PSYKHE
2015, Vol. 24, 1, 1-11
doi:10.7764/psykhe.23.2.724


ISSN 0717-0297
www.psykhe.cl

# ¿Pruebas Tradicionales o Evaluación Invisible a Través del Juego? Nuevas Fronteras de la Evaluación Cognitiva

# Traditional Assessment or Invisible Assessment Using Games? New Frontiers in Cognitive Assessment

Ricardo Rosas
Pontificia Universidad Católica de Chile

Francisco Ceric
Universidad del Desarrollo

Andrés Aparicio, Paulina Arango, Rodrigo Arroyo, Catalina Benavente, Pablo Escobar, Polín Olguín, Marcelo Pizarro, María Paz Ramírez, Marcela Tenorio y Soledad Véliz
Pontificia Universidad Católica de Chile

Se aborda el problema de ansiedad relacionado con las evaluaciones tradicionales, que pueden afectar los resultados y subestimar el rendimiento de los sujetos. Las evaluaciones invisibles permiten evaluar a los sujetos sin que estos tengan la sensación de ser evaluados. Se desarrollaron pruebas de evaluación invisible en soporte tableta táctil, considerando los dominios cognitivos de inteligencia, cálculo y lectura, las que fueron aplicadas a 337 niños, entre kínder y tercero básico, de 3 colegios particulares subvencionados de Santiago, Chile. Los colegios fueron seleccionados por conveniencia, incluyendo a todos los niños cuyos padres firmaron consentimiento informado. La muestra final se distribuyó al azar entre los dominios. Se observaron correlaciones entre las pruebas de evaluación tradicional e invisible. Los niños reportaron una preferencia por las pruebas de evaluación invisible por sobre las tradicionales. Según un análisis mixto de varianza, los niños con bajo rendimiento escolar obtuvieron mejores resultados en las pruebas de evaluación invisible que en las tradicionales. Los resultados sugieren que es posible evaluar dominios cognitivos con instrumentos no tradicionales, permitiendo estos el acceso al rendimiento real de los sujetos.

*Palabras clave*: evaluación, ansiedad, juego

This paper addresses the problem of anxiety related to traditional assessment, which can affect assessment outcomes and underestimate the performance of subjects. Invisible assessment makes it possible to evaluate subjects without making them feel like they are being evaluated. Invisible evaluation tests were developed for touch screen tablets for 3 cognitive domains: intelligence, calculation, and reading. These tests were applied to 337 children from kindergarten through third grade, who attended 3 mixed-funding schools in Santiago, Chile. The schools were convenience sampled and all the children whose parents signed the informed consent form were included. The final sample was randomly distributed among the domains. Correlations between traditional assessment tests and invisible assessment tests were observed. Children reported a preference for invisible assessment over traditional assessment. Subjects with low academic performance obtained better scores on invisible assessment tests than on traditional tests, according to a mixed factors analysis of variance. These findings suggest that it is possible to assess cognitive domains with non-traditional instruments and that they can reveal the real academic performance of subjects.

*Keywords*: assessment, anxiety, games

En este artículo abordamos el problema de la evaluación de dominios cognitivos en niños que están en edad escolar. Específicamente, nos preguntamos si es posible evaluar por medio de juegos presentados en tabletas digitales, dominios cognitivos, como lectura, cálculo e inteligencia, en niños que atraviesan los primeros años de educación formal y, además, si dicha evaluación se asemeja a la evaluación tradicional en sus tasas de predicción de adaptación escolar. Asimismo, trataremos el tema de la preferencia de los niños por alguna de estas actividades y su rendimiento diferencial en estas pruebas, de acuerdo a su rendimiento promedio.

---

Ricardo Rosas, Escuela de Psicología, Pontificia Universidad Católica de Chile, Santiago, Chile; Francisco Ceric, Facultad de Psicología, Universidad del Desarrollo, Santiago, Chile; Andrés Aparicio, Paulina Arango, Rodrigo Arroyo, Catalina Benavente, Pablo Escobar, Polín Olguín, Marcelo Pizarro, María Paz Ramírez, Marcela Tenorio y Soledad Veliz, Centro de Desarrollo de Tecnologías de Inclusión (CEDETi UC), Escuela de Psicología, Pontificia Universidad Católica de Chile.

Esta investigación fue financiada por el Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) a través del proyecto 1110975 a cargo del primer autor. La participación de Marcela Tenorio en esta investigación fue posible gracias al apoyo de FONDECYT Postdoctoral 3031612.

La correspondencia relativa a este artículo debe ser dirigida a Ricardo Rosas D., Centro de Desarrollo de Tecnologías de Inclusión (CEDETi UC), Pontificia Universidad Católica de Chile, Avda. Vicuña Mackenna 4860, Edificio MIDE UC, piso 2, Macul, Santiago, Chile. E-mail: rrosas@uc.cl

La evaluación cognitiva se ha realizado tradicionalmente en formato de lápiz y papel o con apoyo de material concreto (e.g., Cubos en las subpruebas Wechsler). Este tipo de evaluación tiene ventajas en razón de sus conocidos buenos índices de confiabilidad y validez (Wechsler, 2003, 2008), pero tiene desventajas dadas por la cantidad de tiempo, recursos que consumen y los posibles factores de invalidez de los resultados relacionados con la ansiedad o desinterés de los niños presentes en largas y rutinarias sesiones de evaluación (Eum & Rice, 2011; Nie, Lau & Liau, 2011; Urhahne, Chao, Florineth, Luttenberger & Paechter, 2011). ¿A qué factores de invalidez nos referimos? Aludimos especialmente a aquellos relacionados con una posible subestimación de capacidades en niños que, por su historia de rendimientos negativos, tienen una percepción de autoeficacia baja en situaciones de evaluación tradicional, la que incide en sus resultados. Bandura (2006) señala que es esencial que las personas crean que sus acciones tendrán los efectos deseados. Aquellos estudiantes con creencias de baja competencia tienen predominantemente metas de logro caracterizadas por la evitación de las tareas, debido al miedo al fracaso que les representan las situaciones de evaluación, que les generan ansiedad (Elliot & Pekrun, 2007; Putwain & Symes, 2012).

En función de lo anterior, ¿qué modificaciones se requieren sobre la situación de evaluación tradicional para lograr superar sus carencias y ofrecer mejores predicciones de rendimiento escolar? Nuestra respuesta es evaluar los mismos dominios de una manera diferente, es decir, implementando situaciones que llamaremos de *evaluación invisible a través del juego* en las que, de manera menos explícita, logramos explorar las mismas dimensiones accedidas por pruebas tradicionales. En esta investigación entendimos la ansiedad como un factor que es observable conductualmente (Beck, Emery & Greenberg, 1985), especialmente en la situación de evaluación, y que el sujeto puede verbalizar. Consideramos que el cambio en la situación de evaluación genera un cambio en la percepción del evaluado ante la evaluación, que es observable conductualmente.

