

Aplicaciones de la tecnología a la evaluación psicométrica

Applications of Technology to Psychometric Evaluation

Andrés Alberto Burga León 

Universidad de Lima, Lima, Perú

ORCID: <http://orcid.org/0000-0003-0388-4238>

Recibido 15-05-19 **Revisado** 05-08-19 **Aprobado** 29-10-19 **En línea** 11-11-19

*Correspondencia

Email: andresburgaleon@gmail.com

Citar como:

Burga León, A. (2019). Aplicaciones de la tecnología a la evaluación psicométrica. *Propósitos y Representaciones*, 7(SPE), e318. doi: <http://dx.doi.org/10.20511/pyr2019.v7nSPE.318>

Resumen

El presente trabajo tiene como objetivo señalar la importancia y aplicabilidad que tiene la tecnología y en especial el uso de la computadora, en la evaluación psicométrica. Para ello, se presentan, de manera introductoria, tres temas que a juicio del autor son algunos de los más destacados. Específicamente se presentan los aspectos centrales de los ítems mejorados con tecnología, los bancos de ítems y los test adaptativos computarizados. En cuanto a los ítems mejorados con tecnología, se enfatiza su importancia para ampliar la representatividad y fidelidad con la cual se trata de medir un constructo, señalando diferentes recursos impresos y en internet que pueden ser consultados para ver más ejemplos de este tipo de ítems. Al tratar el tema del banco de ítems se resalta el rol que cumplen en la actualidad, que va mucho más allá de ser un almacén organizado de ítems, para convertirse en un verdadero sistema informático que permite elaborar, revisar y aplicar los ítems. Finalmente, se presentan los aspectos principales de un test adaptativo computarizado, que teniendo como base un banco de ítems, permite en tiempo real ir estimando la medida de un rasgo latente y aplicar ítems adecuados según dicha estimación. También se hace referencia a una alternativa vinculada al software libre, que permite implementar un test adaptativo computarizado.

Palabras clave: Tecnología; Aplicación informática; Evaluación; Psicometría.

Summary

The objective of this work is to point out the importance and applicability of technology, especially the use of computers, in psychometric evaluation. To this end, three topics are presented, in an introductory manner, which, in the author's opinion, are some of the most outstanding. Specifically, the central aspects of items improved with technology, item banks and computerized adaptive tests are presented. As for technology-enhanced items, emphasis is placed on their importance for broadening the representativeness and fidelity with which a construct is measured, pointing out different printed and internet resources that can be consulted to see more examples of this type of items. When dealing with the issue of the item bank, the role that they currently play is highlighted, which goes far beyond being an organized store of items, to become a true computer system that allows the items to be elaborated, reviewed and applied. Finally, the main aspects of an adaptive computerized test are presented. Based on a bank of items, it allows, in real time, to estimate the measurement of a latent feature and apply appropriate items according to said estimate. Reference is also made to an alternative linked to free software, which allows the implementation of an adaptive computerized test.

Key words: Technology; Computer Applications; Assessment; Psychometrics.

Introducción

Es indudable que los avances tecnológicos que experimentamos en estos tiempos son cada vez mayores. Es usual encontrar cientos de noticias y páginas web que nos hablan sobre los avances tecnológicos y su impacto en la sociedad. Si bien es muy difícil ponerse de acuerdo en cuál es el avance tecnológico más importante de los últimos 100 años, es innegable el enorme impacto que la computadora y la informática en general han tenido en la sociedad. Recordemos que una de las primeras computadoras electrónicas fue denominada ENIAC (Electronic Numerical Integrator and Computer), cuyo desarrollo empezó en la universidad de Pennsylvania en 1943 y estuvo a cargo de John Mauchly y Presper Eckert, quienes la culminaron en 1946 con un costo de casi medio millón de dólares (Benov, 2014).

Si bien las computadoras han impactado en diferentes esferas de nuestras vidas, la psicología no es ajena a ello. De este modo, en 1971 se crea la *Society for Computers in Psychology*¹, cuyo propósito principal es incrementar y difundir el conocimiento sobre el uso de la computadora en la investigación psicológica. Dicha sociedad se ha centrado en desarrollar

temas como: modelos computacionales sobre el procesamiento cognitivo y la conducta, herramientas computacionales para el análisis y la recolección de datos, interacción persona – computador, representación del conocimiento en humanos y máquinas, aprendizaje de máquinas, métodos y herramientas para la investigación basada en internet y uso de la tecnología para mejorar la evaluación.

