

Ensayo de Aplicación práctica, el Test Informatizado de Memoria Memopoc*

Practical Application of Memopoc informatized Memory Test: A test of the procedure

Nieves Schade¹
Pilar Hernández²
Bárbara Elgueta³

Resumen

El presente estudio tuvo como objetivo la implementación informatizada de un test convencional. Específicamente mediante un modelo estadístico descriptivo se realizó una aplicación piloto de una prueba para evaluar memoria: MEMOPOC, con el fin de determinar la aplicabilidad del instrumento en formato computacional. Se evaluaron dos aspectos: los técnicos y formales del instrumento. La muestra estuvo constituida por 36 niños, que cursaban de kinder a Cuarto Básico, de un colegio municipalizado de la comuna de Concepción. Los resultados indican que hay algunas falencias técnicas necesarias de corregir y a su vez muestra los beneficios que significó el cambio del test a un sistema computarizado. La discusión está en base a las proyecciones del test y los aportes de los test informatizados en la evaluación psicométrica.

Palabras clave: test informatizados, evaluación, memoria

Abstract

The present study attempted to implement a computerized version of a conventional test. Specifically, through a descriptive statistical model, a pilot application of the test MEMOPOC for evaluating memory was performed in order to determine the applicability of a computerized format. Two aspects were evaluated: the technical and the formal. The sample was constituted by 36 children, from kindergarten to fourth grade elementary, of a municipal school in the district of Concepción. The results indicate that there are few technical errors to correct and also showed the benefits that meant changing to a computerized system. The discussion is based on the projections of the test and benefits of computerized tests in psychometric evaluation.

Key words: computerized test, evaluating, memory.

* El presente trabajo es producto de un proyecto de Investigación financiado por la Dirección de Investigación de la Universidad de Concepción.

¹ Dr. en Psicología Univ. Pont. de Salamanca, España y docente del departamento de Psicología, Universidad de Concepción. email: nschade@udec.cl

² Docente Departamento de Psicología, Facultad de Ciencias Sociales. Universidad de Concepción

³ Psicóloga Instituto Diego Portales. Carrera de Servicio Social

Introducción

En años recientes el uso de los computadores ha impactado en casi todos los ámbitos de la vida de las personas y la psicología no ha sido una excepción. Dentro de las áreas del quehacer del psicólogo, una en las que se ha ido incorporando es en la psicometría (Viswesvaran, 2000). Varias ventajas han sido atribuidas al uso de los computadores en la evaluación psicológica. Los tests informatizados, como son denominados actualmente, reportan ventajas como rapidez, uso de estímulos tridimensionales, formatos interactivos, control riguroso del tiempo de reacción y de presentación de los estímulos, no ocupan espacio, el resultado es inmediato, reducen el tiempo de dedicación de los profesionales a esta tarea. (Viswsvaran, 2000; Huff y Sirece, 2001).

Además, el manejo adecuado de la computación genera una mayor seguridad por lo que aumenta la estabilidad, fiabilidad y validez de las mediciones. Al ser las instrucciones y los medios de presentación estandarizadas esto hace que sean más ventajosos para la evaluación de determinados pacientes. Por ejemplo, los estudios muestran que los tests cognitivos basados en la computación ofrecen un 60% de ahorro en tiempo que la administración tradicional con población de pacientes psiquiátricos (Schatz y Brown dyke, 2002)

En general, estos avances tecnológicos hacen cada vez más accesible la elaboración de pruebas con un nivel máximo de automatización, al respecto Muñiz y Hambleton (1999) creen que "el uso del computador en psicometría permitirá utilizar ítems más complejos y cercanos a la realidad, que incluso se puede simular, lo que conllevaría posibles mejoras en la validez predictiva, al acercarse más el test a la realidad criterial que se pretende predecir" (pp. 24).

A pesar de las ventajas antes señaladas, los psicólogos han sido relativamente lentos

en embarcarse en la aplicación de las tecnologías en su propio trabajo (Schatz y Brown dyke, 2002), aunque podemos encontrar avances importantes en los países más desarrollados como España, donde se aprecia una autentica revolución en cuanto al diseño y aplicación de los test psicológicos y educacionales a nivel informático. Incluso en algunas universidades existen grupos de investigadores que estudian aspectos específicos como construcción y gestión de bancos de ítems, medidas cognitivas informatizadas de visualización espacial, control y tasa de exposición de los ítems, entre otros (Olea y Ponsada, 1998). En Estados Unidos también se aprecia un uso amplio de estos incluso cuentan con planes gubernamentales de fomentos de estas tecnologías en diferentes ámbitos, como en las fuerzas armadas y educación.