## ¿Qué Es una Evaluación Invisible?

Por evaluación invisible entendemos una forma de evaluación en la que el evaluado no es consciente de que está siendo evaluado, ya que sus contenidos están encubiertos en una actividad diferente, por ejemplo, un juego.

Este concepto es similar al de *stealth assessment* o evaluación encubierta, que hace referencia a evaluaciones que se encuentran tan profundamente incorporadas en el contexto de aprendizaje/evaluación que se vuelven invisibles (Shute, 2011). Shute utiliza videojuegos para evaluar habilidades cognitivas, ya que, a su juicio, su fortaleza se encuentra en su capacidad para apoyar la emergencia de habilidades complejas, lo que permite evaluar esas habilidades mientras se juega. Por lo tanto, la incorporación de medidas de evaluación integrada al juego, como tareas de tutorial o puntajes que proveen feedback sobre el desempeño del jugador, conforman el mecanismo ideal para evaluar estas competencias. La evidencia necesaria para evaluar estas habilidades es provista por el juego mismo, la cual puede ser contrastada con los productos de una actividad, por ejemplo, la norma de ambientes educativos (Shute, 2011).

Hay otras investigaciones que comparan pruebas en formato de juego y tradicionales para evaluar diferentes conocimientos o constructos psicológicos. Por ejemplo, Desrochers, Pusateri Jr. y Fink (2007) demuestran que es posible obtener la misma media de desempeño en asignaturas de educación superior contra test tradicionales de lápiz y papel, siendo la diferencia entre ambos tipos de evaluación la experiencia subjetiva del evaluado. Bajo otra metodología, McPherson y Burns (2007) evaluaron un juego de computador orientado hacia la velocidad de procesamiento de información, obteniendo una correlación de 0,84 con una prueba de lápiz y papel estandarizada. Rosas y Bravo (2009) presentan evidencia a favor de una prueba lúdica medida por tecnología que permite la evaluación de la lectura inicial (Jugando con las Letras). En su análisis, su prueba y la Prueba de Comprensión Lectora de Complejidad Lingüística Progresiva (Alliende, Condemarin & Milicic, 2000) alcanzan una correlación de 0,79, valor más que deseable al realizar análisis de evidencia de validez.

Verhaegh, Fontijn, Aarts y Resing (2013) comparan la relación entre un juego, "Tap the Little Hedgehog", y varias subpruebas de la Escala de Inteligencia de Wechsler para Niños (WISC-III-NL). El juego utilizó la plataforma TagTiles, una consola física con interfaz analógica, que ha sido utilizada para evaluar habilidades cognitivas no verbales, como memoria de trabajo y razonamiento espacial, en niños entre 8 y 10 años. Los principales resultados de la investigación fueron los siguientes: (a) se observaron correlaciones significativas entre los resultados del juego y varias subpruebas de WISC-III-NL (Búsqueda de Símbolos, Completación de Figuras, Construcción con Cubos, Ensamblaje de Objetos, Retención de Dígitos y Vocabulario), (b) los niños

que participaron en la intervención disfrutaron los juegos y (c) los niños realizaron las tareas de manera independiente.

Nuestra concepción de evaluación invisible es muy parecida en el concepto a la de stealth assessment de Shute (2011), pero más parecida en su implementación a la de Verhaegh et al. (2013). Bajo evaluación invisible entendemos toda evaluación que cumpla con los siguientes requisitos: (a) sea invisible como "evaluación" para la persona evaluada (condición que se cumple aun en el caso de que la persona sea instruida para contestar una evaluación, ya que la naturaleza de la tarea hace que la olvide rápidamente y se involucre en el juego) y (b) permita obtener un puntaje que cumpla con los requisitos de construcción de pruebas psicométricas definidas en el estándar internacional (American Educational Research Association [AERA], American Psychological Association & National Council in Measurement in Education, 2002).

El cumplimiento del primer requisito implica dos diferencias fundamentales con la evaluación tradicional: (a) durante el proceso de evaluación invisible el sujeto no activa los esquemas o scripts (Schank & Abelson, 1977) de una situación de evaluación, lo que puede ser considerado una ventaja, especialmente para la evaluación de niños con malas experiencias o bajos resultados habituales en situaciones de evaluación y (b) las demandas cognitivas son esencialmente diferentes, ya que la evaluación invisible supone el despliegue de contenidos automatizados y procesos implícitos, al contrario de la tradicional, que normalmente pregunta por contenidos no automatizados o en proceso de aprendizaje por medio de procesos explícitos. Entonces, si bien durante ambos tipos de evaluaciones, la tradicional y la invisible, desplegamos procesos que requieren de la inversión de recursos, como la atención (Keogh & French, 2001), los sistemas de memoria (Eysenck, Derakshan, Santos & Calvo, 2007) y las funciones ejecutivas en términos de flexibilidad y control (Diamond & Lee, 2011), la focalización de esos recursos en ambos tipos de evaluación es totalmente diferente: en la tradicional el foco es el contenido (explícito); en la invisible es el proceso o la actividad realizada (en la cual el contenido está implícito o automatizado). Para la evaluación de contenidos implícitos es necesario tener presentes las trayectorias de desarrollo de las habilidades evaluadas (Farran & Karmiloff-Smith, 2012). En esta perspectiva se entiende que para lograr el desarrollo de una habilidad es necesario que haya otras habilidades o dominios que se formen antes. De esta manera, tenemos habilidades consideradas precursores de la habilidad, como es el caso de la conciencia fonológica para la lectura, en la que se considera que debe estar el precursor presente para que se logre el aprendizaje de la lectura (Georgiou, Parrila & Papadopoulos, 2008). En otros casos, hay dominios que son compartidos por distintas habilidades o que están a la base de otras habilidades o aprendizajes más complejos. Es el caso, por ejemplo, de la discriminación musical y la conciencia fonológica, que comparten el procesamiento auditivo, base necesaria para ambas habilidades (Anvari, Trainor, Woodside & Levy, 2002; Bolduc & Montésinos-Gelet, 2005).

## Nuestra Investigación

Coincidimos con Shute (2011) en que el juego es el mejor contexto para hacer una evaluación invisible, pero nuestro objetivo fundamental, en línea con la propuesta de Verhaegh et al. (2013), consistió en probar si en este contexto de evaluación se pueden desarrollar pruebas psicométricas con similares propiedades que las tradicionales y con similares valores predictivos en relación a los rendimientos de criterio. Sin embargo, a diferencia de la propuesta de Verhaegh et al. (2013), nos interesó probar si esto es posible de implementar en entornos completamente virtuales, como son los videojuegos.

El juego representa, entonces, nuestra mejor posibilidad para crear contextos de evaluación invisible. Apoyado en accesos implícitos al conocimiento, nos permitiría evaluar tanto aquello que el niño está seguro de saber como aquello que está ahí pero que no logra acceder desde el nivel consciente. Este contexto, además de entretenido, generaría una sensación de bienestar, pues supera la tensión y aburrimiento producidos por las pruebas tradicionales.

