

# Especificación del algoritmo para un Test Adaptativo Informatizado de Analogías Verbales<sup>1</sup>

## Algorithm Specification for a Computerized Adaptive Test of Verbal Analogies

Gabriela Lozzia<sup>2</sup> Horacio Attorresi<sup>3</sup>

Instituto de Investigaciones de la Facultad de Psicología de la Universidad de Buenos Aires, Argentina

(Rec: octubre de 2012 – Acep: noviembre de 2012)

### Resumen

Se presentan los pasos seguidos en el diseño de un Test Adaptativo Informatizado de Analogías Verbales. El algoritmo adaptativo se determinó teniendo en cuenta las características de un Banco de Ítems de Analogías Verbales construido a partir de la Teoría de Respuesta al Ítem, los objetivos de evaluación, la población por evaluar y las características del software. El Banco está compuesto por 64 ítems unidimensionales calibrados según el Modelo Logístico de Tres Parámetros, informativos en todo el rango del rasgo y libres de funcionamiento diferencial entre géneros. Se programó el algoritmo adaptativo con el FastTEST Professional Testing System utilizando las siguientes especificaciones: a) inicio aleatorio entre niveles levemente inferiores a la media del rasgo, b) estimación del rasgo por Máxima Verosimilitud Condicional, c) selección progresiva de los ítems con el Método de Máxima Información de Fisher y d) criterio mixto de finalización.

Palabras clave: Test Adaptativo Informatizado, Banco de Ítems y Analogías Verbales.

### Abstract

This paper presents the steps followed in the design of a Computerized Adaptive Test to measure Verbal Analogies. The adaptive algorithm was determined upon the characteristics of a Verbal Analogies' Item Bank constructed on the basis of Item Response Theory, the assessment objectives, the studied population and software features. The Bank consists of 64 unidimensional items calibrated with the Three Parameter Logistic Model, which are informative throughout the entire latent trait's range and free of differential item functioning between genders. The adaptive algorithm was programmed via FastTEST Professional Testing System using the following specifications: a) random start between levels slightly below average trait, b) Conditional Maximum Likelihood Trait Estimation, c) progressive items selection by Maximum Fisher Item Information, and d) combined stopping rule.

Keywords: Computerized Adaptive Tests, Item Bank and Verbal Analogies.

<sup>1</sup> La investigación que se presenta en este artículo fue realizada con subsidios de la Universidad de Buenos Aires (UBACyT N° 20020100100346) y de la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT PICT 2011 N° 0826), en el marco de la Beca Interna de Postgrado Tipo I otorgada a la Lic. Gabriela Lozzia por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

<sup>2</sup> Correspondencia a: Av. Lope de Vega 1507 Dpto. 2, (1407) Ciudad Autónoma de Buenos Aires, Argentina. Te: (011) 4568-6172. E-mail: glozzia@psi.uba.ar y gabrielalozzia@gmail.com

<sup>3</sup> E-mail: hatorre@psi.uba.a

## Introducción

Actualmente la psicometría ha retomado su interés por procedimientos de evaluación en función de las características de las personas gracias al desarrollo de los *Tests Adaptativos Informatizados* (TAI, traducción de la expresión inglesa *Computerized Adaptive Test*, CAT). Este tipo de procedimiento de evaluación no era viable dentro del marco de la Teoría Clásica de Tests por la imposibilidad de establecer comparaciones entre las puntuaciones de las personas obtenidas a partir de diferentes ítems. Los avances de la tecnología informática posibilitaron aplicar los nuevos modelos psicométricos de la *Teoría de Respuesta al Ítem* (TRI) a la construcción de *Bancos de Ítems* (BIs) y obtener a partir de ellos instrumentos que presentaran únicamente los reactivos que fueran altamente informativos para estimar el nivel de habilidad de cada individuo (Wainer et al., 2000). Olea y Ponsoda (1996) señalan que se trata de pruebas cuyos ítems se seleccionan según el nivel de competencia que progresivamente va manifestando la persona mediante un algoritmo aplicado en una computadora.

Son muchos los tests convencionales para los cuales existen versiones adaptativas y es frecuente, tanto en Estados Unidos como en Europa, su uso en selección de personal, admisión a centros educativos, exámenes de licenciatura o certificación y en clínica (Bartram y Hambleton, 2006; Forbey y Ben-Porath, 2007; Hol, Vorst y Mellenbergh, 2008; van der Linden, 2008). En Argentina, si bien existen algunas aplicaciones de la TRI (Abal, Lozzia, Picón Janeiro, Galibert y Aguerri, 2007; Attorresi, Lozzia, Abal, Galibert y Aguerri, 2009; Tornimbeni, Pérez y Olaz, 2008), hay pocos desarrollos acerca de la construcción de BIs (Lozzia, Abal, Aguerri, Galibert y Attorresi, 2006; Lozzia et al., 2012; Lozzia, Galibert, Aguerri y Attorresi, 2005) y ninguno de TAIs. Por ello se inició un proyecto que busca profundizar el conocimiento teórico y metodológico de la TRI al crear un TAI de Analogías Verbales. Dicho TAI sería aplicado con fines diagnósticos al estudio de uno de los componentes de la competencia intelectual de estudiantes de psicología.