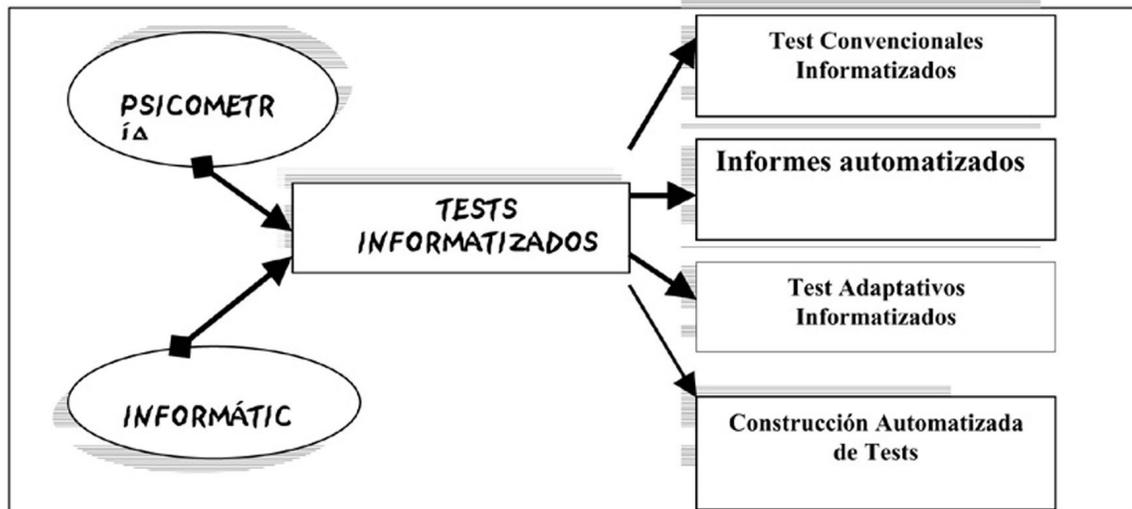
En nuestro país el panorama es muy distinto, de acuerdo a un estudio realizado por Concha (2003) acerca de la utilización de los tests informatizados en el quehacer del psicólogo, en el cual participaron 106 psicólogos, a quienes se les aplicó un cuestionario en dos versiones, una de papel- lápiz y otra vía Internet. Los resultados mostraron que estos profesionales usan muy poco los tests informatizados (80% nunca y el 3,8% siempre). Las razones que ellos plantean son: el desconocimiento, la carencia de recursos en cuanto equipamiento, infraestructura, software y falta de entrenamiento. Las pocas aplicaciones se aprecian más en el ámbito laboral y los tests informatizados más usados son el 16PF- 5, seguido por el MMPI; Lüscher y Raven respectivamente. Uno de los aspectos que merece mencionarse de este trabajo es que muchos psicólogos creen que el uso de éstos, es reduccionista y enfría la relación médico paciente, pero el cuestionamiento que ellos señalan, pone en evidencia su desconocimiento en el uso de los tests informatizados. Pero también, es importante señalar que un grupo presenta disposición a trabajar con ellos y ven como fortaleza la facilitación del trabajo como psicólogo, ahorro de tiempo, mayor

control en el proceso global y limitación en la existencia de errores.

Los tests informatizados (TICs), estrictamente hablando, son aquellos que utilizan el computador como medio de presentación de ítems, de entrada de las respuestas y de análisis e interpretación de los rendimientos (Olea, Ponsada y Prieto, 1999).

Según Muñiz y Hambleton, (1999), la aplicación de tecnología computacional en la evaluación psicométrica tiene cuatro ámbitos generales de influencia: a) Presentación Informatizada de test convencionales; b) Elaboración automatizada de informes; c) Test adaptativos informatizados y d) Construcción automatizada de los tests. Ver Figura 1.

Figura1 : Aplicaciones de la computación a la evaluación psicométrica.



A continuación se revisará someramente cada uno de estos por separado.

a) Presentación informatizada de Tests convencionales

Bunderson y otros (1989), afirman que la primera generación de tests informatizados correspondió a la implementación de test convencionales de papel y lápiz en un computador y la aplicación de los mismos a través del teclado y la pantalla.

Si bien esta forma de presentación de los tests clásicos no cambió nada de su esencia, el nuevo formato aportó toda una serie de ventajas instrumentales entre las cuales se destacan: presentación precisa y controlada de los ítems a través de la pantalla, corrección y archivo automatizado de las respuestas, eficiencia y rapidez a la hora de dar les

resultados. En Estados Unidos por ejemplo, los test convencionales informatizados más utilizados en la práctica clínica son: MMPI, MMPI-2, Rorschach, WAIS- R, WAIS III, WISC - R, WISC III.

b) Elaboración automatizada de informes

Esta aplicación de la informática a la evaluación psicométrica está muy relacionada con la anterior. Al obtenerse de modo automatizado las puntuaciones de las pruebas, es una extensión "natural" el que se pueda automatizar la interpretación de los datos y que se emita el informe correspondiente.

Muñiz y Hambleton (1999) plantean que en sí mismos, los informes generados por computador no tienen nada de malo, ya que "liberan al psicólogo de las tareas más automáticas, tediosas y burocráticas" (pp. 26),

pero añaden que esto no debe mal interpretarse con la insensatez que implica el que un computador sustituya plenamente la labor de diagnóstico del psicólogo. Un informe psicométrico obtenido automáticamente sólo constituye un buen punto de partida y complemento instrumental para el psicodiagnóstico, ya que es labor privativa del profesional la de contextualizar los datos y conformarlos como un todo coherente a la realidad clínica del evaluado. En 1986 la American Psychological Association (APA) dicta una serie de directrices éticas sobre el uso de esta clase de informes. En ella se hace hincapié en el hecho que los informes generados por computador deben utilizarse únicamente en conjunción con el juicio profesional.

c) Tests Adaptativos Informatizados

Los Tests Adaptativos Informatizados son el resultado de la aplicación computacional de un nuevo modelo de medición de aptitudes, es decir la teoría de la Respuesta al ítem.