Es por este interés en el uso de la tecnología dentro de la psicología que se desarrollan eventos como el International Computer Assisted Assessment (CAA) Conference², evento dedicado al *e-assessment* cuya primera edición fue lanzada hace más de 20 años. Tal y como se puede apreciar en la página web de dicha conferencia, en el año 2016 cambia su nombre a Technology Enhanced Assessment, cuya edición del año 2018 tuvo como objetivo presentar el estado del arte y diferentes perspectivas sobre la concepción, desarrollo e implementación de la evaluación mejorada con tecnología, en diferentes sectores como la educación primaria, secundaria y superior, así como su contribución al aprendizaje a nivel profesional e informal.

Por todo lo antes mencionado, es que encontramos en la literatura términos similares como *Computer Assisted Assessment*, *Computer Aided Assessment* o *Technology Enhanced Assessment*, utilizados para destacar el rol que la computadora tiene en la simulación y evaluación de variables que de otra manera serían difíciles de cuantificar (Weiss, 2013). Es innegable el rol destacado que cumple la computadora en la administración, calificación e interpretación de resultados; buscando fundamentalmente el beneficio de la persona evaluada. Al respecto, Cohen y Swerdlik (2009) destacan algunas ventajas que trae el uso de la tecnología, como un menor tiempo transcurrido entre la aplicación del instrumento de evaluación y el reporte de resultados (puede ser incluso inmediata); evita los errores de calificación propios de los seres humanos; posibilita la aplicación de diversas medidas de seguridad de tipo tecnológico (uso de contraseñas, cifrado de datos), en contraste con las medidas tradicionales que implicaban básicamente mantener los test bajo llave; posibilidad de adaptar el contenido y longitud del test según las características de quien lo responde. Estos mismos autores no dejan de señalar que, a pesar de dichas ventajas, también existen algunos inconvenientes, como el tener que dedicar un tiempo considerable a la lectura de la documentación del software y hardware o libros complementarios sobre la interpretación del test; la posibilidad de que exista un error de software o hardware el cual puede provenir de fuentes difíciles de detectar como un *glitch*; la posibilidad de hackear la seguridad de este tipo de test; pérdida de datos por eventos circunstanciales como los virus de computadora; además no todos tienen la misma experiencia de aplicación del test y puede resultar difícil, para las personas no familiarizadas con el tema, comprender que existen los resultados comparables, a pesar de haberse aplicado diferentes ítems.

Si bien son varios los temas que pueden tratarse al escribir sobre las aplicaciones de la computadora en la evaluación psicológica (con énfasis en el enfoque psicométrico), nos centraremos en tres de ellos: los ítems mejorados con tecnología, el banco de ítems y los test adaptativos computarizados.

Ítems mejorados con tecnología

Inicialmente, este tipo de ítems fueron denominados formatos innovadores (Sireci & Zenisky, 2006), pero en muchos casos son formatos preexistentes que han sido mejorados gracias a la tecnología, permitiendo la inclusión de aspectos como gráficos de alta calidad, audio, video e hipervínculos; por lo cual el nombre sugerido, últimamente, por Sireci y Zenisky (2016) es ítems mejorados con tecnología. En este sentido, es importante considerar que este no es el único nombre con el que se les conoce, pues también se les llama facilitados por la tecnología, computarizados innovadores o digitales.

La gran ventaja al utilizar la tecnología proporcionada por una computadora es que gracias a ello se pueden realizar actividades que son más similares a situaciones de la vida real, mucho mejor de lo que permite el formato de papel y lápiz, facilitando además su calificación

(Sireci & Zenisky, 2006). También presentan como ventaja una menor probabilidad de acertar por azar (depende del tipo de ítem); además de favorecer el involucramiento de los evaluados con la tarea, pues la consideran más auténtica (Bryant, 2017). Otro aspecto positivo que señalan Strain-Seymour, Way y Dolan (2009), es que en general proveen información diagnóstica más rica: pues pueden mostrar no solo el producto, sino además el proceso de resolución seguido.