Se eligió evaluar el razonamiento verbal porque es una capacidad cognitiva requerida en la resolución de problemas simples y complejos, tanto de orden intelectual como de situaciones cotidianas. Esta capacidad es clave para el éxito en la universidad y en la vida profesional (Lohman, 2004). Se operacionalizó mediante ítems de analogías verbales que miden la habilidad para reconocer y discriminar relaciones entre pares

de palabras (Lozzia, Picón Janeiro y Galibert, 2008). Esta habilidad correlaciona con el factor ideativo de la comprensión verbal (Yela, 1987) y es útil para obtener el perfil intelectual de los estudiantes de una gran variedad de carreras tanto humanísticas como técnicas (Kuncel, Hezlett y Ones, 2004; Sternberg, 1988). En efecto, muchos investigadores encontraron que el rendimiento en analogías representa una de las mejores medidas de la comprensión verbal y el pensamiento analítico (Goswami, 2001; Sternberg, 1988; 2001). Los ítems están formados por un par de palabras base entre las cuales existe algún tipo de relación y cuatro opciones de pares de palabras. La resolución de los mismos implica seleccionar entre las opciones el par que presenta la relación más próxima a la que existe entre las palabras del par base (Galibert, Aguerri, Pano, Lozzia y Attorresi, 2005).

Todo TAI requiere para su funcionamiento de dos componentes:

1. *Un banco de ítems calibrados desde un modelo de la TRI*. Se trata de un conjunto de reactivos que miden un mismo rasgo. Sus propiedades psicométricas son conocidas; es decir, sus parámetros están estimados en una misma escala (calibrados). A partir del BI se selecciona el conjunto de ítems más apropiado para cada individuo.
2. *Un algoritmo adaptativo informatizado*. Todos los tests son administrados siguiendo una serie de reglas que determinan los ítems por responder y su orden de presentación. Al conjunto de estas especificaciones se lo denomina algoritmo del test (*testing algorithms*) (Thissen y Mislevy, 2000). En el caso de los TAI, se trata de un algoritmo adaptativo, ya que su característica distintiva es que la selección dinámica de los ítems se realiza en función del nivel de rasgo que va manifestando el evaluado al completar el test. Las especificaciones básicas de un algoritmo adaptativo indican qué procedimiento se seguirá en cada uno de sus cuatro componentes: estrategia de inicio, método estadístico para estimar el nivel de rasgo, procedimiento para la selección de ítems y estrategia de finalización.

No se puede diseñar un TAI mientras no haya sido calibrado un BI. En investigaciones anteriores se calibró un Banco de Ítems de Analogías Verbales que sirve de base para generar diversos tipos de tests (e.g., TAIs, Tests Paralelos, Tests referidos al Criterio, Tests con Características Prefijadas) y permite evaluar en estudiantes universitarios la habilidad para reconocer

y discriminar relaciones (para más información ver Lozzia et al., 2012). Este BI cumple con las características que debe tener para ser utilizado como base de un TAI: incluir ítems informativos a lo largo de todo el rango del rasgo. El BI está compuesto por 64 ítems unidimensionales calibrados con el Modelo Logístico de Tres Parámetros. Son reactivos que no evidenciaron DIF por género, presentaron una capacidad discriminativa adecuada (parámetro  $a > 0.65$ ) y un nivel de acierto por azar cercano a lo esperable para ítems con cuatro opciones de respuesta. El BI contiene una cantidad suficiente y variada de ítems que permite evaluar con precisión los niveles de habilidad comprendidos entre -1.75 y 3.00 (Lozzia et al., 2012).

El objetivo de este artículo es metodológico: presentar los pasos seguidos en la especificación del algoritmo para obtener un TAI basado en el Banco de Ítems de Analogías Verbales. Se desarrollan las distintas posibilidades que presentan los componentes del TAI, indicando para cada uno de ellos el fundamento de las decisiones tomadas para el de Analogías Verbales.

### Desarrollo

Para diseñar el TAI de Analogías Verbales se siguieron los siguientes pasos:

- Seleccionar el software adecuado.
- Determinar las especificaciones para cada uno de los componentes del algoritmo adaptativo (manera de comenzar y finalizar la prueba, forma de seleccionar progresivamente los mejores ítems y método de estimación de los niveles de rasgo).
- Programar el software.
- Almacenar los ítems del BI base del TAI y sus características psicométricas en el software.
- Comprobar el correcto funcionamiento del TAI diseñado.

A continuación se desarrolla cada uno de estos puntos indicando las decisiones técnicas tomadas para su especificación.

#### *Selección del software*

Se examinaron los programas de computación que posibilitaban la implementación práctica del TAI y se eligió el software más apropiado para este trabajo. Este fue un paso importante dado que el tipo de programa conseguido y sus características determinarían muchos

de los aspectos a considerar en el diseño del TAI. Entre los programas informáticos comercializados se eligió uno de los más recientemente presentados en el mercado, que supera las limitaciones de los anteriores. Este fue el FastTEST Professional Testing System de Assessment Systems Corporation en su versión 2.0 (Weiss, 2008).

#### *Especificación de los componentes del algoritmo adaptativo*

A partir de las posibilidades de programación que brinda el software FastTEST Pro se estuvo en condiciones de definir las características que tendría el algoritmo adaptativo. Por lo tanto, se especificaron cada uno de los componentes del TAI: modo de inicio, método de estimación de los niveles de rasgo, estrategia para la selección progresiva de los ítems y modo de finalización. Para ello se tuvieron en cuenta las diversas opciones que se pueden elegir para establecer estos requerimientos, así como la conveniencia de cada una para los objetivos específicos de este TAI, sus ventajas e inconvenientes (Olea y Ponsoda, 2003; van der Linden, 2008; van der Linden y Glas, 2007).