El concepto clásico de la Teoría de Respuesta al Ítem es la Curva Característica del ítem, y corresponde a la función matemática que relaciona la probabilidad de acertar a un ítem con la capacidad de los sujetos. *"...la teoría clásica se centra en las propiedades de la puntuación global del test (x), mientras que la Teoría de Respuesta a los Ítems se centra en las particularidades de cada ítem(q)."* (Prieto y Delgado, 1999, pp. 208)

No es el objetivo de esta revisión el realizar una descripción acabada de la Teoría de Respuesta a los Ítems, sin embargo es necesario connotar la relación de este modelo con el desarrollo de los tests informatizados. A partir de esta teoría se generan pruebas psicológicas de medición de aptitudes, que en función del tipo de respuesta de cada individuo (estrategias de solución, tiempo de latencia de la respuesta, omisiones, tipos de errores más frecuentes, etc.), asumen disposiciones particulares para cada caso. Este procedimiento de auto selección de los ítems es

posible gracias a programas computacionales que gestionan Bancos de Ítems a partir de ciertas reglas preestablecidas.

En resumen, un Test Adaptativo Informatizado consiste en un banco de ítems con propiedades conocidas y un procedimiento para seleccionar los más apropiados para la habilidad de cada persona. El procedimiento de selección es un algoritmo informático que presenta ítems más difíciles o más fáciles según se acierten o se fallen los anteriores.

Su principal ventaja sobre los Tests convencionales es la eficiencia: permite obtener medidas más precisas con el mismo número de ítems o conseguir la misma precisión con casi la mitad de elementos (Wainer, 1990; Olea y Ponsada, 1996).

Lord 1980 (citado en Muñiz y Hontangas, 1999) agrega dentro de las ventajas de estas pruebas: a) mayor flexibilidad para que los examinados puedan acceder a la realización del test; b) posibilidades casi ilimitadas de incorporar nuevos tipos de ítems, incluyendo gráficos complejos, sonidos, videos, etcétera; c) resultados inmediatos de las pruebas y d) mejora de la seguridad de la prueba.

En general, este tipo de test ha alcanzado un desarrollo muy amplio (Chang y Ansley, 2003). Sin embargo McBride (1998), se refiere a tres problemas fundamentales que presentan esta clase de test. En primer lugar, los examinados no suelen tener oportunidad de modificar sus respuestas, una vez que estas hayan sido emitidas. En segundo lugar, investigaciones demuestran que es necesario mantener bancos de ítems muy numerosos para solventar la confiabilidad de las mediciones. En tercer lugar, Rocklin, O'Donnell y Holst (1995) creen que lo relacionado con la situación afectiva y motivacional derivada del éxito o fracaso en las respuestas es un factor interviniente de gran peso. Las personas difieren en sus reacciones ante la dificultad del test, ya sea por la orientación motivacional hacia el éxito o el fracaso, la

ansiedad que experimentan ante la situación de evaluación (Rocklin y O'Donnell, 1987; Vispoel, Herindrickson y Bleiler, 2000, Vispoel, Clough, Bleiler, Hendrickson y Ihrig, 2002) o el concepto que tienen de sí mismas (Vispoel, Rocklin y Wang, 1994). La suma de estos elementos conspira contra la validez de las medidas obtenidas. Rocklin, y O'Donnell y Holst (1995), se propusieron modificar el proceso de selección de los ítems al permitir a los sujetos seleccionar el grado de dificultad de los ítems para que así se pueda adaptar la situación de evaluación a su estado afectivo y emocional. Por otra parte, estimaron necesario presentar alguna clase de retroalimentación ante las respuestas para darle mayor sentido a la evaluación.

El resultado de estas propuestas es la aparición de los Test Autoadaptados Informatizados. Estos tests se describen como una variación de los test adaptados informáticos que permiten reducir la ansiedad y así mejorar la ejecución en la prueba. Estos difieren de los anteriores, en que el examinado puede elegir el nivel de dificultad durante la administración del test (Pitkin y Vispoel 2001). Otra dificultad que se ha visto hace alusión a uso de estos con algunos pacientes psiquiátricos quienes presentan algunos obstáculos como fobia frente al computador, inadecuada percepción con respecto a la utilidad de estos y otros presentan muchas reticencias a su uso (Schatz y Browndyke, 2002).

d) Construcción automatizadas de Tests

La construcción de los tests psicológicos se está beneficiando grandemente de la implementación de procedimientos informatizados. Una de las razones es que permite que la tediosa labor de analizar ítems y ensamblar pruebas se realiza casi exclusivamente mediante programas computacionales especializados en el análisis estadístico.

Existe un grupo de programas que conforman software especializados para la construcción de pruebas. Entre los más popula-

res encontramos MICROCAT, CONTEST, OTD, etc. Estos programas permiten el ensamblaje y optimización

Hasta aquí hemos revisado las cuatro aplicaciones de la computación a la evaluación. Un aspecto que merece mencionarse se relaciona con la estimación de las características psicométricas de confiabilidad y validez que debe cumplir cualquier test. Muchos investigadores para estimar la validez de un test basado en un computador hacen comparaciones entre la versión informatizada y la versión de papel y lápiz. Pero los tests basados en el computador presentan diferencia que se deben principalmente a este nuevo formato, por lo cual se sugiere evaluar la validez de constructo. Existen pocos estudios con respecto a la validez del constructo (el test representa el constructo por el cual esta siendo medido), existiendo la posibilidad que los test presenten una variación en las mediciones producto de un constructo irrelevante, este constructo irrelevante es el medio por el cual esta siendo presentado el test. Esto se relaciona con la familiaridad con las tecnologías, formato, diseño del test, etc. Por lo tanto, los autores señalan que la validez de constructo debe ser evaluada y medida y que el objetivo al desarrollar un test basado en la informática es aumentar la validez de constructo y disminuir la validez irrelevante (Kristen y Sireci, 2001).