Dado este formato computarizado, se produce una nueva forma de interacción entre la persona y el instrumento de evaluación, llevando a que se deba considerar, tal y como lo señalan Sireci y Zenisky (2006), el dispositivo de entrada y la acción necesaria para resolver un ítem. En ese sentido, los dispositivos de entrada más usados son el ratón (*mouse*) y el teclado, aunque existen otras posibilidades como la pantalla táctil, cámara, lápiz óptico, palanca de mando (*joystick*), mando de videojuegos (*gamepad*), etc. En cuanto a las acciones de respuesta Sireci y Zenisky (2006) señalan que el ratón permite desplazar el cursor y hacer clic en lugares específicos que desencadenan acciones (seleccionar, desplegar una lista), coger y arrastrar, marcar un punto de inserción de texto. Por su parte, el teclado permite ingresar texto, números, caracteres especiales o desplazarse por la pantalla, aunque consideramos que esta función se realiza de forma más eficiente con el botón de desplazamiento (*scroll*) del ratón.

Estas nuevas características de los ítems hacen que se desarrollen taxonomías alternativas para clasificarlos, tal y como lo hacen Parshall, Harmes, Davey y Pashley (2010) quienes consideran aspectos como: (a) Estructura de evaluación, que implica el tipo de ítems, que pueden ser de selección o construcción de respuesta; (b) acción de respuesta, que se refiere a la acción física requerida para responderlo y la interfaz utilizada para dar la respuesta; (c) inclusión de medios, pudiendo ser solo texto o involucrar gráficos, audio, video, animaciones; (d) interactividad, que es el grado en el cual un ítem responde ante una acción del usuario; (e) complejidad, que está dada por el tipo de tareas que deben realizarse para resolver un ítem; (f) fidelidad, que es el grado con el cual reproducen actividades y situaciones propias del constructo de interés; y (g) Método de calificación, que puede ser dicotómica, politómica, compleja (definir criterios de respuesta).

Veamos ahora algunos ejemplos de este tipo de ítems mejorados con tecnología. Al respecto, tal y como se señaló con anterioridad, es posible pasar a formato computarizado los ítems tradicionales, como los de opción múltiple. La diferencia radica en que la alternativa de respuesta suele ser seleccionada mediante un clic con el ratón:

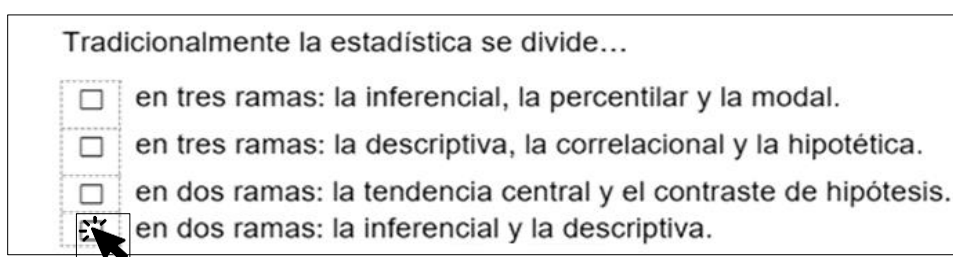


Figura 1. Ejemplo de ítem tipo opción múltiple, presentado por computadora.

Sin embargo, esa no es la única alternativa para presentar ítems de opción múltiple. Ese mismo ítem presentado en la Figura 1, puede estar elaborado en un formato tipo lista desplegable, a la cual se accede mediante un clic con el ratón:

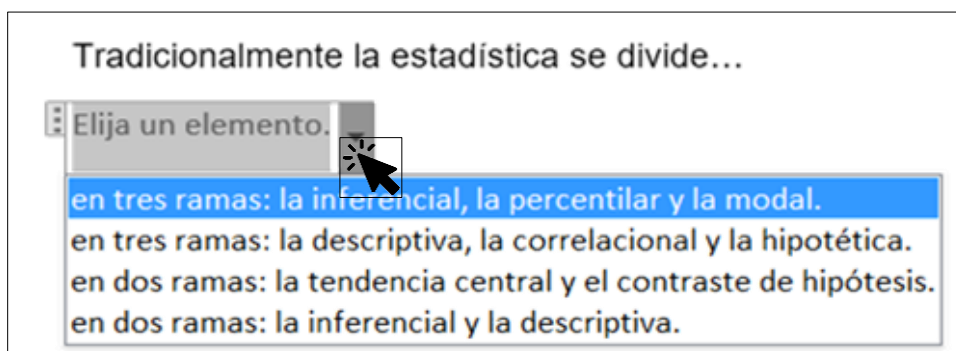


Figura 2. Ejemplo de ítem tipo opción múltiple, usando una lista desplegable.

Por supuesto, la riqueza del formato innovador de ítems va mucho más allá de lo presentado en las Figuras 1 y 2, pues existen muchas posibilidades como la presentada a continuación:

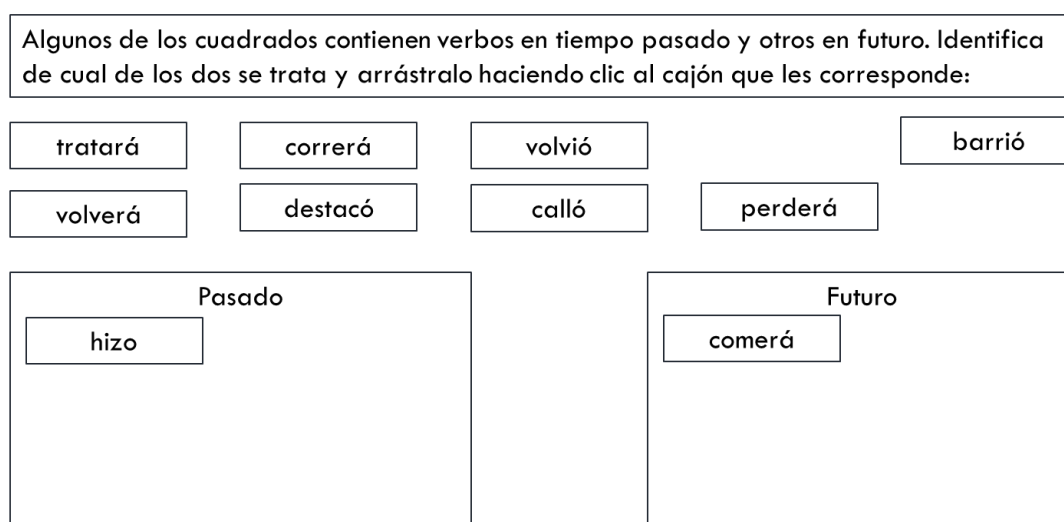


Figura 3. Ejemplo de ítem mejorado con tecnología, que se basa en la clasificación de elementos usando las opciones de arrastrar y soltar.

El lector interesado en ver más ejemplos sobre este tipo de ítems puede consultar la obra de Sireci y Zenisky (2016). Si bien se considera que el capítulo escrito por dichos autores es de una calidad excelente, una limitación es que los ítems computarizados aparecen en formato impreso. Por ello, si se desea una versión real de este tipo de ítems, se puede utilizar alguno de los recursos en línea señalados a continuación:

QTIWorks³ Es una plataforma en código abierto que permite manejar y aplicar ítems mediante una computadora bajo el estándar QTI (Question & Test Interoperability) v2.1.

The Mathematics Common Core Toolbox⁴. Permite apreciar diversos ejemplos de ítems en matemáticas bajo el Common Core State Standards (CCSS). El CCSS es una iniciativa conjunta de 48 estados norteamericanos que hay desarrollado un conjunto claro de estándares en las áreas de Matemática y Lengua (inglés), literatura y artes, desde preescolar hasta décimo segundo grado. También se pueden encontrar ejemplos de ítems mejorados con tecnología⁵.

Shared Common Core Questions⁶. Siguiendo también los Common Core State Standards presentan diferentes ejemplos de ítems para evaluar matemática y lenguaje en este hipervínculo.

*Assessments the next generation*⁷. Este sitio web considera que cada vez más los estudiantes se están enfrentado a ítems mejorados con tecnología, por lo cual necesitan adquirir nuevas habilidades vinculadas a la resolución de dichos ítems.