Basándonos en los antecedentes presentados, las preguntas que guiaron nuestra investigación fueron las siguientes:

1. ¿Es posible diseñar pruebas de evaluación invisible (PEI) a través del juego en los dominios de cálculo, lectura e inteligencia?
2. ¿Es la evaluación invisible basada en juego válida concurrentemente con la evaluación tradicional?
3. ¿Es la evaluación invisible basada en juego similar a la evaluación tradicional al predecir el criterio de adaptación escolar?
4. ¿Es la evaluación basada en juego menos percibida como situación de "evaluación" que la evaluación tradicional?
5. ¿Generan las PEI un efecto de facilitación de la tarea en comparación a las pruebas tradicionales (PET)?

## Método

Nuestro estudio fue cuasi-experimental y transversal.

### Participantes

En el estudio participaron voluntariamente 337 niños y niñas distribuidos en una muestra aleatoria estratificada por género y grado escolar, desde kínder hasta tercero básico (ver Tabla 1). Los niños y niñas atendieron indistintamente a PEI y PET en algunas de las tres habilidades de interés (cálculo, lectura o inteligencia). La muestra total fue dividida en tres grupos de manera aleatoria en cada uno de los dominios cognitivos que serían evaluados. De este modo, un tercio de los participantes realizó las pruebas de inteligencia, otro tercio las de matemáticas y otro las de lenguaje. Todos los participantes son alumnos de tres colegios particulares subvencionados de la Región Metropolitana de Santiago de Chile, esto es, de nivel socio-económico medio. Los colegios fueron seleccionados por conveniencia.

Tabla 1
*Distribución de los Participantes del Estudio, Según Sexo y Grado Escolar*

| Género Grado | Kínder | Primero | Segundo | Tercero | Total |
|---|---|---|---|---|---|
| Femenino | 41 | 40 | 42 | 45 | 168 |
| Masculino | 41 | 45 | 44 | 39 | 169 |
| Total | 82 | 85 | 86 | 84 | 337 |

### Instrumentos

**Cuestionario de evaluación cualitativa de las pruebas.** Diseñamos este cuestionario para esta experiencia. Lo usamos para preguntar a los niños si consideraban que las evaluaciones PET y PEI eran similares o distintas, si alguna se parecía más a las pruebas que rendían en el colegio, cuál de las pruebas consideraban más entretenida, en cuál pensaban que les había ido mejor, si les habían gustado las actividades y cuál de ellas les había gustado más. Además de lo anterior, preguntamos también sobre la experiencia previa en tablets, ya que las pruebas de evaluación invisible utilizaban este soporte.

**Pruebas de evaluación invisible.**

***Prueba de evaluación invisible-cálculo (PEI-C).*** Es un juego cuyo objetivo es explorar el rendimiento en tareas de conteo. En estudios previos se han identificado como precursores del aprendizaje de las matemáticas los desempeños en tareas relacionadas con el componente viso-espacial de la memoria de trabajo y en tareas de conteo y numeración (Passolunghi, Vercelloni & Schadee, 2007). A través de una metáfora lúdica y atractiva, diseñamos e implementamos un juego que evalúa la velocidad de reconocimiento de secuencias numéricas. En esta tarea el evaluado debe observar una grilla que tiene desde 9 hasta 25 números y presionarlos en orden ascendente. Esta grilla funciona como una cerradura de una jaula en la que está atrapado un monstruo que debe ser liberado por el jugador. Independientemente del desempeño del sujeto en esta prueba, al completar todos los números que aparecen en la grilla, el evaluado libera al monstruo que está encerrado en su interior (ver Figura 1a).

Se consideró como criterios de dificultad la cantidad de estímulos en cada grilla y la distancia entre los números que componen la secuencia. De un total de seis grillas, las dos primeras contenían nueve números, las dos siguientes 16 y las últimas 25. Además de lo anterior, en las grillas de mayor dificultad había una mayor distancia entre los números consecutivos. El indicador general de esta prueba se obtiene a partir del siguiente procedimiento: a los sujetos que realizan correctamente la tarea se les considera el tiempo de ejecución total; a los que no la realizan correctamente se les asigna el tiempo del sujeto que tuvo el mayor tiempo de ejecución en la tarea. Estos tiempos se estandarizan para cada uno de los seis niveles de la prueba y luego se estandarizan nuevamente tomando en consideración el total de los niveles. Se obtuvo un 0,90 de evidencia de confiabilidad, a través del método de comparación de mitades.

***Prueba de evaluación invisible-lectura (PEI-L).*** En esta actividad la base teórica es que las habilidades básicas de procesamiento auditivo están relacionadas con las habilidades en música y lectura, ya que ambos aprendizajes requieren del reconocimiento de los sonidos y de su estructura. En específico, las investigaciones apoyan una asociación directa entre discriminación tonal y conciencia fonológica (Anvari et al., 2002; Benasich & Tallal, 2002; Bolduc & Montésinos-Gelet, 2005; Forgeard et al., 2008; Moreno et al., 2011). En línea con lo anterior, desarrollamos un juego en el que el sujeto observa a un gato sobre una tina de baño y escucha dos sonidos de notas musicales. La tarea del sujeto consiste en identificar si las notas son iguales o distintas. Si contesta correctamente, el gato cae al agua y se limpia; si contesta incorrectamente, el gato se escapa del baño (ver Figura 1b). Se trata, entonces, esencialmente, de una prueba de conciencia fonológica.

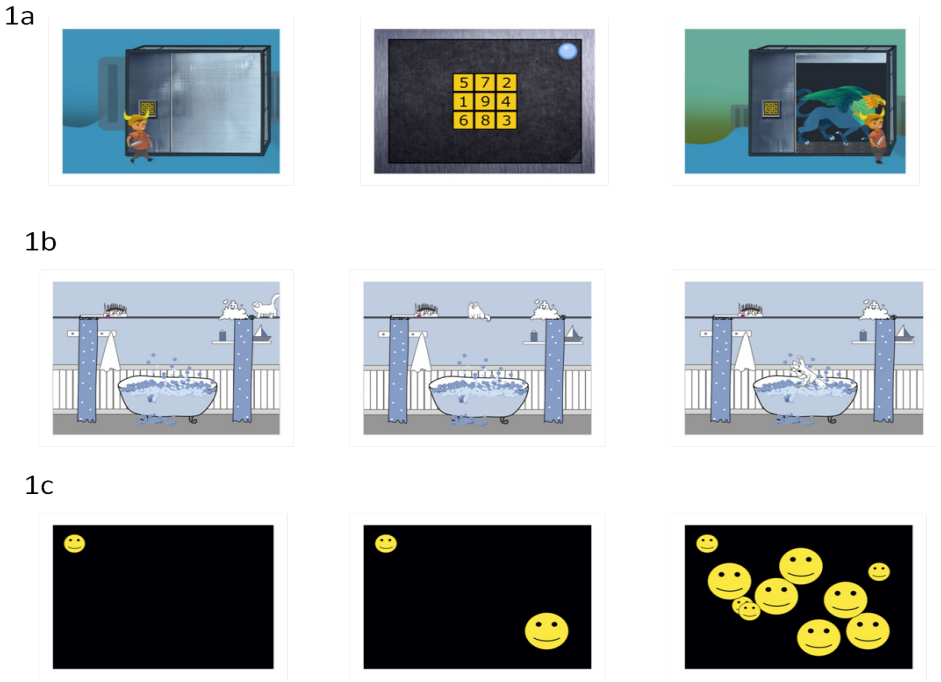


*Figura 1.* Interfaces de las Pruebas de Evaluación Invisible: (a) prueba de evaluación invisible-cálculo, (b) prueba de evaluación invisible-lectura, (c) prueba de evaluación invisible-inteligencia.