Finalmente, es importante señalar que dada la necesidad de prevenir contra el uso inadecuado de los tests informatizados la American Psychological Association (APA) en 1986 publicó directrices al respecto basándose en las normas establecidas por el código de ética de los psicólogos de EE.UU. y en los estándares generales de aplicación de pruebas informatizadas de instituciones como la American Educational Research Association y el National Council of Measurement in Education.

De la revisión realizada se puede concluir que en Chile existen muy pocas aplicaciones

y usos de los tests informatizados y menos investigación en el tema, a pesar de los amplios beneficios reconocidos en la informatización de tests que estos reportan.

El presente estudio pretende contribuir en el desarrollo de este tema. El objetivo de la investigación fue la implementación informatizada de un test convencional. Específicamente mediante un modelo estadístico descriptivo realizar una aplicación piloto de la prueba para evaluar memoria: MEMOPOC, con el fin de determinar la aplicabilidad del instrumento en formato computacional. A este respecto se consideró evaluar dos niveles: Aspectos técnicos y formales del instrumento

Método

Muestra

1. SUJETOS

La muestra fue intencionada no probabilística estratificada por edad, curso y sexo, constituida por 36 niños, que cursaban de kinder a Cuarto Básico, con un rendimiento escolar promedio, todos pertenecientes a una escuela municipalizada de la comuna de San Pedro, Octava Región.

2. INSTRUMENTOS

Prueba de Memoria para Posiciones, Colores y Objetos (MEMOPOC): Prueba creada por Rehbein (1984) y estudiada sus características de fiabilidad y validez por Schade (1994). Test que mide memoria para posiciones, colores y objetos. El test está compuesto por un tablero magnético de 61 por 46 cm., sobre el cual se encuentran marcados 30 puntos negros en 5 filas. Se presentan estímulos consistentes en dos grupos de 15 círculos de maderas (fichas), con un imán en su superficie inferior. Constan de tres subpruebas: posiciones, colores y objetos. Las fichas de posiciones son todas de color rojo óxido, las de colores son todas de distintos colores y la de objetos varía en formas, carentes de un nom-

bre o identificación. El test presenta un adecuado nivel de fiabilidad ($K20= 0.80$), consistencia interna y los juicios de los expertos indican que es un test que mide lo que pretende medir, es decir, presenta un adecuada validez de contenido.

Este test es atractivo y muy amigable para los niños pero presenta como dificultad es que es muy pesado, es decir, difícil de manipular y trasladar, por lo cual se consideró que informatizarlo podría suplir estas dificultades.

Prueba de Memoria para Posiciones, Colores y Objetos (MEMOPOC, versión informatizada). Esta prueba esta basada en la anterior en cuanto al diseño de las actividades, los tiempos de presentación de los estímulos y lapsos que median entre ellas. Al sujeto se le presentan las tres tareas posibles las cuales pueden elegir por donde empezar, tiene actividades que le permiten ensayar y ver si ha comprendido las instrucciones. Luego comienza con la tarea elegida, aparece la primera ficha en el tablero y luego la segunda y el niño cada vez debe hacer un clic en la ficha que va apareciendo nueva, así sucesivamente hasta que se equivoca. Al final aparece el puntaje obtenido por el niño y si quiere volver hacer la prueba. Las posiciones de los estímulos en el tablero, cada secuencia son productos del azar, para evitar aprendizaje de las secuencias cada vez que hace el test.

Cuestionario: El cuestionario constaba de dos parte una que evaluaba los aspectos técnicos y otra los aspectos cualitativos de la prueba como son: la exigencia de la actividad de acuerdo a la edad y la motivación frente a la tarea. Las respuestas eran de tipo cerradas en escala Lickert y otras si y no.

3. PROCEDIMIENTO

Los niños fueron invitados en forma individual a la sala de computación del colegio, se les explicó someramente de lo que se trataba y se les solicitó su colaboración. Los niños debían ejecutar el test y luego responder el cuestionario.

Resultados

Los resultados que se presentan a continuación están divididos en dos partes, los que se relacionan con los aspectos técnicos y aquellos que tienen que ver con los aspectos cualitativos del instrumento.

Aspectos Técnicos de cada subprueba

- TIEMPO APROPIADO PARA LEER LAS INSTRUCCIONES

La pregunta que determina la evaluación de este aspecto es "¿Alcanzaste a leer las instrucciones?".

De acuerdo con los criterios fijados en este estudio, la calidad de la presentación de las instrucciones alcanza niveles aceptables sólo

para la subprueba de Objetos. Tanto para Posiciones como para Colores, los porcentajes observados de aprobación (alcanzan a leer "todo" o "bastante") no alcanzan a ser significativamente mayores que los de los sujetos que encuentran deficiente dicho aspecto (alcanzan a leer "más o menos" o peor).

Existe correlación positiva y significativa, a un nivel de significación del 5%, entre la evaluación de la calidad de la presentación de las instrucciones en la subprueba de Posición y las variables de Edad ($\tau=0,35$) y Curso ($\tau=0,38$), en otras palabras se aprecia que los niños de menor edad y menor escolaridad tienden a encontrar insuficiente el tiempo destinado a la lectura de las instrucciones. Los valores para Chi cuadrado se presentan en la tabla 1

Tabla 1
Prueba de Chi cuadrado para valores dicotomizados de respuestas a ¿alcanzaste a leer u oír las instrucciones? según subtest.