*Sitio web de Kathleen Scalise*⁸. Trabajó en la Universidad de Oregon, desarrollando una taxonomía basada en la categorización de 28 tipos distintos de ítems aplicados mediante computadora.

A pesar de los múltiples ejemplos que pueden encontrarse sobre este tipo de ítems, autores como Bryant (2017) señalan que aún hay algunos temas que deben ser investigados con mayor detalle. Por ejemplo, ¿qué tipos de ítems implican qué capacidades cognitivas, de tal manera que son más eficientes para evaluar ciertos contenidos?; ¿cómo se comparan las propiedades psicométricas de estos ítems con otros tipos de ítems?; ¿qué evidencias hay de su funcionamiento diferencial considerando el nivel socioeconómico (vinculado con el acceso a tecnología)? En esta misma línea, una preocupación que aparece es que la literacidad computacional puede ser un elemento de varianza irrelevante para el constructo. Sin embargo, si hay buenos tutoriales, este no parece ser un aspecto preocupante (Sireci & Zenisky, 2016).

Para concluir esta sección, se sugiere al lector interesado en conocer algunas recomendaciones para elaborar este tipo de ítems, revisar el artículo publicado por Parshall y Harmes (2009).

Banco de ítems

Cuando escuchamos la palabra banco, usualmente pensamos en una entidad financiera, que se encarga, entre otras cosas, de almacenar nuestro dinero. Y es precisamente en esta acepción que puede definirse qué es un banco de ítems. Al respecto Vale (2006) lo considera como un repositorio organizado de ítems; almacena información pertinente sobre dichos ítems; ligada a la calibración de estos mediante un modelo de Teoría de Respuesta al Ítem (Molina, Pareja & Sanmartín, 2008), la cual luego permite construir test, ya sean comunes a todas las personas evaluadas (lineales), o adaptativos. En este sentido, es importante considerar que, en un banco, los ítems están organizados de una manera lógica (Vale, 2006), lo cual permite recuperarlos considerando diversos filtros vinculados a sistemas de estructuración transversales (comunes a todos los ítems) o anidados (específicos de ciertos ítems). Respecto a esta organización de los ítems, es importante señalar que no hay una sólo forma de estructurarlos; hay que buscar una manera funcional y consensuada con los propietarios del banco, considerando el uso específico que se le va a dar.

Entre las diversas características de los ítems que se utilizan para almacenarlos y luego recuperarlos mediante filtros, pueden señalarse las siguientes (Muckle, 2016; Vale, 2006): (a) identificador, usualmente un código alfanumérico; (b) clasificación, como estructura jerárquica del ítem, por ejemplo: nutrición / grupos de alimentos / grasas / grasas saturadas; (c) ancestros, se refiere a los ítems que han servido de base para elaborarlo; (d) tipo de ítem, que puede ser opción múltiple, para relacionar, respuesta corta, etc.; (e) respuesta correcta o rúbrica de calificación; (f) racionalidad de las alternativas en el caso de los ítems de opción múltiple; (g) parámetros según un modelo de teoría de Respuesta al Ítem, como: dificultad, discriminación, pseudoadivinación, umbrales; (h) propiedades según la Teoría Clásica de los Tests, como: tasas de acierto, funcionamiento de distractores, etc.; (i) elaborador, que incluye fecha y datos de contacto u otra información de identificación; (j) revisor, que incluye fecha y datos de contacto u otra información de identificación; (k) historial de aplicación: fechas y operativos; entre otros que se consideren pertinentes.

En cuanto al desarrollo e implementación de un banco de ítems, actualmente se está tratando de promocionar un estándar denominado Question and Test Interoperability (QTI), desarrollado por IMS Global Learning Consortium⁹. Este constituye un modelo de datos,

especificado mediante XML (eXtensible markup language), que se utiliza para crear y compartir información en la web y otros medios digitales. De este modo, se busca que los ítems puedan ser intercambiables entre diferentes sistemas informáticos. Además, se coloca cada vez más énfasis en el manejo de los bancos de ítems a través de internet (*web-based ítem banking tools*, WIT), pues así son compatibles con la mayoría de los navegadores de internet disponibles y no requieren la instalación de programas adicionales, salvo Java o Silverlight (Muckle, 2016).