La gradiente de dificultad de este juego se determinó a partir de la distancia de las notas. En los estímulos más sencillos la distancia entre las notas era mayor, por lo que resultaba más fácil identificar los sonidos como distintos; en los ítems de mayor dificultad, la distancia entre las notas era menor, por lo que resultaba más difícil identificar si es que había una diferencia de tono entre ambas. El indicador de esta prueba se obtiene a partir del número de aciertos del sujeto en la prueba. Como evidencia de confiabilidad, los resultados indicaron una consistencia interna, estimada con alfa de Cronbach, de 0,70.

***Prueba de evaluación invisible-inteligencia (PEI-I).*** Para esta actividad consideramos aspectos relacionados con la amplitud del procesamiento de información. Este juego también está apoyado por estudios previos que han identificado una relación entre la memoria de trabajo y la inteligencia general o factor *g* (Colom, Abad, Quiroga, Shih & Flores-Mendoza, 2008; Engle, Tuholski, Laughlin & Conway, 1999). En este juego el evaluado observa una pantalla donde aparecen estímulos con forma de cara sonriente (ver Figura 1c). La tarea consiste en identificar el estímulo nuevo. La prueba consta de tres intentos con 15 estímulos cada uno. El orden de aparición de los estímulos es el mismo para cada uno de los tres intentos. La dificultad en esta prueba está dada por la distancia entre el nuevo estímulo y los anteriores (un ítem es más fácil cuando hay una mayor distancia entre los estímulos y más difícil cuando la distancia es menor) y por la diferencia de tamaño (los ítems que tienen un tamaño diferente al ítem previo se identifican más fácilmente que los que son similares).

El indicador de esta prueba se obtuvo considerando el número de aciertos de los intentos 2 y 3, debido a que el primero tiene el objetivo de servir como ítem de práctica. Solo se consideraron los casos en que la diferencia de aciertos entre los intentos 2 y 3 fuera igual o mayor a 0. De estos intentos, se consideró aquel que tuviera el mejor puntaje. Como evidencia de confiabilidad, el análisis de consistencia interna, por medio de alfa de Cronbach, arrojó un valor de 0,75.

**Pruebas de evaluación tradicional.**

***Prueba de evaluación tradicional-cálculo (PET-C).*** Para evaluar cálculo aplicamos algunas subpruebas de aprovechamiento de la Batería III Woodcock-Muñoz (Muñoz-Sandoval, Woodcock, McGrew & Mather, 2005): (a) cálculo, que evalúa la habilidad del sujeto para realizar cálculos matemáticos; (b) fluidez en matemáticas, que evalúa la habilidad del sujeto para resolver ejercicios simples de suma, resta y multiplicación de manera rápida; (c) problemas matemáticos, que evalúa la habilidad de analizar y resolver problemas matemáticos y (d) conceptos cuantitativos, que evalúa el conocimiento de símbolos, vocabularios y conceptos matemáticos.

***Prueba de evaluación tradicional-lectura (PET-L).*** Para evaluar lenguaje aplicamos algunas subpruebas de aprovechamiento de la Batería III Woodcock-Muñoz (Muñoz-Sandoval et al., 2005): (a) identificación de letras y palabras, que evalúa la habilidad del sujeto para identificar letras y palabras; (b) comprensión de textos, que evalúa la habilidad del sujeto para entender lo que está leyendo durante el proceso de lectura; (c) análisis de palabras, que evalúa la habilidad del sujeto para el análisis fónico y estructural y (d) vocabulario sobre dibujos, que evalúa el desarrollo de lenguaje y el conocimiento de palabras.

***Prueba de evaluación tradicional-inteligencia (PET-I).*** Para evaluar inteligencia aplicamos las subpruebas Construcción con Cubos y Analogías de la versión chilena de WISC-III. Estas subpruebas son las que tienen la mayor correlación de las escalas verbales y ejecutivas con el indicador global de la batería en los rangos etarios considerados para la investigación (Ramírez & Rosas, 2007).

## Procedimiento

Todos los procedimientos que aplicamos fueron aprobados por el Comité de Ética de la Escuela de Psicología de la Pontificia Universidad Católica de Chile y por la comisión del Fondo que financió el proyecto.

El primer paso fue contactar a las escuelas a fin de presentar el proyecto de investigación a las autoridades del establecimiento. Posteriormente, firmamos un convenio que establecía las condiciones de colaboración entre el equipo de investigación y los profesionales de las escuelas. El equipo de investigación se comprometió a aplicar y corregir todas las pruebas y realizar el análisis posterior de datos. Además, el equipo investigador se comprometió a realizar una charla con temas de interés para los profesionales de cada establecimiento, luego de recoger los datos. Las escuelas se encargaron de contactar a los padres, informarlos sobre la investigación y hacerles llegar el consentimiento informado que debían firmar para participar en el proyecto. Los apoderados de los 337 niños y niñas participantes firmaron el consentimiento, autorizando la participación de su hijo o hija en el estudio. Adicionalmente al consentimiento informado, pedimos un asentimiento a cada niño y niña.

En cada escuela solicitamos los promedios de notas finales de los estudiantes en las asignaturas de lenguaje, matemáticas y promedio total de notas.

Se consideró una aplicación contrabalanceada entre las PEI y las PET. Luego de realizar estas pruebas, aplicamos el cuestionario de evaluación cualitativa.

## Análisis de Datos

Para realizar las comparaciones entre las PET y las PEI construimos un indicador único para cada batería. Para construir este indicador convertimos a puntajes Z los puntajes brutos obtenidos en cada una de las subpruebas de la batería. Posteriormente, reconvertimos estos puntajes Z a nivel de subprueba a puntajes Z a nivel de batería, con lo que obtuvimos un indicador único para cada una de las PET.

Los promedios de notas de los alumnos fueron comparados con los rendimientos en las PET y PEI por medio de correlaciones de Pearson para obtener evidencia de validez concurrente entre las pruebas. Los tiempos de aplicación de las PET y las PEI fueron comparados con un análisis de varianza de mediciones repetidas.