La selección de una determinada estrategia depende de múltiples factores como: objetivos del TAI (certificación, diagnóstico, selección, etc.), consecuencias de la toma de decisiones basadas en los resultados del TAI (e.g., TAI de certificación), nivel de precisión deseado en la estimación del rasgo, características de la población objetivo (e.g., niños, ancianos, personas con algún tipo de discapacidad), longitud del TAI y si ésta es fija o variable, tamaño y composición del BI, actualización de los parámetros de los ítems, seguridad del BI (i.e., riesgo de divulgación de ítems), tiempo disponible de evaluación, fatiga de los evaluados, otras restricciones adicionales (e.g., contenido de los ítems, tiempo de respuesta, posibilidad de omitir, revisar y/o corregir las respuestas), disponibilidad de información previa acerca del rendimiento del evaluado que pueda utilizarse para optimizar el TAI, etc.

#### *Estrategia de inicio*

Especifica de qué modo seleccionar el primer ítem por presentar. Se analizaron distintas variantes. Se descartaron en primer lugar las que no correspondían a los objetivos del TAI como: a) comenzar con ítems cercanos al punto de corte (no era un TAI de acreditación), b) utilizar datos externos como información de los evaluados predictora de su nivel de rasgo (no se

disponía de la misma), o c) que el propio examinado eligiera su nivel inicial entre un conjunto de valores cualitativos preestablecidos (los evaluados no tenían conocimiento sobre su nivel en analogías verbales).

Como se conocía la distribución de la habilidad, gracias a las muestras con las que se calibró el BI de Analogías Verbales, la media de esta distribución podría ser una estimación razonable para el  $\theta$  inicial al no disponer de ninguna otra orientación (Embretson y Reise, 2000; Thissen y Mislevy, 2000). Sin embargo, se decidió comenzar con un  $\theta$  levemente inferior a la media para asegurar una primera experiencia satisfactoria que bajara la ansiedad ante la evaluación (Embretson y Reise, 2000). Además, este procedimiento ayudaría a reducir los índices de exposición de los ítems de dificultad media. Por ello, ésta es una de las estrategias de inicio más utilizadas (e.g., Xing y Hambleton, 2004).

Con respecto al modo de inicio, el programa FastTEST Pro brinda dos posibilidades: a)  $\theta$  inicial igual para todos los evaluados, o b) elección aleatoria del  $\theta$  inicial dentro de un intervalo. El primer reactivo que presentará el software es el más informativo para el  $\theta$  asignado. Para ampliar la diversidad de ítems aplicados en los primeros estadios del TAI, se programó el software para que asignara a cada evaluado un nivel de rasgo inicial seleccionado al azar entre niveles de  $\theta$  superiores a -1.0 e inferiores a -0.5. Son seis los ítems del BI cuya máxima información estaba comprendida dentro del intervalo mencionado; por lo tanto, había una cantidad suficiente de reactivos que permitiría la selección aleatoria (Embretson y Reise, 2000).

### *Método de estimación del nivel de rasgo*

Este procedimiento se aplica luego de responder cada ítem del TAI y es el que determina cuál es el valor de la escala  $\theta$  que más se ajusta al patrón de respuestas emitido hasta ese momento por el evaluado. Se trata de un caso de estimación condicional ya que el proceso de estimación está condicionado a los parámetros conocidos (calibrados) de los ítems que conforman el BI que es base del TAI. El método elegido brindará la estimación de  $\theta$  que cumplirá tres funciones: ser la base para la selección del siguiente ítem por presentar, determinar la finalización del TAI cuando la estimación alcanza cierta precisión y ofrecer el nivel final del individuo en el rasgo (Muñiz, 1997).

Con respecto a la estimación progresiva del nivel de rasgo tras cada respuesta al TAI, se estudiaron los dos métodos clásicos: el procedimiento de Máxima Verosimilitud (ML), que consiste en maximizar la

función de verosimilitud del patrón de respuestas (Lord, 1980), y los bayesianos (Owen, 1975), que añaden a la función de verosimilitud información acerca de la distribución a priori de la habilidad de la población.

Generalmente, los TAIs que utilizan métodos bayesianos cuentan con información relevante y segura que, al incorporarla a la función de verosimilitud, les permite mejorar las estimaciones (Embretson y Reise, 2000). Los procedimientos bayesianos tienen la ventaja de que pueden aplicarse después de responder al primer ítem y ante patrones de respuestas constantes y, además, reducen el error estándar del estimador ya que se dispone de más información sobre la distribución de  $\theta$ . Sin embargo, presentan una limitación importante relacionada con la adecuación de la distribución a priori y el sesgo. El  $\theta$  estimado no depende únicamente del rendimiento del evaluado, sino también de los valores de media y varianza que se estipulan para la distribución a priori de  $\theta$ . Por ello, se prefiere ML que es un estimador asintóticamente eficiente e insesgado. Lord (1986) demostró que el sesgo es mínimo si la dificultad de los ítems se ajusta al nivel de  $\theta$  del evaluado. Como esto es lo que sucede en los TAIs, se considera adecuado utilizar ML en estos casos.

Se programó el FastTEST Pro de modo que la estimación se realizara por el procedimiento ML. La estimación de  $\theta$  obtenida será un valor numérico comprendido en el rango de -4 a 4 que determina el siguiente ítem a seleccionar (el más informativo para el nivel estimado hasta ese momento). Este proceso se repite hasta que finaliza la evaluación. Para cada  $\hat{\theta}$  se obtendrá el Error Estándar del Estimador,  $EEE(\hat{\theta})$ . Éste es el valor inverso de la raíz cuadrada de la información que aportan para el último nivel de rasgo estimado los ítems presentados hasta ese momento (Olea, Abad, Ponsoda y Ximénez, 2004).