Si bien la función original de un banco de ítems estaba orientada hacia su almacenamiento, en la actualidad su funcionalidad se amplía a la construcción y revisión de ítems, además de la aplicación de estos. Por ello, la especificación de un programa para manejar un banco de ítems puede incluir lo siguiente (Molina, Pareja & Sanmartín, 2008; Muckle, 2016; Vale, 2006; Weiss, 2013): (a) almacenamiento de diversos medios digitales (imágenes, video, audio) y la metadata que permitirá luego recuperar los ítems; (b) almacenamiento de las especificaciones del test; (c) edición y visualización, incluyendo la inserción de caracteres especiales y ecuaciones; (d) seguimiento de los cambios realizados a los ítems (historial); (e) búsqueda y selección de ítems mediante operadores lógicos, considerando diversos campos; (f) exportación de búsquedas (a Excel u otros similares), dando información sobre características de los ítems; (g) exportación de ítems para armar pruebas de lápiz y papel (p. ej. pdf, Word); (h) aplicación y recolección de respuestas a ítems; y (i) calibración de ítems.

Además, es importante considerar la seguridad del banco de ítems, aplicando alguna estrategia como las señaladas por Muckle (2016) y Vale (2006), las cuales implican tener un acceso limitado mediante contraseñas, tarjetas o dispositivos electrónicos o biometría; la desconexión de usuarios que permanecen inactivos por algún tiempo; y, finalmente, el almacenamiento de información encriptada, llamada también cifrado de archivos.

Se finaliza esta sección, señalando que, si bien existen diversas alternativas de software comercial para implementar un banco de ítems, siempre es posible diseñar un sistema propio, como lo tiene planificado la Oficina de Medición de la Calidad de los Aprendizajes (UMC) del Ministerio de Educación del Perú.

Test adaptativos computarizados

Desde nuestra perspectiva, los test adaptativos computarizados (TAC) son una de las aplicaciones más interesantes y útiles que la tecnología aporta a la evaluación psicométrica. De manera general, implican que mientras se aplica el test computarizado, los ítems son adaptados a la persona que los responde (Molina, Pareja & Sanmartín, 2008; Piton-Gonçalves & Aluísio, 2015). Para ello, se utilizan algoritmos complejos, asociados a los modelos de Teoría de Respuesta al Ítem para estimar la medida del rasgo latente de la persona evaluada y seleccionar los ítems a aplicar en tiempo real (van der Linden & Pashley, 2010). Es decir, la secuencia de ítems administrados depende de las respuestas a los ítems anteriores. Por ello, para aplicar de modo eficiente este tipo de test, se necesita contar con un buen banco de ítems, tarea que no es sencilla. Sin embargo, a pesar de la dificultad que puede implicar el armado de un banco de ítems, son innegables las ventajas que se encuentran asociadas al uso de test adaptativos computarizados, entre las cuales podemos señalar las siguientes (Kingston, Scheuring & Kramer, 2013; Weiss, 2013): (a) se pueden utilizar y calificar tipos de ítems que no son tan fáciles de aplicar o son muy costosos en formato impreso; (b) evitan aplicar ítems no informativos, que son demasiado fáciles o difíciles para las personas evaluadas y por lo tanto no contribuyen con información para estimar la cantidad del rasgo o atributo latente que poseen; y (c) en general, son más confiables que un test lineal, considerando el número de ítems y el tiempo de aplicación.

Para ilustrar el proceso general que sigue un TAC, podemos usar como apoyo la figura 4. En primer lugar, se da una estimación inicial de la medida del rasgo latente de una persona y su error estándar de medición. Considerando dicha estimación inicial, la computadora, aplicando diferentes algoritmos, selecciona el siguiente ítem a aplicar. Usando la información obtenida

mediante ese nuevo ítem, se vuelve a estimar la medida del rasgo latente y su error de medición. Considerando esa nueva estimación del rasgo latente se selecciona un nuevo ítem, cuya información se agregará a la preexistente, para estimar la medida del rasgo latente. Este proceso iterativo seguirá hasta que se cumpla con una regla de finalización, la cual se encuentra implementada dentro del proceso del TAC. Además, es importante resaltar, tal y como aparece en la figura 4, que uno de los objetivos buscados con un TAC es ir reduciendo el error de medición conforme se cuenta con más información (más ítems aplicados) para estimar la medida del rasgo latente.