¿Alcanzaste a leer/ oír las instrucciones?	Frecuencia y % de Respuesta Observada		Chi-cuadrado observado p (esperada) = 0,5
	Si	No	
Posición	23 63,9%	13 36,1%	2,778
Colores	22 61,1%	14 38,9%	1,778
Objetos	28 77,8%	8 22,2%	11,111 **

*: significativo para $\alpha = 0,05$

** : significativo para $\alpha = 0,01$

- CLARIDAD DE LA VOZ

La pregunta que determina la evaluación de este aspecto es "¿estuvo clara la voz?"

Se trata de una cualidad técnica pobremente evaluada. Los porcentajes de aprobación y desaprobación son estadísticamente

equivalentes en las 3 subpruebas, no obstante los peores resultados son para la subprueba de posición. La apreciación es similar sin importar grupos de edades, sexo, curso u horas semanales de acceso a computador. Los valores para Chi cuadrado se presentan en la Tabla 2

Tabla 2

Prueba de Chi cuadrado para valores dicotomizados de respuestas a ¿estuvo clara la voz? según subtest

¿Estuvo clara la voz?	Frecuencia y % Respuesta Observada		Chi-cuadrado observado p (esperada) = 0,5
	Si	No	
Posición	16 44,4%	20 55,6%	0,444
Colores	19 52,8%	17 47,2%	0,111
Objetos	19 52,8%	17 47,2%	11,111 **

* : significativo para $\alpha = 0,05$

** : significativo para $\alpha = 0,01$

• TAMAÑO ADECUADO DE LA LETRA DE LAS INSTRUCCIONES

La pregunta que determina la evaluación de este aspecto es ¿está bien el tamaño de la letra?''.

Aspecto evaluado muy positivamente con porcentajes superiores al 80% de aprobación en las 3 subpruebas. Se encontraron correla-

ciones significativamente positivas entre este aspecto y las variables de edad ($\tau=0,45$ con posiciones; $0,43$ con colores y $0,37$ con objetos) y escolaridad ($\tau=0,49$ con posiciones; $0,45$ con colores y $0,38$ con objetos). Estos valores indican que los niños de menor edad y de menor escolaridad tienden a evaluar menos positivamente este aspecto en las 3 subpruebas. Los valores para Chi cuadrado en la Tabla 3.

Tabla 3

Prueba de Chi cuadrado para valores dicotomizados de respuestas a ¿Está bien el tamaño de la letra? según subtest

¿Está bien el tamaño de letra?	Frecuencia y % Respuesta Observada		Chi-cuadrado observado p (esperada) = 0,5
	Si	No	
Posición	30 83,3%	6 16,9%	16,000**
Colores	33 91,7%	3 8,3%	25,000**
Objetos	30 83,3%	6 16,9%	16,000**

* : significativo para $\alpha = 0,05$

** : significativo para $\alpha = 0,01$

• **TIEMPO ADECUADO PARA INDICAR CON UN CLICK EN LOS BOTONES DE DIÁLOGO**

La pregunta que determina la evaluación de este aspecto es “¿te alcanzó el tiempo para hacer click?”.

Porcentajes de aprobación en torno al 75% obtuvo este aspecto lo que corresponde a una buena evaluación por parte de los sujetos de la muestra, sin influir edad, sexo, curso ni horas semanales de práctica en computador. Los valores para Chi cuadrado en la Tabla 4.

Tabla 4
Prueba de Chi cuadrado para valores dicotomizados de respuestas a ¿te alcanzó el tiempo para hacer click? según subtest

¿Te alcanzó el tiempo para hacer click?	Frecuencia y % Respuesta Observada		Chi-cuadrado observado p (esperada) = 0,5
	Si	No	
Posición	26 72,2%	10 27,8%	7,111**
Colores	28 77,8%	8 22,2%	11,111**
Objetos	28 77,8%	8 22,2%	11,111**

* : significativo para a = 0,05

** : significativo para a = 0,01

• **SUFICIENCIA DEL NÚMERO DE ENSAYOS**

La pregunta que determina la evaluación de este aspecto es “¿te faltó ensayar más?”.

En general, cerca de aproximadamente el 50% de los niños se sintió disconforme con la cantidad de ensayos previos a la ejecución completa de cada una de las pruebas. Los

resultados más notables se ven en la prueba de objetos, donde el porcentaje de rechazo es un poco mayor. Esta evaluación es similar para todos los grupos de edad, sexo, cursos y no se relaciona significativamente con la cantidad semanal de horas de acceso a computador. Los valores para Chi cuadrado se aprecian en la Tabla 5.

Tabla 5
Prueba de Chi cuadrado para valores dicotomizados de respuestas a ¿te faltó ensayar más? según subtest

¿Te faltó ensayar más?	Frecuencia y % Respuesta Observada		Chi-cuadrado observado p (esperada) = 0,5
	Si	No	
Posición	16 44,4%	20 55,6%	0,444
Colores	19 52,8%	17 47,2%	0,111
Objetos	20 55,6%	16 44,4%	0,444

* : significativo para a = 0,05

** : significativo para a = 0,01

• DIFERENCIACIÓN DE COLORES (SUBTEST COLORES)

La pregunta que determina la evaluación de este aspecto es “¿diferenciaste bien un color de otro?”.

Los resultados indican que en la subprueba de Colores los tonos usados en

la programación eran discriminados adecuadamente por la mayoría de los niños de la muestra sin importar edad, sexo, curso ni la cantidad de horas semanales de acceso al computador con un 95% de seguridad. Los valores para Chi cuadrado en la Tabla 6.