A fin de conseguir evidencia de confiabilidad para las PEI, realizamos un análisis de consistencia interna con coeficiente alfa de Cronbach para las actividades de lectura e inteligencia y prueba de comparación de mitades para la actividad de matemática. Para obtener evidencia de validez, realizamos análisis descriptivos y pruebas de correlación para cada una de las PEI con sus respectivas PET. Los datos fueron analizados con el programa SPSS 18.

## Resultados

Presentaremos los resultados siguiendo el mismo orden de las preguntas que guiaron esta investigación.

### ¿Es Posible Diseñar Pruebas de Evaluación Invisible a Través del Juego en los Dominios de Cálculo, Lectura e Inteligencia?

Por la descripción entregada de los instrumentos y por las propiedades psicométricas de los mismos, la respuesta parece ser afirmativa. En la Tabla 2 aparecen los estadísticos descriptivos por curso en las tres pruebas construidas. El incremento de valores con la edad puede considerarse un buen indicador de evidencia de validez de la prueba, ya que se trata de pruebas cognitivas cuya complejidad disminuye con la edad. Los índices de confiabilidad son muy buenos para la prueba PEI-C y aceptables para la pruebas PEI-I y PEI-L.

Tabla 2
*Estadísticos Descriptivos (Promedio y Desviación Estándar) y Confiabilidad de las PEI por Grado Escolar*

| Grado | Kínder | Primero | Segundo | Tercero | Confiabilidad |
|---|---|---|---|---|---|
| Prueba | *M* (*DE*) | *M* (*DE*) | *M* (*DE*) | *M* (*DE*) | |
| PEI-C | -0,97 (0,75) | -0,16 (0,86) | 0,34 (0,77) | 0,85 (0,65) | 0,90 |
| PEI-L | 7,81 (2,68) | 8,88 (2,44) | 10,32 (2,65) | 11,19 (3,05) | 0,70 |
| PEI-I | 9,12 (4,00) | 11,81 (6,18) | 13,50 (7,02) | 15,48 (8,25) | 0,75 |

*Notas.* PEI-C = PEI cálculo; PEI-L = PEI lectura; PEI-I = PEI inteligencia.
Los valores en las pruebas están expresados en escalas que no son comparables inter-pruebas.

### ¿Es la Evaluación Invisible Basada en Juego Válida Concurrentemente con la Evaluación Tradicional?

En la Tabla 3 presentamos las correlaciones lineales de Pearson entre las PEI y las respectivas PET. Los resultados de este análisis indican que existen correlaciones entre ellas.

Tabla 3
*Correlaciones Entre las PEI, PET y Promedio de Notas*

| | PEI-C | PEI-L | PEI-I | PET-C | PET-L | PET-I |
|---|---|---|---|---|---|---|
| PET-C | 0,72** | | | | | |
| PET-L | | 0,37** | | | | |
| PET-I | | | 0,52** | | | |
| Notas matemáticas | 0,14 | | | 0,33* | | |
| Notas lenguaje | | 0,17 | | | 0,49** | |
| Promedio general | | | 0,16* | | | 0,33* |

*Nota.* PEI-C = PEI cálculo; PEI-L = PEI lectura; PEI-I = PEI inteligencia; PET-C = PET cálculo; PET-L = PET lectura; PET-I = PET inteligencia.
* $p < 0,05$, ** $p < 0,01$

**¿Es la Evaluación Invisible Basada en Juego Similar que La Evaluación Tradicional, al Predecir el Criterio de Adaptación Escolar?**

No exactamente. Si todas las correlaciones son directas, las PET tienen mejor valor predictivo que las PEI sobre las notas escolares, ya que todas las correlaciones entre PET y notas son significativas. En el caso de las PEI, solo la evaluación de inteligencia mostró una correlación significativa con las notas escolares. En la Tabla 3 se muestran las correlaciones de las PEI y PET con los promedios finales de matemáticas, lenguaje y general, como criterios para las variable de cálculo, lectura y habilidad cognitiva, respectivamente.

**¿Es la Evaluación Basada en Juego Percibida Como Situación de Evaluación Tradicional?**

Los resultados muestran que las PEI no son percibidas como evaluación tradicional, mientras que las PET son percibidas como parecidas a las pruebas escolares. Formulamos dos preguntas a los niños para comparar los dos tipos de pruebas. En la primera de ellas les preguntamos: "¿Crees que las dos actividades son parecidas o diferentes?". La mayoría de los niños opinó que las pruebas son diferentes (87,3%). A continuación les preguntamos: "¿Cuál de estas dos actividades se parece más a las pruebas que te hacen en el colegio?". Casi la totalidad de los niños (96,4%) opinó que la prueba tradicional es más parecida a las pruebas escolares tradicionales. Al preguntar a los niños sobre cuál tipo de actividades les gusta más, el 84% de los niños se inclinó por las actividades de las PEI y más del 86% las encontró más entretenidas que las PET. Los resultados anteriores permiten concluir que las PEI son efectivamente percibidas como diferentes y menos parecidas a una evaluación tradicional que las PET.

En la Tabla 4 mostramos los resultados de los tiempos totales de aplicación de las PET y las PEI. Los tiempos de las PEI son significativamente inferiores a los de las PET, lo que probablemente explique en parte el hecho de que sean consideradas como actividades no propias del contexto escolar.

Tabla 4
*Tiempos Totales de Aplicación en Minutos de las PET y PEI*

| Dominio | PET *M* (*DE*) | 95% IC PET | PEI *M* (*DE*) | 95% IC PEI | *F* | *p* | $\eta^2$ |
|---|---|---|---|---|---|---|---|
| Lectura | 18,2 (5,8) | [17,0, 19,4] | 9,5 (3,1) | [8,9, 10,1] | $F(1, 93) = 191{,}09$ | < 0,001 | 0,47 |
| Cálculo | 26,7 (7,9) | [24,2, 29,0] | 7,7 (4,8) | [6,3, 9,1] | $F(1, 42) = 161{,}54$ | < 0,001 | 0,70 |
| Inteligencia | 2,0* | | 0,2 (0,3) | | | | |

* Este es un valor estimado para la prueba de Construcción con Cubos que considera los tiempos totales de los tres primeros ítems, que son los mínimos necesarios para cumplir el criterio de suspensión. La aplicación de la PEI toma 1/6 del tiempo de aplicación de la PET. En este cálculo no se incluyó el tiempo de aplicación de Analogías. En caso de agregar el tiempo, la diferencia entre ambas pruebas sería mayor.

### ¿Generan las Pruebas de Evaluación Invisible un Efecto de Facilitación de la Tarea en Relación a las Pruebas Tradicionales?

