para la validación del *dip*, es que la muestra no es representativa del país, al concentrarse en la Región Metropolitana.

Resultados

Las medias y desviación estándar indican que existen diferencias por sexo, donde las niñas se desempeñan mejor que los niños en todos los ejes de aprendizaje (véase la Tabla 4 para ver las medias por sexo). Estas diferencias son estadísticamente significativas, en especial en los ejes de razonamiento lógico-matemático $F(1,331) = 11,88$, $p = 0,001$ y cuantificación $F(1,290) = 8,35$, $p = 0,004$ (véase la Tabla 5 para las diferencias estadísticamente significativas por sexo).

Tabla 4
Medias por género y eje de aprendizaje

Sexo		<i>dip</i> eje RLM	<i>dip</i> eje C	<i>dip</i> eje CO	<i>dip</i> eje IL	<i>dip</i> eje IE
Femenino	Media	26,7	14,4	2,3	15,9	3,0743
	N	163	163	163	163	148
	Desv. típ.	7,8	4,5	0,8	4,0	1,9
Masculino	Media	23,7	13,3	2,1	14,8	2,4
	N	170	170	170	170	144
	Desv. típ.	7,8	4,9	1,0	4,3	2,0

Nota: RLM = relaciones lógico matemáticas, C = cuantificación, C = comunicación oral, IL = iniciación a la lectura, IE = iniciación a la escritura.

Tabla 5
Diferencias por género y eje de aprendizaje

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
<i>dip</i> eje Razonamiento Lógico-matemático * sexo	Intergrupos (combinadas)		728,9	1	728,9	11,88	0,001**
	Intragrupos		20307,1	331	61,3		
	Total		21036,1	332			
<i>dip</i> eje cuantificación* Sexo	Intergrupos (combinadas)		97,8	1	97,8	4,34	0,038*
	Intragrupos		7451,0	331	22,5		
	Total		7548,8	332			
<i>dip</i> eje comunicación oral * Sexo	Intergrupos (combinadas)		4,3	1	4,3	4,86	0,028*
	Intragrupos		294,2	331	,889		
	Total		298,5	332			
<i>dip</i> eje iniciación a la lectura * sexo	Intergrupos (combinadas)		96,1	1	96,1	5,42	0,021*
	Intragrupos		5872,6	331	17,7		
	Total		5968,8	332			
<i>dip</i> eje iniciación a la escritura * sexo	Intergrupos (combinadas)		32,9	1	32,9	8,35	0,004**
	Intragrupos		1142,8	290	3,9		
	Total		1175,7	291			

Se observaron diferencias por dependencia administrativa y ejes de aprendizaje. Los centros de dependencia municipal presentaron menores puntajes brutos totales en la prueba *dip* en comparación con los centros particulares subvencionados y pagados ($M = 51$, $SD = 12,3$). Estas diferencias se repiten en todos los ejes de aprendizaje (véase la Tabla 6 para reportes de medias por dependencia administrativa). Un análisis posterior indicó que las diferencias entre centros no son estadísticamente significativas para los ejes de comunicación oral $F(2,33) = 0,59$, $p = 0,551$ e iniciación a la escritura $F(2,28) = 0,70$, $p = 0,496$ (véase la Tabla 7 para un reporte de diferencias significativas por dependencia administrativa y eje de aprendizaje).

Tabla 6
Reporte de medias por dependencia administrativa

Tipo de colegio		dip	dip eje RLM	dip eje C	dip eje CO	dip eje IL	dip eje IE
		Puntaje					
Municipal	Media	51,2	23,1	12,5	2,2	14,4	2,5
	N	131	131	131	131	131	111
	Desv. tip.	12,3	7,4	4,0	,9	3,6	1,9
Particular subvencionado	Media	57,8	26,1	14,7	2,2	16,3	2,8
	N	96	96	96	96	96	86
	Desv. tip.	14,9	8,4	5,4	1,0	4,4	2
Particular pagado	Media	57,9	26,8	14,8	2,3491	15,7	2,8
	N	106	106	106	106	106	95
	Desv. tip.	13,9	7,6	4,6	,90546	4,4	2,1

Nota: RLM = relaciones lógico-matemáticas, C = cuantificación, CO = comunicación oral, IL = iniciación a la lectura, IE = iniciación a la escritura.