Tabla 6
Prueba de Chi cuadrado para valores dicotomizados de respuestas a ¿diferenciaste bien un color de otro? según subtest

¿Diferenciaste bien un color de otro?	Frecuencia y % Respuesta Observada		Chi-cuadrado observado p (esperada) = 0,5
	Si	No	
	26 72,2%	10 27,8%	7,111**

* : significativo para a = 0,05

** : significativo para a = 0,01

• CLARIDAD DE LAS IMÁGENES (SUBTEST OBJETOS)

La pregunta que determina la evaluación de este aspecto es “¿te parecieron claras las imágenes?”

Sólo 4 sujetos calificaron negativamente este aspecto. El 89% de la muestra encontró que las imágenes eran bastante o totalmente claras. En esta apreciación no influyo edad, sexo, curso ni horas semanales frente al computador. Los valores para Chi cuadrado en la Tabla 7.

Tabla 7
Prueba de Chi cuadrado para valores dicotomizados de respuestas a ¿te parecieron claras las imagenes? según subtest

¿Te parecieron claras las imágenes?	Frecuencia y % Respuesta Observada		Chi-cuadrado observado p (esperada) = 0,5
	Si	No	
	32 88,9%	4 11,1%	21,778**

* : significativo para a = 0,05

** : significativo para a = 0,01

• DISCRIMINACIÓN DE POSICIÓN DE FICHAS EN EL ESPACIO (SUBTEST POSICIONES)

La pregunta que determina la evaluación de este aspecto es: "¿pudiste encontrar la nueva ficha en el espacio?"

Dentro de las tareas específicas que deter-

minan el logro en las diferentes subpruebas, la capacidad para encontrar fichas nuevas en el espacio es la que presenta menor grado de desarrollo. El 45% de los sujetos tuvo problemas serios para realizar esta tarea sin importar edad, sexo, curso u horas de acceso semanal a computador. Los valores para Chi cuadrado en la Tabla 8.

Tabla 8

Prueba de Chi cuadrado para valores dicotomizados de respuestas a ¿pudiste encontrar las fichas nuevas en el espacio? según subtest

¿Pudiste encontrar las fichas nuevas en el espacio?	Frecuencia y % Respuesta Observada		Chi-cuadrado observado p (esperada) = 0,5
	Si	No	
	20 55,6%	16 44,4%	0,444

* : significativo para $\alpha = 0,05$

** : significativo para $\alpha = 0,01$

Aspectos cualitativos no formales de la prueba en general

• CALIFICACIÓN DE LA EXTENSIÓN DE LA ACTIVIDAD

La pregunta que determina la evaluación de este aspecto es: ¿Fue muy larga la actividad?

Aún cuando el 80% de los sujetos encontró que la duración de la prueba era adecuada, los niños mas pequeños y los de menor curso tendieron a evaluar como más larga la tarea encomendada ya que se encontró una correlación Tau de $-0,31$ con edad y de $-0,33$ con curso (significativa con un 95% de seguridad). Los valores para Chi cuadrado en la Tabla 9.

Tabla 9

Prueba de Chi cuadrado para valores dicotomizados de respuestas a ¿fue muy larga la actividad? según subtest

¿Fue muy larga la actividad?	Frecuencia y % Respuesta Observada		Chi-cuadrado observado p (esperada) = 0,5
	Si	No	
	7 19,4%	29 80,6%	14,211**

* : significativo para $\alpha = 0,05$

** : significativo para $\alpha = 0,01$

• CANSANCIO UNA VEZ FINALIZADA LA EJECUCIÓN

La pregunta que determina la evaluación de este aspecto es: ¿estás cansado(a) después de esta actividad?

En total 32 de 36 sujetos dijeron no encontrarse cansados en lo absoluto luego de la aplicación. En esta evaluación no influyo edad, sexo, escolaridad ni horas semanales de acceso a computador. Los valores para Chi cuadrado en la Tabla 10.

Tabla 10:
Prueba de Chi cuadrado para valores dicotomizados de respuestas a ¿te sentiste cansado(a) después de la actividad? según subtest

¿Te sentiste cansado(a) después de la actividad?	Frecuencia y % Respuesta Observada		Chi-cuadrado observado p (esperada) = 0,5
	Si	No	
	3 8,3%	33 91,7%	22,444**

* : significativo para a = 0,05

** : significativo para a = 0,01

• ADECUACIÓN DEL TIEMPO PREVISTO PARA REALIZARLA

La pregunta que determina la evaluación de este aspecto es: ¿alcanzaste a responder en el tiempo que se te dio?

El 83% de los sujetos respondió a esta pregunta “bastante” o “totalmente”. Esta variable correlaciona positiva y significativamente con edad (Rtau=0,38) y con curso (Rtau=0,39), lo que señala que a mayor edad y más alta escolaridad aparece mayor adecuación al tiempo previsto para responder. Los valores para Chi cuadrado en la Tabla 11.

Tabla 11
Prueba de Chi cuadrado para valores dicotomizados de respuestas a ¿alcanzaste a responder en el tiempo que se te dio? según subtest

¿Alcanzaste a responder en el tiempo que se te dio?	Frecuencia y % Respuesta Observada		Chi-cuadrado observado p (esperada) = 0,5
	Si	No	
	30 83,3%	6 16,7%	16,000**

* : significativo para a = 0,05

** : significativo para a = 0,01

• GRADO DE ABURRIMIENTO

La pregunta que determina la evaluación de este aspecto es: ¿te sentiste aburrido en algún momento?

Esta característica esta particularmente bien evaluada: 34 niños de 36 señalaron no haberse aburrido nunca sin influir edad, sexo, escolaridad ni horas de acceso al computador.