Para resolver el problema de que ML no proporciona estimaciones finitas ante patrones de respuesta constante, el programa FastTEST Pro opera de la siguiente manera: si se da una respuesta correcta,  $\theta$  se establece en 4; mientras que para las respuestas incorrectas,  $\theta$  se establece en -4. Este procedimiento tiende a forzar a un patrón de respuesta mixta con mayor rapidez que los *stepzises* más pequeños.

En cuanto a la estimación final del nivel de rasgo del evaluado, algunos TAI utilizan un método distinto al empleado para las estimaciones provisionales (e.g., Segall y Moreno, 1999). En estos casos, a partir del patrón de respuestas final obtenido se realiza la estimación con el método propuesto. Mientras que si el procedimiento de estimación final es el mismo que el

utilizado en la estimación provisional, el  $\theta$  provisional obtenido tras responder al último ítem se convertirá automáticamente en el  $\theta$  final. Para este TAI, se decidió continuar con una estimación ML. Del mismo modo que en la estimación provisional de  $\theta$ , ML resulta más adecuado ya que los procedimientos bayesianos pueden derivar en problemas de equidad en la evaluación (dos personas con el mismo patrón de respuestas podrían no obtener idéntico resultado final). Los estudios con datos reales y simulados sugieren que la estimación ML es adecuada para determinar el nivel final de rasgo y que un buen indicador de su precisión es el  $EEE(\hat{\theta})$  (Thissen y Mislevy, 2000).

A la hora de considerar el puntaje  $\hat{\theta}$  estimado para una persona, se tendrá en cuenta la precisión a través del  $EEE(\hat{\theta})$ . La salida del programa FastTEST Pro brinda el  $\hat{\theta}$ , su  $EEE(\hat{\theta})$  y el intervalo de confianza de 95.5%.

#### *Procedimiento para la selección de ítems*

Los procedimientos actuales siguen una *Estrategia de Nivel Múltiple con Ramificación Variable o Flexible*, que permite actualizar el nivel  $\theta$  estimado para el evaluado después de responder a cada ítem que se le presenta y en función de este nivel seleccionar el siguiente ítem. Esto hace que el TAI sea adaptativo. El reactivo es elegido en función tanto de sus propiedades psicométricas (parámetros de la TRI) como del nivel de rasgo estimado para el examinado en ese punto de la evaluación (Barrada, Olea y Ponsoda, 2004).

Se consideraron las dos estrategias más difundidas para seleccionar los ítems a partir de una estimación provisional de  $\theta$ : a) máxima información de Fisher y b) máxima precisión esperada (o su equivalente: mínima varianza posterior esperada).

El método de máxima precisión esperada pertenece a la familia de criterios de selección bayesiana (Owen, 1975), que eligen el ítem que minimiza la varianza de la distribución a posteriori de la habilidad. Esta distribución es actualizada con la presentación de cada ítem e incluye toda la información disponible en cuanto a la distribución de  $\theta$  en la población y las respuestas dadas a los ítems ya presentados. Esta estrategia se corresponde con el uso del procedimiento bayesiano para la estimación del nivel del rasgo. Pero fue descartado el uso de información a priori en el TAI de Analogías Verbales.

El método de máxima información de Fisher (Lord, 1980) elige el ítem, entre los que aún no se han administrado, cuya Función de Información (FI) proporciona

el mayor valor para el nivel de habilidad estimado hasta el momento, o lo que viene a ser lo mismo, el que minimiza el error estándar, y en consecuencia maximiza la precisión, para la estimación actual de la habilidad del evaluado. Un ítem aportaría más información para un determinado nivel de  $\theta$  cuanto más elevado sea su parámetro de discriminación, menor sea su parámetro  $c$  y más cerca se encuentre el parámetro de dificultad del nivel de  $\theta$  (Olea y Ponsoda, 2003). Este criterio de selección es utilizado por la mayoría de los TAI y los software diseñados para administrarlos, no se ve afectado por el orden de presentación de los ítems e impone menos restricciones ya que no hace suposiciones sobre la distribución del rasgo a priori (Thissen y Mislevy, 2000).

El software FastTEST Pro sólo utiliza el procedimiento de máxima información de Fisher para seleccionar los ítems. Pero permite decidir entre su aplicación directa o su aplicación con el método de control de exposición por introducción de un componente aleatorio en la selección (i.e., eligiendo un ítem al azar entre los más informativos). Sin embargo, este método repercute en la eficiencia del TAI (la reduce) y da lugar a TAI un poco más largos. Por ello, se seleccionó la primera opción.

No se fijó un tiempo límite para la presentación y respuesta a cada ítem, ya que los TAI son más eficientes cuando se administran sin restricción de tiempo (Zickar, Overton, Taylor y Harms, 1999) y, por otro lado, los modelos clásicos de la TRI no toman en consideración los efectos de la presión del tiempo sobre el funcionamiento del ítem (Segall y Moreno, 1999).

#### *Estrategia de finalización*

Se determinó un criterio para dar por terminada la evaluación ya que, de lo contrario, la presentación de ítems seguiría hasta administrar todos los reactivos del BI. Si bien se puede fijar la longitud, los TAI cuentan con la posibilidad novedosa de finalizar la prueba cuando se ha alcanzado determinado nivel de precisión (fijado por el evaluador) en la estimación del rasgo. Se espera que, a medida que se presenten más ítems, el  $\theta$  estimado se acerque de manera cada vez más precisa al nivel de rasgo real y el  $EEE(\hat{\theta})$  vaya disminuyendo hasta alcanzar un valor considerado aceptable para dar por terminado el test.