Tabla 7
Reporte de diferencias significativas por dependencia administrativa y eje de aprendizaje

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
<i>dip</i> puntaje * dependencia adm.	Intergrupos	(combinadas)	3575,4	2	1787,7	9,62	0,000**
	Intragrupos		61307,7	330	185,7		
	Total		64883,2	332			
<i>dip</i> eje RLM * dependencia adm.	Inter-grupos	(combinadas)	912,5	2	456,2	7,48	0,001**
	Intra-grupos		20123,5	330	60,9		
	Total		21036,1	332			
<i>dip</i> eje C * dependencia adm.	Intergrupos	(combinadas)	418,8	2	209,4	9,69	0,000**
	Intragrupos		7129,9	330	21,6		
	Total		7548,8	332			
<i>dip</i> eje CO * dependencia adm.	Intergrupos	(combinadas)	1,0	2	0,5	0,59	0,551
	Intragrupos		297,5	330	0,9		
	Total		298,5	332			
<i>dip</i> eje IL * dependencia adm.	Intergrupos	(ombinadas)	212,4	2	106,2	6,08	0,003*
	Intragrupos		5756,3	330	17,4		
	Total		5968,8	332			
<i>dip</i> eje IE * dependencia adm.	Intergrupos	(combinadas)	5,6	2	2,8	0,70	0,496
	Intragrupos		1170,0	289	4,0		
	Total		1175,7	291			

Nota: RLM = relaciones lógico-matemáticas, C = cuantificación, CO = comunicación oral, IL = iniciación a la lectura, IE = iniciación a la escritura.

Evidencia de confiabilidad

El instrumento *dip* evalúa aprendizajes esperados agrupados en 5 ejes (razonamiento lógico-matemático, cuantificación, comunicación oral, iniciación a la escritura, iniciación a la lectura). El Alpha de Cronbach para el total de puntajes del test *dip* es de 0,7.

El eje de cuantificación consta de 21 ítems ($\alpha = 0,8$), el eje de razonamiento lógico-matemático consta de 42 ítems ($\alpha = 0,9$), el eje de comunicación oral consta de 4 ítems ($\alpha = 0,8$) y el eje de iniciación a la lectura consta de 22 ítems ($\alpha = 0,7$). En cuanto al eje de iniciación a la escritura, cabe mencionar que las respuestas de los niños de 3 y 4 años se evalúan con base en los criterios de estructura y contenido, mientras que los niños de 5 y 6 son evaluados en relación con los criterios de estructura, contenido y longitud. Dado que los criterios aplicados a cada edad no son iguales, para el análisis de confiabilidad se dividió a los niños en dos grupos según su edad: el eje de iniciación a la escritura para niños de 3 y 4 años tiene 2 ítems ($\alpha = 0,9$) y 3 ítems para niños entre 5 y 6 años ($\alpha = 0,9$).

La evidencia de confiabilidad para cada eje de aprendizaje se detalla en la siguiente tabla:

Tabla 8
Evidencia de confiabilidad por eje de aprendizaje

	Alpha Conbrach	N.º de ítems
Eje cuantificación	0,8	21
Eje razonamiento lógico-matemático	0,9	42
Eje comunicación oral	0,8	4
Eje iniciación a la lectura	0,7	22
Eje iniciación a la escritura		
3 y 4 años	0,9	3
5 y 6 años	0,9	

Evidencia de validez

Progresión de los puntajes del test *dip* en función de la edad de la muestra

Al comparar los resultados promedio obtenidos por los niños en el test *dip* y en los distintos ejes, se aprecia una progresión creciente y significativa en función de la edad de los niños ($p < 0,001$). Esto es esperable de acuerdo con los aprendizajes que logran los niños a lo largo del desarrollo, lo que da cuenta de un instrumento que permite discriminar entre las distintas edades. En la siguiente tabla se presentan las medias y desviaciones estándar para el puntaje *dip* y cada uno de los ejes de aprendizaje para las distintas edades (véase la Tabla 9 y la Figura 2).