- DISPOSICIÓN A REPETIR LA ACTIVIDAD

La pregunta que determina la evaluación de este aspecto es: ¿harías de nuevo esta actividad?

El 13,8% de los sujetos respondió a esta pregunta "bastante" y el 86,1% restante respondió "totalmente", no obstante son los niños de mayor edad y de más alta escolaridad quienes están más proclives a repetir el test y seguir practicando. Las correlaciones observadas son de 0,35 con edad y 0,33 con escolaridad.

Discusión

Respecto a los resultados de la investigación piloto, el primer hallazgo a destacar se refiere a los aspectos cualitativos no formales del instrumento.

El grado de exigencia de la tarea para los evaluados fue razonablemente tolerable para la gran mayoría. Se aprecia que, en la perspectiva de toda la muestra, la prueba no provocó agotamiento entre los evaluados, no fue catalogada como extensa y el tiempo previsto para ejecutarla se ajustó a los requerimientos de niños entre 5 y 10 años. No obstante, se observó que en niños de menor edad la exigencia de las tareas propuestas demandó un uso más intensivo de sus capacidades lo que se expresó con un grado mayor de agotamiento al finalizar la prueba si se le compara con niños de mayor edad.

En cuanto al grado de motivación que la ejecución informatizada generó en los sujetos, se observó que la prueba no provocó aburrimiento significativo en ninguno de los niños evaluados; es más, la totalidad de la muestra se mostró dispuesta a repetir la actividad en un futuro cercano.

En suma, la prueba fue asumida como un desafío entretenido, atractivo y motivante. Esto en sí mismo hace pensar en lo oportuno de generar material educativo y psicométrico en formato computacional. El formato

multimedia implica el uso combinado de diversos modos sensoriales (visual, táctil y auditivo) lo que genera en los niños un alto grado de interés, particularmente si la exigencia de la tarea es coherente con su nivel de habilidades.

En cuanto a los aspectos técnicos del instrumento se observaron ciertas falencias que es necesario destacar. Lo primero que se debe mencionar es la relativa confusión que se generó en torno a la comprensión de las instrucciones. Las tres subpruebas presentaron, en mayor o menor medida dificultades en este aspecto (lectura de instrucciones y claridad de la voz), particularmente la prueba de posición la cual pareciera ser más compleja para los niños en función de la clase de habilidades memorísticas que mide

Una vía de solución a esta falencia nace observando la manera en que los niños se enfrentan a nuevos juegos de computación. Estos por lo regular presentan pantallas de animación que demuestran la forma correcta de operar. Sería un aporte interesante el que la tarea a desarrollar en cada subprueba sea modelada previamente mediante una demostración animada, acompañada de las instrucciones en formato de audio, ya que al parecer el formato de texto complica la comprensión, particularmente a aquellos niños más pequeños que no manejan adecuadamente la lectura.

Un aspecto muy positivo resultó la capacidad de interacción de los niños con la prueba. Los cuadros de diálogo funcionaron de buena manera lo que se refleja en no encontrar problemas de importancia en el tiempo destinado a indicar con un *click* las opciones preferidas. Se trató de lograr interactividad a través del diseño de cuadros de diálogo atractivos en color y agregando diseños que expresaran en nivel visual la contingencia del contenido (p.ej.: "caritas" sonrientes en los cuadros de felicitaciones, sorpresa en el error en los ensayos, tristeza en las finalizaciones de las pruebas, etc.). Este logro en interacti-

vidad también es reflejo de las capacidades de los niños relativas al dominio básico de computadores. A pesar de su poca experiencia con los computadores a los niños les bastó con una práctica semanal compartida frente al computador para alcanzar un aceptable nivel de dominio del "mouse" y la lógica de los botones de diálogo.

Un aspecto menos positivo tiene que ver con la suficiencia del número de ensayos, previo a la ejecución puntuable de la tarea. Los niños de todas las edades señalaron como insuficientes las posibilidades de practicar que la prueba le ofrecía. La forma de enfrentar la prueba sin la supervisión directa de un evaluador lleva a que los niños exploren mediante ensayo y error. Esta vez la motivación por la tarea juega en contra de los objetivos de evaluación ya que los niños se sienten "obligados" a probar los límites y "secretos" de la prueba, lo cual es imposible de lograr en tan pocos ensayos.

Este punto sirve para determinar los caminos que se abren en el desarrollo del MEMO-POC. Por una parte, para ser usada como instrumento de medición psicométricamente adecuado, una vez solucionadas ciertas deficiencias técnicas, la prueba puede convertirse en un eficaz medio de evaluación de trastornos en la memoria visual, ya no sólo en niños sino también en adultos, asociado por ejemplo, a tareas como el diagnóstico del grado de compromiso cognitivo que puedan estar provocando enfermedades degenerativas del sistema nervioso central (Alzheimer, Accidentes Vasculares Encefálicos, etc.), y sin descartar el uso potencial en el terreno de la selección de personal en empleos que requieran un dominio preciso de las habilidades memorísticas visuales de corto plazo (operadores de sistemas de radar, controladores de calidad, etc.)

Por otra parte, la prueba tiene otro potencial de desarrollo, más allá de la medición psicométrica. Al diseñar el instrumento en su forma de autoaplicación se está creando una eficaz y atractiva batería de entrenamiento y estimulación de habilidades memorísticas específicas, particularmente útil para niños con trastornos específicos de aprendizaje, sujetos en rehabilitación de daño cerebral, etc. Una presentación de este tipo permitiría que fuera puesto a disposición de alumnos y profesores, vía Internet en el contexto del proyecto ENLACES del Ministerio de Educación.