Un TAI de longitud fija será más corto que su versión convencional, ya que es más eficaz (i.e. requiere administrar menos ítems). Con este procedimiento

todos los evaluados responderán la misma cantidad de reactivos pero sus estimaciones finales de  $\theta$  tendrán diferente nivel de precisión (Renom y Doval, 1999).

Para este TAI se prefirió utilizar una longitud variable ya que este criterio de parada hace el mejor uso del algoritmo adaptativo (Embretson y Reise, 2000). Dicha estrategia utiliza de forma óptima el BI, dado que cada evaluado recibe el largo mínimo de test (i.e., el menor número de ítems) necesario para estimar de forma precisa su nivel de rasgo. Se buscaba conseguir el objetivo, altamente valorado desde el punto de vista estadístico, de obtener estimaciones del nivel de habilidad con el mismo error de estimación para todos los evaluados. Sin embargo, alcanzarlo exigía que el BI tuviera una FI lo suficientemente alta y uniforme para que esto sea posible (Embretson y Reise, 2000). La distribución de la FI del BI de Analogías Verbales no era uniforme a lo largo de todos los niveles de  $\theta$ ; en consecuencia, sería difícil alcanzar el  $EEE(\hat{\theta})$  prefijado en ciertas zonas del rasgo evaluado. La FI era baja en los valores extremos del rasgo dado que, como suele suceder, fue difícil elaborar muchos ítems con alto valor discriminativo para estos niveles (Martínez Arias, 1995). Si se administraba el TAI bajo estas condiciones existiría una enorme variabilidad en la cantidad de ítems que sería necesario presentar hasta alcanzar el nivel de precisión pretendido, y en los evaluados cuyos niveles de  $\theta$  estimados no alcanzarían la precisión especificada, la presentación de los ítems seguiría hasta agotar todos los reactivos del BI.

Para evitar la administración innecesaria de ítems que no aportan más precisión a la estimación del rasgo se pueden especificar distintos niveles de precisión según el nivel de rasgo o detener el TAI cuando el error decrece menos de 0.01 con la presentación de los siguientes ítems. Pero el programa FastTEST Pro no lo permitía. Por ello, se especificó un criterio de parada mixto. Éste es el más utilizado actualmente ya que conjuga los beneficios de los dos procedimientos, el de longitud fija y el de longitud variable. Se establece a priori un determinado nivel de precisión para que el test se detenga al alcanzarlo. En los casos que la estimación de la habilidad no alcanza la precisión deseada, el TAI finaliza luego de administrar una determinada cantidad de reactivos fijada por el evaluador.

Se buscaba que el TAI de Analogías Verbales evaluara con igual o mayor precisión que su versión convencional. En las pruebas de calibración se obtuvieron índices de confiabilidad ( $\alpha$  de Cronbach) entre .77 y .85. Esto correspondería aproximadamente a un  $EEE(\hat{\theta})$  de entre 0.39 y 0.48 ( $\sqrt{1-\alpha}$ ). Por lo tanto, el  $EEE$

( $\hat{\theta}$ ) meta debía ser igual o inferior a 0.40. Se descartó utilizar un  $EEE(\hat{\theta})$  inferior, pues la FI del BI no era uniformemente alta a lo largo de toda la escala de valores  $\theta$ . Utilizar un error no superior a 0.40 permitía evaluar a los individuos en el rango de -1.20 a 2.40. Este rango era lo suficientemente grande para contener a la mayoría de evaluados. Para las personas con  $\theta$  por fuera de este intervalo, la estimación de su habilidad no alcanzaría la precisión deseada. En estos casos, el TAI finalizaría luego de administrar 32 reactivos. Se eligió este valor ya que es la mitad de la cantidad de ítems de la prueba completa.

Responder el TAI así diseñado podría demandar cerca de media hora y estaría en consonancia con lo sugerido por Zickar et al. (1999) de que la evaluación adaptativa no dure más de 60 minutos para que la fatiga no perjudique el rendimiento. Por otro lado, no se fijó un tiempo límite de evaluación para no introducir factores externos (e.g., rapidez de respuesta, familiaridad con el manejo computacional, estrategias de respuesta, características de personalidad) que contaminaran la medición del constructo.

El criterio de parada mixto así definido para el TAI (finaliza cuando el  $EEE(\hat{\theta})$  descienda del valor 0.40 o cuando, no alcanzando este nivel de precisión, la persona responda a 32 ítems) sería evaluado al implementar el TAI.

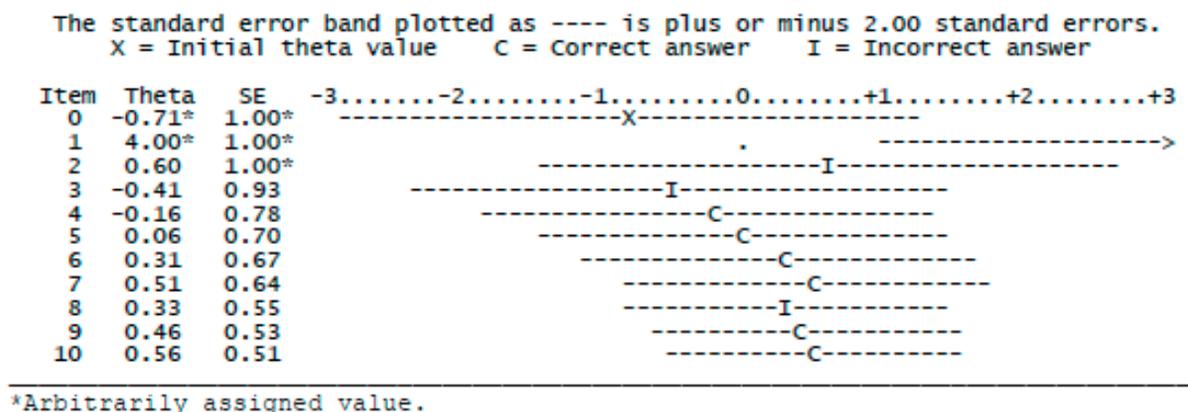
#### *Programación del software y almacenamiento de los ítems del BI base del TAI*