Tabla 9
Progresión de los puntajes promedio obtenidos por edad de los niños

Edad	3 años	4 años	5 años	6 años
	Media (DE)	Media (DE)	Media (DE)	Media (DE)
Eje razonamiento lógico-matemático	17,8 (5,7)	24,1 (6,2)	29,3 (6,4)	32,3 (5,6)
Eje cuantificación	6,7 (2,3)	9,2 (2,7)	11,7 (2,8)	13,5 (2,8)
Eje comunicación oral	3,6 (2,7)	6,7 (2,9)	8,3 (3,0)	10,3 (3,2)
Eje iniciación a la lectura	9,0 (2,8)	9,8 (2,9)	11,6 (3,0)	14,0 (2,9)
Eje iniciación a la escritura	0,8 (1,1)	3,3 (1,8)	3,3 (1,8)	3,8 (1,9)

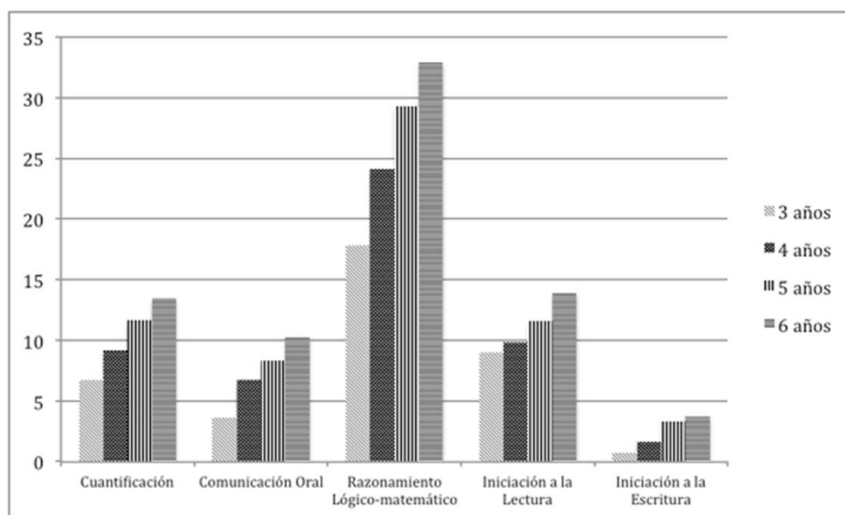


Figura 2. Progresión de los puntajes promedio obtenidos, por edad de los niños y por eje de aprendizaje evaluado.

Evidencia de validez concurrente

Los resultados obtenidos por la muestra en el test *dip* fueron correlacionados con otras pruebas que evalúan constructos similares. En este caso, se comparó el rendimiento de *dip* con dos pruebas diferentes: a) ABCDeti (Prueba de Habilidades Lectoras Iniciales, Rosas et al., 2011) y b) PLAEP-R (Perfil de Logros de Aprendizajes en la Educación Parvularia, versión revisada, Fundación Integra, 2009). Además, para evaluar el efecto de la plataforma tecnológica como potencial factor de invalidez en la población rural, se aplicó una versión de *dip* en papel y en tablet a una muestra de 26 niños de una escuela rural.

En la Tabla 10 se presentan los resultados de la correlación entre el total de puntajes de la prueba *dip* y el total de puntajes de la prueba ABCDeti. Los resultados indican que la prueba presenta evidencia de validez para la evaluación de precursores de la lectura ($r = 0,7$ $n = 12$ $p = 0,01$).

Tabla 10
Evidencia de validez concurrente entre *dip* y una prueba de evaluación de habilidades lectoras iniciales

		<i>dip</i>	ABCDeti
<i>dip</i>	Correlación de Pearson	1	0,7*
	Significancia (bilateral)		0,01
	N	333	12
ABCDeti	Correlación de Pearson	0,7*	1
	Significancia (bilateral)	0,01	
	N	12	14

*La correlación es significativa al nivel 0,05 (bilateral).