Como esperamos demostrar, creemos que sí. Para responder esta pregunta dividimos a los niños en dos grupos: los de alto rendimiento (en el percentil 75 o superior) y los de bajo rendimiento (en el percentil 25 o inferior) en las notas de lectura, matemáticas y promedio general al fin del año escolar. Para comparar los rendimientos en las PET y PEI de los grupos de alto y bajo rendimiento estandarizamos los puntajes obtenidos en las pruebas determinadas (inteligencia, cálculo y lectura). Luego, comparamos los promedios estandarizados obtenidos en las PET y las PEI. En el dominio de lectura encontramos un efecto principal del nivel de rendimiento, $F(1, 53) = 16{,}46$, $p < 0{,}001$, $\eta_g^2 = 0{,}14$, y una interacción entre el nivel de rendimiento y el tipo de prueba en el dominio de lectura, $F(1, 53) = 4{,}16$, $p = 0{,}046$, $\eta_g^2 = 0{,}03$ (ver Tablas 5 y 6 para el detalle de las comparaciones en lectura).

Tabla 5
*Puntajes Estandarizados por Nivel de Rendimiento en Lectura*

| Nivel de rendimiento | *M* | *DE* | 95% IC |
|---|---|---|---|
| Alto | 0,30 | 0,56 | [0,02, 0,58] |
| Bajo | -0,51 | 0,89 | [-0,79, -0,23] |

Tabla 6
Puntajes Estandarizados por Nivel de Rendimiento y Tipo de Evaluación en Lectura

| Nivel de rendimiento | *Tipo de evaluación* | *M* | *DE* | 95% IC |
|---|---|---|---|---|
| Alto | PET | 0,47 | 0,57 | [0,10, 0,83] |
| Alto | PEI | 0,13 | 1,02 | [-0,24, 0,49] |
| Bajo | PET | -0,71 | 1,27 | [-1,07, -0,34] |
| Bajo | PEI | -0,31 | 1,04 | [-0,67, 0,06] |

Como puede apreciarse en la Figura 2, en todas las evaluaciones los niños de bajo rendimiento obtienen, comparativamente, mejores resultados en las PEI que en las PET y, al contrario, los niños de alto rendimiento obtienen comparativamente mejores resultados en las PET que en las PEI, a pesar de que en promedio los niños de alto rendimiento obtienen siempre mejores resultados tanto en las PEI como en las PET.

## Discusión

### ¿Qué Debemos Entender Bajo Evaluación Invisible?

Nuestra propuesta de evaluación invisible es conceptualmente más parecida a la de Shute (2011), pero en contenidos, más parecida a la de Verhaegh et al. (2013), porque, si bien el propósito de nuestras evaluaciones es que el sujeto evaluado no sienta que lo está siendo, el objetivo central de nuestra investigación fue entregar evidencia de que sí es posible desarrollar instrumentos de evaluación invisible, con base en criterios psicométricos aceptados por la comunidad de psicólogos. Este último punto implica necesariamente someter a los instrumentos a los requerimientos estandarizados de calidad psicométrica establecidos en los estándares de AERA et al. (2002).

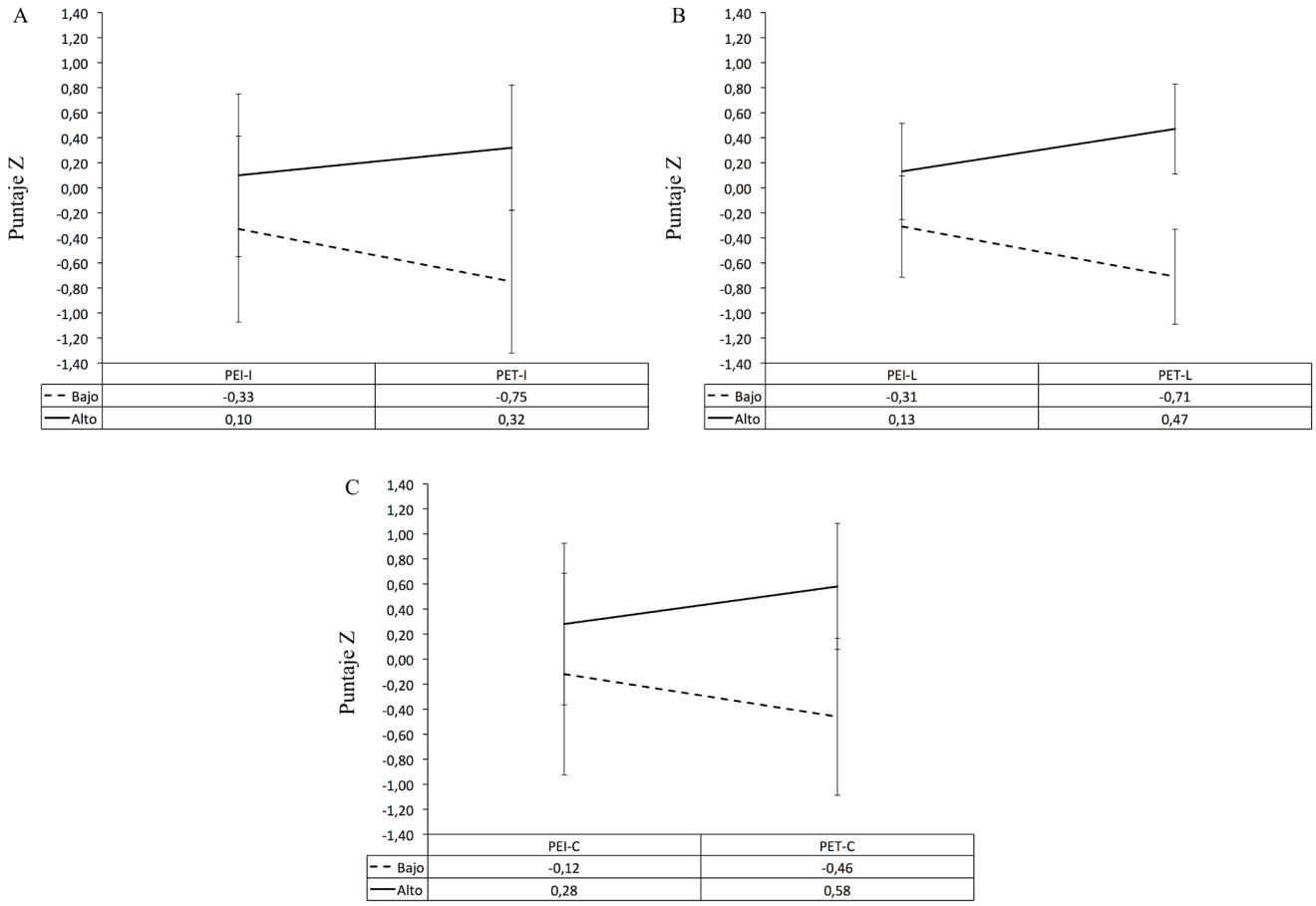


*Figura 2.* Comparación de medias estandarizadas obtenidas entre las PEI y las PET: A = inteligencia, B = lectura y C = cálculo. El eje horizontal indica los puntajes Z obtenidos por cada grupo y el eje vertical, el tipo de prueba utilizada.