Una vez definidos los componentes del algoritmo adaptativo se realizó la programación del software FastTEST Pro. Ésta también incluyó las especificaciones adicionales necesarias para la implementación práctica del TAI: a) la forma de identificación de cada persona, información que constituirá la denominación de su archivo de datos, b) el modo de ingreso de la repuesta por parte del evaluado, c) el formato de presentación de los ítems y de los ejemplo de prueba, d) las instrucciones del TAI, e) los reportes que se obtendrán para describir el desempeño del evaluado y f) la información que debe archivararse tras la ejecución (Lozzia et al., 2009). Todas estas especificaciones y los ítems del Banco de Analogías Verbales se ingresaron en el FastTEST Pro. Se buscó que el programa tuviera una apariencia sencilla, de manera que la diferencia en experiencias con computadoras no impactaran en los resultados (Abal, Lozzia, Aguerri y Galibert, 2010).

El diseño del TAI en el programa FastTEST Pro se realizó por módulos que luego fueron ensamblados,

Figura 1.

Ejemplo del gráfico de rendimiento en el TAI brindado por el programa FastTEST Pro.



formando una sesión de evaluación completa ya que este software no permitía hacerlo en un solo bloque. Para este TAI se crearon cuatro módulos: instrucciones, test, fin de prueba y reporte. A continuación se describe cómo se procedió en el diseño de cada uno y, finalmente, en el ensamblado de la sesión de evaluación que permitiría administrar el TAI de Analogías Verbales.

*Módulo de instrucciones.* Se plasmó la misma información que contenía la hoja de instrucciones utilizada en las pruebas de calibración (Lozzia et al., 2012), pero añadiendo las instrucciones necesarias para realizar el TAI. A saber: cómo avanzar de pantalla, cómo responder a los ítems, imposibilidad de omitir la respuesta y de retroceder para modificarla y el feedback sobre la resolución de los ejemplos.

*Módulo de fin de prueba.* Esta pantalla sólo podía ser cerrada por el supervisor de la evaluación mediante una clave. Con ello se aseguraba que el evaluado no pudiera entrar a otras aplicaciones o acceder al reporte de resultados sin autorización.

*Módulo del test.* Para crear este módulo, primero se ingresaron al programa FastTEST Pro los 64 ítems que formaban el BI con sus características específicas y se les dio el formato que permitiera su administración informatizada. Para cada ítem se almacenó la siguiente información: identificación, texto del ítem, fecha de creación y autor, pruebas en las que se lo administró y su posición, respuesta correcta, parámetros de la TRI (*a*, *b* y *c*), índices clásicos (dificultad y correlación ítem-test corregido) y notas (incluía datos de la muestra en la que fue calibrado, los resultados del análisis del DIF y otros

índices como el de discriminación entre grupos extremos y su pesaje factorial). Luego se creó el módulo de test. Para ello se seleccionó: a) la lista de ítems que constituiría la base del TAI, b) el tipo de test que se deseaba crear: *Maximum information adaptive* (TAI con selección de ítems por el método de máxima información), c) la estimación inicial de  $\theta$  obtenida al azar dentro del intervalo -1.0 a -0.5, d) el método de máxima verosimilitud condicional para la estimación provisional de  $\theta$  y la final, e) el procedimiento de máxima información de Fisher sin restricciones para la selección adaptativa de los ítems, y f) el criterio de parada mixto: cuando el  $EEE(\hat{\theta})$  desciende del valor 0.40 o cuando, no alcanzando este nivel de precisión, la persona responde a 32 ítems.

*Módulo de reporte.* En primer lugar informa los datos de la sesión de evaluación y del test y luego los datos del evaluado. Brinda un gráfico de líneas de punto del progreso ítem por ítem del examinado a través del TAI. Este gráfico de rendimiento (ver Figura 1) presenta la escala de valores  $\theta$  en un rango de -3 a +3. Bajo ésta, se indica el  $\theta$  inicial asignado (X) y luego los  $\theta$  estimados tras cada respuesta (C ó I según la respuesta dada haya sido correcta o incorrecta). La línea punteada muestra el intervalo correspondiente al  $\theta \pm 2EEE(\hat{\theta})$  (los errores de estimación son llamados por el programa *SE* por *Standard Error*).

A continuación del gráfico de rendimiento, se presentan los resultados totales de la evaluación: el  $\theta$  final estimado con su error de estimación, el intervalo de confianza correspondiente al  $\hat{\theta} \pm 2EEE(\hat{\theta})$  y la cantidad de ítems sobre la cual se basan estos resultados.

Por último, el informe incluye una línea para cada reactivo administrado, siguiendo el orden en que fueron presentados al evaluado.