PLAEP-R (2009) es una herramienta de evaluación en formato tradicional cuya batería estandarizada solo se encuentra disponible para niños de la Fundación Integra y es utilizada para apoyar a la Fundación en la toma de decisiones a nivel regional e institucional. PLAEP-R fue estandarizada para niños de entre 1 y 5 años pertenecientes a los establecimientos Integra a nivel nacional. Esta herramienta presenta evidencias de confiabilidad; sin embargo, no presenta evidencias de validez con un criterio externo, en parte porque hasta el año 2012 era el único instrumento que existía en Chile que evaluaba aprendizajes esperados (Fundación Integra, 2009). En la Tabla 11 se presentan los resultados de la correlación entre *dip* y PLAEP-R.

Tabla 11
Evidencia de validez concurrente entre *dip* y PLAEP-R

		Total PLAEP-R_Bruto	Total <i>dip</i> -Bruto	Puntaje <i>dip</i> cuantificación	Puntaje <i>dip</i> comunicación oral	Puntaje <i>dip</i> razonamiento lógico-matemático	Puntaje <i>dip</i> iniciación a la lectura	Puntaje <i>dip</i> iniciación a la escritura
Total PLAEP-R_Bruto	Correlación de Pearson	1	0,520**	0,121	0,582**	0,591**	0,011	0,052
	Sig. (bilateral)		0,001	0,465	0,000	0,000	0,947	0,752
	N	39	39	39	39	39	39	39

**La correlación es significativa al nivel 0,01 (bilateral)

Como se aprecia en la Tabla 11, hay una correlación significativa, aunque moderada entre los puntajes totales de las escalas ($r = 0,520$, $n = 39$, $p = 0,001$). Este resultado entrega evidencia preliminar de validez concurrente. La correlación moderada puede explicarse al mirar la relación entre los ejes de aprendizaje de *dip* y los puntajes totales de PLAEP-R; se evidencia una relación moderada y altamente significativa en comunicación oral ($r = 0,582$, $n = 39$, $p = 0,000$) y razonamiento lógico-matemático ($r = 0,591$, $n = 39$,

$p = 0,000$); sin embargo, no es significativa en la escala de cuantificación ($r = 0,121$, $n = 39$, $p = 0,465$), iniciación a la lectura ($r = 0,011$, $n = 39$, $p = 0,947$) y a la escritura ($r = 0,052$, $n = 39$, $p = 0,752$). De este modo, a pesar de derivarse de los mismos conceptos, *dip* y PLAEP-R evalúan aspectos separados de aprendizajes esperados en los ejes antes indicados. Si bien se requieren más investigaciones para comprobar esto, se propone que la prueba *dip* evalúa competencias lectoras iniciales que no se evalúan en PLAEP-R, por lo tanto, actúa en esta dimensión como una prueba complementaria al núcleo de lenguaje verbal de PLAEP-R. En cuanto a la iniciación a la escritura, se requiere investigación adicional para determinar la validez de tal escala. Se concluye que ambas pruebas no comparten la dimensión de cuantificación a partir de una revisión ítem a ítem de ambas pruebas, por lo que es más adecuado considerar la relación como evidencia de validez discriminante, resultado que demuestra la especificidad de la prueba *dip*.

Resulta necesario, para futuras investigaciones, comparar ambas pruebas con un tercer criterio, de modo de determinar la naturaleza de la evaluación en las escalas de cuantificación e iniciación a la escritura.

Comparación entre aplicación de *dip* por medio digital y en papel

Con el fin de descartar posibles efectos de invalidez debidos al formato de presentación de la prueba, se hizo un microestudio en el que se comparó, en una muestra de 26 niños provenientes de una escuela rural, su rendimiento tanto en formato digital como en formato en papel. La primera modalidad consistió en la presentación de la prueba en su formato estándar, mientras que la segunda correspondió a una adaptación de la prueba, presentada en un formato impreso. Se excluyeron los ítems que, por su naturaleza, no eran posibles de adaptar al formato de lápiz y papel (aquellos que indican movimiento, por ejemplo).