En general, la experiencia de diseñar y poner a prueba un test psicológico de ejecución en formato informatizado tiene en sí misma elementos importantes de destacar. No existen en nuestro entorno profesional psicólogos especializándose en esta clase de aplicaciones computacionales. Este hecho hace que la experiencia de este trabajo resulte pionera, al menos para el contexto regional. Por otra parte, el trabajo coordinado con ingenieros programadores/diseñadores crea una nueva área de cooperación interdisciplinaria de interesante proyección

El logro fundamental de este estudio consistió en transformar una prueba muy poco dúctil (difícil de transportar, cara en materiales, etc.) y en la práctica inaplicable, en un instrumento transportable, amigable, fácil de aplicar y de gran atractivo para los niños.

Los pasos subsecuentes son corregir las falencias detectadas y proceder a la determinación de las características psicométricas del instrumento, validez de constructo y fiabilidad, en la confianza que se han controlado un número importante de fuentes de variación en los puntajes causadas por errores de presentación del instrumento.

Referencias Bibliográficas

- BUNDERSON, C., INTUYE, D. Y OLSEN J. (1989) The four generations of computerized Educational measurement. En R. L. Linn (Eds) , *Educational Measurement* Londres: McMillan.
- CHANG, S., Y ANSLEY T. (2003) A comparative Study of item exposure control methods In computerized adaptive testing. *Journal of educational Measurement*. **40**, 71–103.
- CONCHA (2003) *Utilización de tests informatizados en el quehacer del psicólogo*. Seminario para optar al grado de Licenciado en Psicología. Universidad de Concepción. Concepción Chile.
- HUFF, K., Y SIRECE, S. (2001) Validity in computer based testing. *Educational Measurement, Issues and Practice*. Vol. 20, 16–27.
- GREEN, B., BOCK, R., HUMPHREYS, L. , LINN., R., Y RECKASE, M. (1984) Evaluation plan For the computerized adaptive tests. *Journal of Educational Measurement*. 21, 347–360.
- MCBRIDE, J. (1998) Innovation in computer-based ability testing: promise, problems, and perils. En M. D. Hakel (Eds) *Beyond multiple choices*. Londres; LEA
- MUÑIZ J. Y HAMBLETON R. (1999) Evaluación psicométrica de los tests informatizados. En Olea J., Ponsada, V. y Prieto G. (Eds). *Test Informatizados*. Madrid: Pirámide
- OLEA J., Y HONTAGAS, P. (1999). Tests Informatizados de primera generación. En Olea J., Ponsada, V. y Prieto G. (Eds). *Test Informatizados* Madrid: Pirámide
- OLEA J., Y PONSODA, V. (1996) Tests adaptativos informatizados. En C. Vizcarro y J. A. León (Eds) *Nuevas tecnologías para el aprendizaje*. Madrid: Pirámide
- OLEA J. Y PONSADA, V. (1998) Test informatizados y adaptativos informatizados: Investigación en España. *Revista Electrónica de Investigación y Evaluación Educativa*, Vol. 4, N° 2.
- OLEA J., PONSADA, V., Y PRIETO, G. (1999) *Test Informatizados: fundamentos y Aplicaciones*. Madrid: Pirámide
- PITKIN A., Y VISPOEL W. (2001) Differences between self-adaptated and computerized Adaptive test: A meta-analysis. *Journal of Educational Measurement*, 38,235– 247
- PRIETO G. Y DELGADO, A. R. (1999) Medición cognitiva de las aptitudes En Olea J.,
- PONSADA, V. Y PRIETO G. (Eds). *Test Informatizados* Madrid: Pirámide.
- REHBEIN, L. (1984) *Long term effects of early hippocampectomy in the monkey*. Tesis Doctoral no publicada, Northeastern University, Boston. USA.
- ROCKLIN, T Y O'DONNELL A. M. (1987) Self-adapted testing: A performance-improving Variant of computerized adaptive testing. *Journal of Educational Psychology*, 77, 368–372.
- ROCKLIN, T., O'DONNELL A. M. Y HOLST P. (1995) Effects and underlying mechanisms of self-adapted testing. *Journal of educational Psychology*, 77, 368–372.
- SCHADE, N. (1993) *Validación y Fiabilidad de un Test de Reconocimiento Acumulativo*. Tesis Doctoral. Editorial Kadmos: Universidad Pontificia de Salamanca. España.
- SCHATZ, P., Y BROWNDYKE, J. (2002) Applications of computer based neuropsychological **Assessment**. *The Journal of Educational Measurement*, Vol 37, 21–38.

- VISPOEL, W. P. HERINDRICKSON, A. Y BLEILER, T. (2000) Limiting answer review and change on computerized adaptive vocabulary test: Psychometric and attitudinal results. *Journal of Educational Measurement*, 37, 21–38.
- VISPOEL, W. P. CLOUGH, S., BLEIR, T., HENDRICKSON A., Y IHRIG, D., (2002) Can Examinees use judgements of item difficulty to improve proficiency estimate On computerized adaptative vocabulary test. *Journal of Educational Measurement*, 39, 311–339.
- VISPOEL, W., ROCKLIN, T. Y WANG, T. (1994). Individual differences and test Administration procedure: A comparison of fixed-item, computerized-adaptive and self-adaptive testing. *Applied Measurement in Education*, 7 (1) 53–79.
- VISWESVARAN, C. (2000) Innovations in computerized assessment. *Personnel Psychology*. Vol 53, 260–263.
- WAINER, H. (1990) *Computerized adaptive testing: a primer*. Hillsdale, NJ: LEA.