*La sesión de evaluación.* Una vez que cada uno de los módulos estuvo listo y revisado, se procedió a insertarlos en la sesión de evaluación. Ésta comenzaría con el módulo de instrucción. Continuaría con el test propiamente dicho, es decir, el TAI. Luego de la pantalla que indica que la evaluación ha finalizado, y por medio de la introducción de una clave por parte del evaluador, se presentaría el módulo de reporte que muestra el desempeño del evaluado en el TAI. Una vez ensamblada la sesión de evaluación, se chequeó su funcionamiento. Al comprobar su correcto accionar, se concluyó con la etapa de diseño del TAI. Por tanto, el TAI de Analogías Verbales estaba listo para ser administrado.

### Discusión

Como señalan Olea et al. (2004, p. 519), “el desarrollo de un TAI hasta que se encuentra operativo es un proceso laborioso y exige conocimientos y destrezas técnicas importantes [...] en el contenido sustantivo de la prueba y en informática”. Como se observó en este trabajo, a la hora de diseñar el TAI fue necesario tomar muchas decisiones teóricas, metodológicas y computacionales. Se comenzó con la definición de cada uno de sus componentes y finalizó con la programación del software FastTEST Pro a partir de las especificaciones y criterios seleccionados. El TAI desarrollado utiliza: a) una estrategia de comienzo que selecciona el  $\theta$  inicial aleatoriamente (evita repetir la misma secuencia en diferentes estudiantes) entre niveles levemente inferiores a la media del rasgo para asegurar una primera experiencia satisfactoria que disminuya la ansiedad ante la evaluación, b) el método de máxima verosimilitud condicional para estimar después de cada respuesta el nivel de rasgo y la precisión asociada a dicha estimación, c) la estrategia de máxima información de Fisher para la selección sucesiva de ítems y d) un criterio mixto de finalización al alcanzar un nivel de precisión en la estimación del nivel de rasgo equivalente a un error de estimación menor o igual a 0.4 o, en su defecto, al administrar 32 ítems. Estos valores se eligieron teniendo en cuenta: a) la FI del BI para garantizar que el error fijado como punto de corte pueda ser alcanzado por la mayoría de los evaluados, b) la precisión alcanzada en las versiones convencionales utilizadas en la calibración del BI para fijar el error

de estimación máximo tolerable (el TAI no debe ser menos preciso) y c) el número de ítems administrados en los tests convencionales de calibración para fijar la cantidad máxima de reactivos del TAI (se espera que el TAI, utilizando igual o menor número de ítems, sea igual o más preciso). Los estudios que se realicen con la implementación práctica del TAI indicarán si esta estrategia de finalización es la adecuada.

El objetivo fue obtener una versión adaptativa de la prueba de papel y lápiz de Analogías Verbales que pudiera ser administrada en el contexto universitario. Este trabajo continuará con la obtención de datos empíricos que permitan estudiar la adecuación del algoritmo adaptativo propuesto. Los próximos estudios buscarán aportar información sobre la eficacia y capacidad de pronóstico del TAI, aplicando a un mismo grupo de estudiantes el TAI y el Test Completo de Analogías Verbales compuesto por todos los ítems del BI (van der Linden, 2008). Si el TAI es eficaz debería proporcionar, después de la presentación de un número reducido de ítems, un nivel de habilidad aproximado al que obtendría la persona si respondiera a todos los ítems (Bartram y Hambleton, 2006).

### Referencias

- Abal, F., Lozzia, G., Aguerri, M. y Galibert, M. (2010). La Evaluación mediante Tests Adaptativos Informatizados. Experiencia Subjetiva del Examinado. *Memorias del II Congreso Internacional de Investigación y Práctica Profesional en Psicología, Facultad de Psicología, Universidad de Buenos Aires*, 4, 429-431.
- Abal, F., Lozzia, G., Picón Janeiro, J., Galibert, M. y Aguerri, M. (2007). Dificultades en la difusión y desarrollo de la Teoría de Respuesta al Ítem en Argentina. *Memorias XIV Jornadas de Investigación, Facultad de Psicología, Universidad de Buenos Aires*, 2, 503-505.
- Attorresi, H., Lozzia, G., Abal, F., Galibert, M. y Aguerri, M. (2009). Teoría de Respuesta al Ítem. Conceptos básicos y aplicaciones para la medición de constructos psicológicos. *Revista Argentina de Clínica Psicológica*, 18, 179-188.
- Barrada, J., Olea, J. y Ponsoda, V. (2004). Reglas de selección de ítems en tests adaptativos informatizados. *Metodología de las Ciencias del Comportamiento, Vol. Esp.*, 55-61.
- Bartram, D. y Hambleton, R. (2006). *Computer-based testing and the Internet: Issues and advances*. Chichester, West Sussex: Wiley.
- Embretson, S. y Reise, S. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: LEA.
- Forbey, J. y Ben-Porath, Y. (2007). Computerized adaptive personality testing: a review and illustration with the MMPI-2 Computerized Adaptive Version. *Psychological Assessment*, 19, 14-24.
- Galibert, M., Aguerri, M., Pano, C., Lozzia, G. y Attorresi, H. (2005). Análisis de Ítem de Analogías Verbales mediante la Aplicación de un Modelo Político de la Teoría de Respuesta al Ítem. *Revista Mexicana de Psicología*, 22, 419-431.
- Goswami, U. (2001). Analogical reasoning in children. En D. Gentner, K. Holyoak y B. Kokinov (Eds.), *Perspectives on thought and*