En primer lugar, se comparó, en términos del rendimiento, la ejecución de los niños en las diferentes dimensiones de la prueba (véase la Tabla 12). Como se aprecia en la tabla, ninguna de las diferencias resultó estadísticamente significativa, lo que permite concluir que el formato digital es perfectamente utilizable en poblaciones para las que se suponga que tienen una desventaja por carecer de experiencia con plataformas digitales.

Tabla 12
Comparación de puntajes *dip* en medio digital y *dip* en papel

Ejes	Digital	Analógico	Significación
Razonamiento lógico-matemático	17,8	18,8	NS
Cuantificación	8,2	8,1	NS
Comunicación oral	2,1	2,4	NS
Iniciación a la escritura	2,1	2,6	NS
Iniciación a la lectura	4,9	5,3	NS

En segundo lugar, se estudió la correlación en el rendimiento de ambas versiones y se obtuvieron resultados que permiten sugerir que ambas versiones son paralelas, si se toma la medida de correlación como medida de confiabilidad (véase la Tabla 13). La única escala que debe tomarse con cautela al ser aplicada en poblaciones desaventajadas con tecnologías es la de escritura, que muestra una confiabilidad en el límite de lo aceptable.

Tabla 13
Correlación de puntajes *dip* en medio digital y *dip* en papel

Ejes	Coefficiente de correlación	Significancia (bilateral)	N
Razonamiento lógico-matemático	0,9	0,0	26
Cuantificación	0,8	0,0	26
Comunicación oral	0,9	0,0	26
Iniciación a la escritura	0,7	0,0	26
Iniciación a la lectura	0,8	0,0	26

Discusión

Test *dip* como aporte para la planificación y el apoyo del proceso de enseñanza y aprendizaje

El propósito general de la evaluación en un contexto educacional es ayudar a los docentes y a otros actores involucrados en la educación a tomar decisiones pedagógicas basadas en evidencia para informar y/o ajustar los procesos de enseñanza y aprendizaje (Brassard & Boehm, 2007).

Idealmente, la evaluación es un proceso dinámico y continuo que utiliza una variedad de mediciones y aproximaciones, se enfoca en el contexto de aprendizaje del niño, se utiliza para descubrir sus fortalezas y las áreas de desarrollo emergente, informa sobre estrategias e intervenciones apropiadas y se lleva a cabo con la perspectiva de que el niño va a cambiar. Asimismo, presupone que, mientras más temprano es la intervención mayores son las probabilidades de producir resultados beneficiosos (Brassard & Boehm, 2007). Por último, los padres necesitan estar involucrados en la evaluación del proceso de muchas formas, no solo para proporcionar información acerca del desarrollo del niño y sus necesidades particulares sino también cobrar una mayor conciencia acerca de la importancia de su presencia en el desarrollo (Brassard & Boehm, 2007).

En Chile, la evaluación en educación parvularia ha estado ligada a la medición de desarrollo, inteligencia y aptitud para el ingreso a la educación básica. La mayoría de las evaluaciones se hacen en torno a evaluaciones sumativas al interior de los jardines infantiles y salas cuna (por ej., PLAEP-R). Este uso de pruebas estandarizadas responde a la necesidad de toma de decisiones a nivel nacional e institucional y al reporte de evaluaciones de resultados (sobre el nivel de aprendizaje de los niños al final de un proceso o nivel educativo) que permiten monitorear avances y problemas a mayor escala. Para una comparación detallada entre *dip* y PLAEP-R, véase la Tabla 14.