## ¿Cómo Conciliamos Ambos Requerimientos?

En principio, parece una tarea nada fácil, ya que el concepto de *invisibilidad* de la evaluación impide preguntar de manera explícita lo que se supone que se está evaluando. La solución que dan Verhaegh et al. (2013) es presentar en forma de juego tareas que son casi idénticas a las tareas del WISC, Esto es, que dan un contexto lúdico a una actividad que cognitivamente es idéntica a la de su análogo presentado en contexto no lúdico. Aquí, el juego sirve de *contexto facilitador*, pero el contenido de la evaluación no es estrictamente invisible. Esto es análogo a la manera como otros autores han contextualizado aprendizajes en contextos lúdicos (Rosas et al., 2003), en los cuales, por ejemplo, se presenta una tarea de decodificación lectora en forma de juego, construyendo un puente de letras para liberar a una princesa de un castillo. La tarea indudablemente es más entretenida para los niños, pero sigue siendo una tarea de decodificación lectora.

## ¿Y Cómo Conciliar Ambos Mundos?

La estrategia que escogimos para optimizar los dos requerimientos que parecen contrapuestos fue usar juegos en los que se evalúan precursores de los constructos evaluados (como en cálculo e inteligencia) o constructos que comparten habilidades de procesamiento con la habilidad que se pretende evaluar (como en lectura). Como proponemos, los precursores son destrezas previas automatizadas que permiten evaluar los constructos de interés desde indicadores *invisibles* o diferentes a las destrezas. En el segundo caso, consideramos que, al conocer que el rendimiento en dos habilidades distintas está relacionado porque comparten procesamientos o dominios en su base, será posible extrapolar el rendimiento en un tipo de

habilidad a partir del rendimiento en la otra. Así, volviendo a nuestro ejemplo anterior, al conocer el valor predictivo de las destrezas de discriminación tonal sobre las destrezas de decodificación lectora, una forma de evaluar la consolidación de estas últimas es conocer el grado de automatización de las primeras. Los resultados parecen apoyar esta estrategia: las habilidades evaluadas en los constructos de lectura, cálculo e inteligencia muestran índices psicométricos que permiten estar optimistas de que esta es una alternativa promisoria para la evaluación de constructos psicológicos complejos de una manera rápida y confiable.

Sin embargo, al comparar la correlación de las medidas de evaluación invisible con los criterios de adaptación escolar, resulta evidente que es menor que la obtenida con medidas tradicionales. Este es un resultado que debe ser explorado en mayor profundidad en futuros estudios, ya que es posible que obedezca a dos posibilidades que es preciso clarificar: (a) que sea más difícil predecir adaptación escolar a partir de cualquier medida basada en precursores o habilidades compartidas o (b) que la evaluación invisible sea menos efectiva, porque las evaluaciones tradicionales son más parecidas en formato y contenido a las medidas de adaptación escolar.

### ¿Cuál Es la Diferencia Subjetiva Entre la Evaluación Tradicional y la Invisible?

Las diferencias objetivas entre la evaluación tradicional y la invisible han sido reportadas por varios autores (Shute, 2011; Shute, Ventura, Bauer & Zapata-Rivera, 2009). Las diferencias radican en el tipo de esquemas que activan, los contenidos de la evaluación, la actividad implicada y la forma de responder. El presente estudio entrega, además, algunas pistas acerca de las propiedades subjetivas de este tipo de evaluación, entre las que destaca el ser percibida por los niños como diferente a las evaluación a los que están acostumbrados en la escuela, en el sentido de ser menos parecida a estas y preferirlas en términos afectivos. Estos datos solo vienen a confirmar algo que se da por sabido: que las pruebas en formato tradicional no son del gusto de los niños. ¿Y por qué las evaluaciones tradicionales resultan menos atractivas para los chicos? Probablemente porque les anticipan potenciales consecuencias aversivas: malas notas, castigos, bullying y una larga lista de etcéteras, que les genera ansiedad y estrés al ser sometidos a la evaluación (Eum & Rice, 2011).

Es preciso destacar, por último, una consideración ética importante: los niños son informados de manera clara y explícita que serán evaluados, tanto en la condición de PET como de PEI. Sin embargo, en esta última condición se olvidan rápidamente que se trata de una evaluación.

### ¿Y Tiene Alguna Consecuencia Importante que las PEI Generen Menos Ansiedad que las PET Sobre el Rendimiento de los Niños?

Como esperamos haber demostrado, nuestros resultados muestran una interacción importante entre el nivel general de rendimiento de los estudiantes y el rendimiento en las PET y PEI. Así, los estudiantes pertenecientes al percentil 75 y superior de rendimiento general muestran un rendimiento relativo superior en las PET que en las PEI, lo que demuestra que su rendimiento no está afectado negativamente por la evaluación explícita de los contenidos. Sin embargo, los estudiantes pertenecientes al percentil 25 e inferior en rendimiento general muestran el patrón exactamente inverso: les va mucho mejor en las PEI que en las PET, lo que muestra que, al ser confrontados a una evaluación tradicional, los niños de conocido bajo rendimiento rinden más bajo que su máximo potencial. Esto es coincidente con la literatura que demuestra los efectos de la ansiedad sobre la evaluación (Elliot & Pekrun, 2007; Putwain & Symes, 2012, Urhahne et al., 2011), pero agrega un dato esencial: pareciera que las PEI, por el hecho de ser invisibles, impiden que la ansiedad de los estudiantes interfiera con los resultados de su evaluación, logrando de esta manera, en palabras de Vigotsky, ampliar su zona de desarrollo próximo, al alcanzar un diagnóstico más acertado de su máximo potencial de rendimiento (Vigotsky, 1978).

En otras palabras, lo que nos muestran los resultados es que la evaluación tradicional subestima el rendimiento de estudiantes de bajo rendimiento general y estima correctamente el de estudiantes de alto rendimiento. Dado el interés teórico y práctico que revisten estos resultados y atendiendo a que en el caso de nuestro estudio está establecido a partir de un número restringido de casos —y, por lo tanto, con baja potencia estadística—, es que sugerimos profundizar y replicar estudios que permitan ampliar la evidencia que apoya esta línea argumental. Los resultados sugieren preliminarmente que las PEI deben ser investigadas como alternativas para el tamizaje de diagnóstico de habilidades para niños con un historial de experiencias negativas en las PET. Y, una vez establecido este punto, es preciso indagar más en detalle si las

PEI, en formato de evaluación formal, no terminan siendo igualmente ansiógenas, al conocerse su naturaleza evaluativa. Estas son todas cuestiones que no es posible zanjar de manera conclusiva con los pocos datos disponibles.