- language: *Interrelations in development* (pp. 225-277). London: Cambridge University Press.
- Hol, A., Vorst, H. y Mellenbergh, G. (2008). Computerized adaptive testing of personality traits. *Zeitschrift für Psychologie/ Journal of Psychology*, 216, 12-21.
- Kuncel, N., Hezlett, S. y Ones, D. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, 86, 148-161.
- Lohman, D. (2004). Aptitude for College: The Importance of Reasoning Tests for Minority Admissions. En R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp 41-55). London: Falmer Press.
- Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: LEA.
- Lord, F. (1986). Maximum likelihood and Bayesian parameter estimation in item response theory. *Journal of Educational Measurement*, 23, 157-162.
- Lozzia, G., Abal, F., Aguerri, M., Galibert, M. y Attorresi, H. (2006). Presentación de una Base de Datos para la gestión de un Banco de Ítems de Analogías Verbales. *Investigaciones en Psicología*, 11, 67-82.
- Lozzia, G., Abal, F., Blum, G., Aguerri, M., Galibert, M. y Attorresi, H. (2009). Tests Informatizados. Nuevos desafíos prácticos y éticos para la Evaluación Psicológica. *SUMMA Psicológica UST*, 6, 135-148.
- Lozzia, G., Abal, F., Blum, G., Aguerri, M., Galibert, M. y Attorresi, H. (2012). Test Adaptativo Informatizado de Analogías Verbales: Construcción del Banco de Ítems. *Revista Mexicana de Psicología*. Enviado.
- Lozzia, G., Galibert, M., Aguerri, M. y Attorresi, H. (2005). Construcción de un Banco de Ítem de Razonamiento Verbal. *Interdisciplinaria*, 22, 5-27.
- Lozzia, G., Picón Janeiro, J. y Galibert, M. (2008). La Evaluación del Razonamiento Verbal mediante el Formato de Analogías Verbales. *Memorias de las XV Jornadas de Investigación, Facultad de Psicología, Universidad de Buenos Aires*, 2, 474-476.
- Martínez Arias, R. (1995). *Psicometría: Teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- Muñiz, J. (1997). *Introducción a la Teoría de Respuesta a los Ítems*. Madrid: Pirámide.
- Olea, J., Abad, F., Ponsoda, V. y Ximénez, M. (2004). Un test adaptativo informatizado para evaluar el conocimiento del inglés escrito: diseño y comprobaciones psicométricas. *Psicothema*, 16, 519-525.
- Olea, J. y Ponsoda, V. (1996). Tests adaptativos informatizados. En J. Muñiz (Ed.), *Psicometría* (pp. 730-783). Madrid: Universitat.
- Olea, J. y Ponsoda, V. (2003). *Tests adaptativos informatizados*. Madrid: UNED.
- Owen, R. (1975). A bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association*, 70, 351-356.
- Renom, J. y Doval, E. (1999). Tests Adaptativos informatizados: Estructura y desarrollo. En J. Olea, V. Ponsoda y G. Prieto (Eds.), *Tests informatizados: Fundamentos y aplicaciones* (pp. 127- 162). Madrid: Pirámide.
- Segall, D. y Moreno, K. (1999). Development of the Computerized Adaptive Testing Version of the Armed Service Vocational Aptitude Battery. En F. Drasgow y J. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (pp. 35-65). Mahwah, NJ: LEA.
- Sternberg, R. (1988). *The Triarchic Mind: A New Theory of Human Intelligence*. NY: Viking.
- Sternberg, R. (2001). *How to Prepare for the MAT-Müller Analogies Test*. NY: Barron's Educational Series.
- Thissen, D. y Mislevy, R. (2000). Testing algorithms. En H. Wainer (Ed.), *Computerized adaptive testing: A primer* (2ª ed., pp. 101-133). Mahwah, NJ: LEA.
- Tornimbeni, S., Pérez, E. y Olaz, F. (2008). *Introducción a la psicometría*. Buenos Aires: Paidós.
- van der Linden, W. (2008). Some new developments in adaptive testing technology. *Zeitschrift für Psychologie / Journal of Psychology*, 216, 3-11.
- van der Linden, W. y Glas, C. (2007). Statistical aspects of adaptive testing. En C. Rao y S. Sinharay (Eds.), *Handbook of statistics* (Vol. 27: Psychometrics, pp. 801-838). Amsterdam: North-Holland.
- van der Linden, W. y Pashley, P. (2000). Item selection and ability estimation in adaptive testing. En W. van der Linden y C. Glas (Eds.), *Computerized adaptive testing. Theory and practice* (pp 1-25). Dordrecht: Kluwer Academic Publishers.
- Wainer, H., Dorans, N., Eignor, D., Flaugher, R., Green, B., Mislevy, R., Steinberg, L. y Thissen, D. (2000). *Computerized Adaptive Testing: A Primer* (2ª ed.). Mahwah, NJ: LEA.
- Weiss, D. (2008). *Manual for the FastTEST Professional Testing System, Version 2*. St. Paul, MN: Assessment Systems Corporation.
- Xing, D. y Hambleton, R. (2004). Impact of Test Design, Item Quality, and Item Bank Size on the Psychometric Properties of Computer-Based Credentialing Examinations. *Educational and Psychological Measurement*, 64, 5-21.
- Yela, M. (Ed.). (1987). *Estudios sobre inteligencia y lenguaje*. Madrid: Pirámide.
- Zickar, M., Overton, R., Taylor, L. y Harms, H. (1999). The Development of a Computerized Selection System for Computer Programmers in a Financial Service Company. En F. Drasgow y J. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (pp. 7-33). Mahwah, NJ: LEA.