Tabla 14
Comparación *dip* y PLAEP-R

Características	<i>dip</i>	PLAEP-R
Tipo de evaluación	Con énfasis en la evaluación formativa	Medición a nivel institucional de tipo sumativa (resultados)
Medio de evaluación	Evaluación invisible a través del juego Dispositivo digital de pantalla táctil	Evaluación tradicional Objetos reales
Ámbito evaluado	Dos ámbitos.: núcleos de aprendizaje de relaciones lógico-matemáticas y de cuantificación y lenguaje verbal	Todos. Al menos un ítem por núcleo
Edades	3 a 6 años de edad	11 meses a 5 años 3 meses de edad
Portabilidad	Alta, tablet digital	Limitada; batería de cerca de 15 objetos por edad
Estandarización ítems	Batería se encuentra estandarizada en cuanto a sus componentes (software que se carga en un tablet)	Batería se compone de objetos que se deben reunir en el comercio
Tiempo de aplicación total	20-30 minutos	Entre 35 y 50 minutos dependiendo de la edad
Tipo de corrección	Automática casi en su totalidad para reducir errores de puntuación. Solo 2 ítems de corrección manual	Manual; se hace en el momento en que el niño responde
Tiempo de corrección	Los ítems que requieren corrección manual toman 5 minutos.	Alrededor de 10 minutos por ítem
Acceso a la prueba	Por venta/convenio	Limitada a menos que pertenezca a Integra

Como se ha indicado anteriormente, el resultado del test *dip* aplicado a un niño se compone de los cinco ejes explorados a lo largo de la prueba: razonamiento lógico-matemático, cuantificación, comunicación oral, iniciación a la lectura e iniciación a la escritura. Cada niño obtiene para cada eje evaluado un nivel de desempeño que puede corresponder a lo esperado para la edad o puede dar cuenta de aprendizajes

que están en proceso o que requieren intervención. Todos estos niveles de desempeño están orientados a criterios, y en particular, a los aprendizajes esperados para la edad de cada niño.

Con el objetivo de que el reporte del test *dip* sea un verdadero aporte para la planificación y el apoyo que las educadoras y otros profesionales dan a los niños, se especifica también el nivel de desempeño desagregado por los conceptos relacionales evaluados.

Como *dip* es un test de carácter formativo, y considerando lo expuesto en los párrafos anteriores, es importante que si el puntaje de un niño parece bajo, se utilicen modos complementarios de evaluación para obtener un cuadro más completo.

Tal vez el aporte más significativo de *dip*, y que contribuye directamente al mejoramiento de la calidad de la educación parvularia, sea proporcionar insumos y objetivos cuantificables para tomar decisiones pedagógicas basadas en evidencia. Los resultados obtenidos permitirán a la educadora en aula proponer planes específicos para modificar la planificación de la cobertura curricular con miras a responder adecuadamente a la diversidad de aprendizajes en el aula y/o derivar a un especialista cuando corresponda.

Los desafíos pendientes para *dip*, y que están siendo abordados por los autores en otros proyectos de investigación presentes y futuros, son, entre otros: la necesidad de llevar a cabo estudios para examinar el comportamiento y viabilidad de *dip* para monitorear el logro de aprendizajes esperados en dos y tres momentos distintos del año, realizar investigación que incluya el uso de *dip* como herramienta que guíe la toma de decisiones pedagógicas en conjunto con instrumentos para evaluar funciones ejecutivas para lograr entregar información más específica para el trabajo en aula; e incorporar estudios con una muestra más amplia y representativa a nivel nacional para mejorar la validez de la prueba. De especial consideración en estudios futuros es la incorporación de variables tradicionales asociadas a aprendizaje y desarrollo infantil, tales como el ingreso familiar, para determinar nivel socioeconómico y género. En este estudio las diferencias reportadas en relación con género y dependencia administrativa no son concluyentes, debido a la limitante de representatividad antes mencionada. Sin embargo, ameritan análisis más profundo. Por último, es necesario estudiar una manera de evaluar, de manera complementaria con esta tecnología y respetando el principio de evaluación invisible, aquellos ámbitos de aprendizaje que implican una relación del niño con otros (a saber, el ámbito de formación personal y social que incluye autonomía, identidad y convivencia).