Como limitación de este estudio es importante señalar el efecto que puede tener en los resultados el hecho de que las PET fueran presentadas en formato de lápiz y papel, mientras que las PEI lo fueran en formato tecnológico. Consideramos que el uso de tecnologías puede favorecer el interés de los niños por los instrumentos utilizados, pero que no explica por sí solo su percepción de las tareas como entretenidas. Sabemos que existen PET diseñadas para aplicación en plataforma tecnológica, como las pruebas de evaluación de la atención sostenida bajo el modelo del Test de Ejecución Continua (Conners, 1994) que, a pesar de ser presentadas en computador, son tareas tediosas y monótonas que precisamente buscan forzar el sostenimiento atencional. Es importante que se desarrollen estudios que comparen las PET y PEI en el mismo formato de presentación, para determinar si los factores motivacionales y de interés de los niños frente a un tipo de evaluación sobre otra se mantienen, a pesar del soporte de los instrumentos.

## Referencias


Alliende, F., Condemarín, M. & Milicic, N. (2000). *Abriendo mundos*. Santiago, Chile: Editorial Universitaria.

American Educational Research Association, American Psychological Association & National Council in Measurement in Education (2002). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anvari, S. H., Trainor, L. J., Woodside, J. & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology, 83,* 111-130. doi:10.1016/S0022-0965(02)00124-8

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science, 1,* 164-180. doi:10.1111/j.1745-6916.2006.00011.x

Beck, A., Emery, G. & Greenberg, R. L. (1985). *Anxiety disorders and phobias: A cognitive perspective*. New York, NY: Basic Books.

Benasich, A. A. & Tallal, P. (2002). Infant discrimination of rapid auditory cues predict later language impairment. *Behavioral Brain Research, 136,* 31-49. doi:10.1016/S0166-4328(02)00098-0

Bolduc, J. & Montésinos-Gelet, I. (2005). Pitch processing and phonological awareness. *Psychomusicology: Music, Mind & Brain, 19,* 3-14. doi:10.1037/h0094043

Colom, R., Abad, F. J., Quiroga, M. A., Shih, P. C. & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related construct, but why? *Intelligence, 36,* 584-606. doi:10.1016/j.intell.2008.01.002

Conners, C. K. (1994). *Conners Continuous Performance Test*. Toronto, Canadá: Multi-Health Systems.

Desrochers, M. N., Pusateri Jr., M. J. & Fink, H. C. (2007). Game assessment: Fun as well as effective. *Assessment & Evaluation in Higher Education, 32,* 527-539. doi:10.1080/02602930601116789

Diamond, A. & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science, 333,* 959-964. doi:10.1126/science.1204529

Elliot, A. J. & Pekrun, R. (2007). Emotion in the hierarchical model of approach-avoidance achievement motivation. En P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 57-74). Burlington, MA: Elsevier.

Engle, R. W., Tuholski, S. W., Laughlin, J. E. & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128,* 309-331. doi:10.1037/0096-3445.128.3.309

Eum, K. & Rice, K. G. (2011). Test anxiety, perfectionism, goal orientation, and academic performance. *Anxiety, Stress, & Coping, 24,* 167-178. doi:10.1080/10615806.2010.488723

Eysenck, M. W., Derakshan, N., Santos, R. & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7,* 336-353. doi:10.1037/1528-3542.7.2.336

Farran, E. K. & Karmiloff-Smith, A. (Eds.) (2012). *Neurodevelopmental disorders across the lifespan: A neuroconstructivist approach*. New York, NY: Oxford University Press.

Forgeard, M., Schlaug, G., Norton, A., Rosam, C., Iyengar, U. & Winner, E. (2008). The relation between music and phonological processing in normal-reading children and children with dyslexia. *Music Perception, 25*(4), 383-390. doi:10.1525/mp.2008.25.4.383

Georgiou, G. K., Parrila, R. & Papadopoulos, T. C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology, 100,* 566-580. doi:10.1037/0022-0663.100.3.566

Keogh, E. & French, C. C. (2001). Test anxiety, evaluative stress, and susceptibility to distraction from threat. *European Journal of Personality, 15,* 123-141. doi:10.1002/per.400

McPherson, J. & Burns, N. R. (2007). Gs invaders: Assessing a computer game-like test of processing speed. *Behavior Research Methods, 39,* 876-883. doi:10.3758/BF03192982

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J. & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science, 22,* 1425-1433. doi:10.1177/0956797611416999

Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S. & Mather, N. (2005). *Batería III Woodcock-Muñoz: pruebas de aprovechamiento*. Rolling Meadows, IL: Riverside.

Nie, Y., Lau, S. & Liau, A. K. (2011). Role of academic self-efficacy in moderating the relation between task importance and test anxiety. *Learning and Individual Differences, 21,* 736-741. doi:10.1016/j.lindif.2011.09.005

Passolunghi, M. C., Vercelloni, B. & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability and numerical competence. *Cognitive Development, 22,* 165-184. doi:10.1016/j.cogdev.2006.09.001

Putwain, D. W. & Symes, W. (2012). Achievement goals as mediators of the relationship between competence beliefs and test anxiety. *British Journal of Educational Psychology, 82,* 207-224. doi:10.1111/j.2044-8279.2011.02021.x

Ramírez, V. & Rosas, R. (2007). Estandarización del WISC-III en Chile: descripción del test, estructura factorial y consistencia interna

de las escalas. *Psykhe, 16*(1), 91-109. doi:10.4067/S0718-22282007000100025O

Rosas, R. & Bravo, T. (2009). Jugando con las letras: validación de un instrumento basado en computador para evaluar competencias lectoras iniciales. *Boletín de Investigación Educacional, 24*(1), 17-34.

Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M. Flores, P. … Salinas, M. (2003). Beyond Nintendo: Design and assessment of educational video games for first and second grade students. *Computers & Education, 40,* 71-94. doi:10.1016/S0360-1315(02)00099-4

Schank, R. C. & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Earlbaum.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. En S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503-524). Charlotte, NC: Information Age.

Shute, V. J., Ventura, M., Bauer, M. & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. En U. Ritterfeld, M. Cody & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295-321). New York, NY: Routledge.

Urhahne, D., Chao, S. -H., Florineth, M. L., Luttenberger, S. & Paechter, M. (2011). Academic self-concept, learning motivation, and test anxiety of the underestimated student. *British Journal of Educational Psychology, 81,* 161-177. doi:10.1348/000709910X504500

Verhaegh, J., Fontijn, W. F. J., Aarts, E. H. L. & Resing, W. C. M. (2013). In-game assessment and training of nonverbal cognitive skills using TagTiles. *Personal and Ubiquitous Computing, 17,* 1637-1646. doi: 10.1007/s00779-012-0527-0

Vigotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children — fourth edition (WISC-IV).* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale — fourth edition (WAIS-IV).* San Antonio, TX: Psychological Corporation.

Fecha de recepción: Marzo de 2014.
Fecha de aceptación: Octubre de 2014.