El artículo original fue recibido el 3 de junio de 2013

El artículo revisado fue recibido el 13 de agosto de 2013

El artículo fue aceptado el 3 de septiembre de 2013

Referencias

- Boehm, A. (2001). *Boehm-3 Preschool: Boehm test of basic concepts*. Bloomington, MN: Pearson.
- Boehm, A. (2013). *Assessment to intervention using the Boehm Test of Basic Concepts-Third Edition (Boehm-3)*, 1-10. Recuperado de http://images.pearsonclinical.com/images/Assets/Boehm-3/Boehm_White_paper_RPT.pdf
- Bransford, J., Derry, S., Berliner, D., Hammerness, K., & Beckett, K. (2005). Theories of learning and their roles in teaching. En L. Darling-Hammond, & J. Bransford (Eds.), *Preparing teachers for a changing world: what teachers should learn and be able to do* (pp. 40-87). San Francisco CA: John Wiley & Sons.
- Brassard, M., & Boehm, A. (2007). *Preschool assessment: principles and practices*. Nueva York, NY: The Guilford Press.
- Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas (CPEIP), Ministerio de Educación (Mineduc) (2008). *Marco para la buena enseñanza*. Recuperado de <http://www.docentemas.cl/docs/MBE2008.pdf>
- Fundación Integra (2009). *Perfil de logros de aprendizajes en la educación parvularia versión revisada (PLAEP-R)*. Santiago, Chile.
- Heckman, J., & Masterov, D. (2004). *The productivity argument for investing in young children*. [Documento de trabajo 5]. Recuperado de http://jenni.uchicago.edu/Invest/FILES/dugger_2004-12-02_dvm.pdf
- Kelly-Vance, L., & Ryalls, B. O. (2005). Best practices in play assessment and intervention. En J. Grimes, & A. Thomas (Eds.), *Best practices in school psychology V* (pp. 549-560). Bethesda, MD: National Association of School Psychologists.
- Knoche, L., Sheridan, S., Edwards, C., & Osborn, A. (2010). Implementation of a relationship-based school readiness intervention: a multidimensional approach to fidelity measurement for early childhood. *Early Childhood Research Quarterly*, 25, 299-313. doi: 10.1016/j.ecresq.2009.05.003
- Kolb, B., & Fantie, B. (2009). Development of the child's brain and behavior. En C. Reynolds, & E. Fletcher-Janzen (Eds.), *Handbook of Clinical Child Neuropsychology* (pp. 19-46). Nueva York, NY: Springer Science + Business Media. doi: 10.1007/978-0-387-78867-8_2
- Leseman, P. (2002). *Early childhood and care for children from low-income or minority backgrounds*. OCDE: París.
- Manzi, J., González, R. y Sun, Y. (2011). *La evaluación docente en Chile*. Santiago, Chile: Centro de Medición MIDE UC.
- Ministerio de Educación (Mineduc) (2005). *Bases curriculares de la educación parvularia*. Santiago, Chile: Mineduc.
- OCDE (2012). *Panorama de la educación 2011: Indicadores de la OCDE*. España: Santillana. doi: 10.1787/9789264180482-es
- Rosas, R., Medina, L., Meneses, A., Guajardo, A., Cuchacovich, S. y Escobar, P. (2011). Construcción y validación de una prueba de evaluación de competencia lectora inicial basada en computador. *Pensamiento Educativo*, 48(1), 43-61. doi: 10.7764/PEL.48.1.2011.4
- Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P., Grau, V., Lagos, F., López, X., López, V., Rodríguez, P., & Salinas, M. (2003). Beyond Nintendo: design and assessment of educational video games for first and second grade students. *Computers and Education*, 40(1), 71-94. doi: 10.1016/S0360-1315(02)00099-4
- Shute, V. J., & Kim, Y. J. (2014). Formative and stealth assessment. En J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology (4th Edition)* (pp. 311-321). Nueva York NY: Lawrence Erlbaum Associates, Taylor & Francis Group. doi: 10.1007/978-1-4614-3185-5_25
- Siegler, R. (1998). *Cognition in children*. Upper Saddle River: Prentice Hall.
- Vernon-Feagans, L., Odom, E., Pancsofar, N., & Kainz, K. (2007). Farkas and Hibel, a transactional/ecological model of readiness and inequality: person/process/context interactions. En A. Booth, & A. C. Crouter (Eds.), *Disparities in school readiness. How families contribute to transitions into school* (pp. 61-78). Mahwah, NJ: Erlbaum.
- Welsh, J., Nix, R., Blair, C., Bierman, K., & Nelson, K. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology*, 102(1), 43-53. doi: 10.1037/a0016738