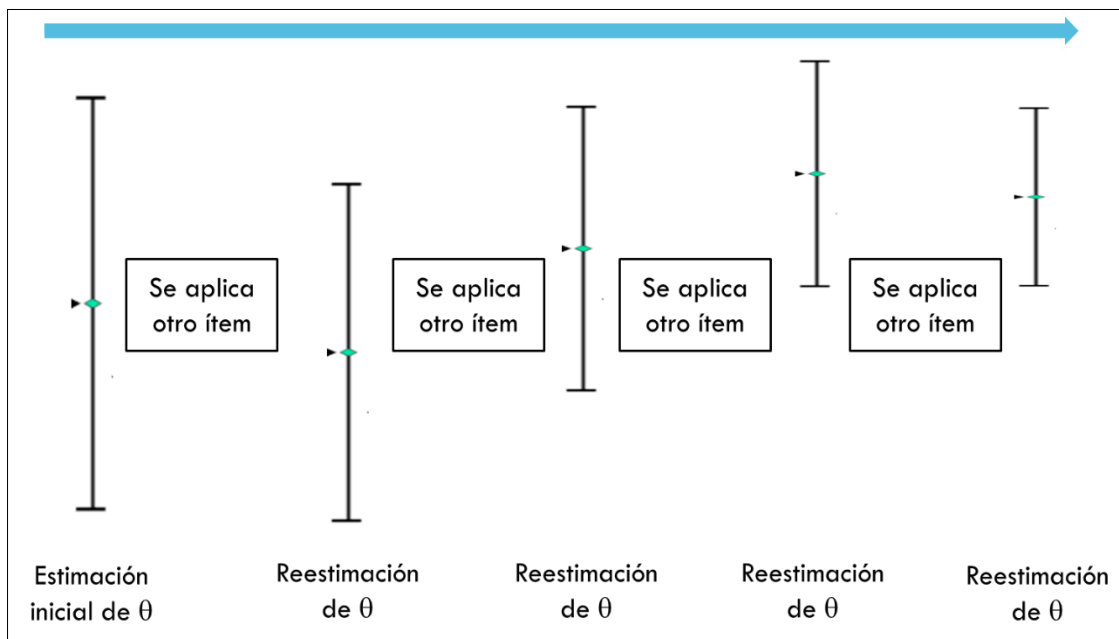


Figura 4. Proceso de estimación del rasgo latente seguido en un test adaptativo computarizado.

Como acaba de ilustrarse, tal y como ocurre con cualquier instrumento de medición psicométrica, se busca obtener un número que estime de la mejor manera posible la cantidad de un rasgo latente de interés. Por ello, en el proceso que sigue un TAC hay tres fases de estimación del rasgo latente (van der Linden & Pashley, 2010):

Inicial. Se realiza antes de empezar la aplicación adaptativa de los ítems contenidos en el banco. Para esta estimación inicial se puede fijar en un valor bajo o alto la medida del rasgo latente, aplicar un test lineal o esperar a que se tenga un número mínimo de ítems aplicados. A pesar de las diferentes alternativas, lo bueno es que una estimación inicial sesgada solo es un problema si se trata de un test corto; con 20 o más ítems el error inicial de estimación tiende a corregirse.

Proceso. Hay varios algoritmos que realizan la estimación de la habilidad de las personas (MLE, WLE, EAP, MAP) y la selección de los ítems (usualmente información máxima de Fisher, aunque hay otras alternativas: Kullback–Leibler, información ponderada por la verosimilitud, métodos bayesianos). Al respecto, van der Linden (1998) encontró que los métodos bayesianos resultan ser bastante eficientes.

Final. se establecen reglas para culminar la aplicación, pues de lo contrario se podrían aplicar incluso todos los ítems del banco. Usualmente esas reglas están vinculadas con el logro de cierta precisión en la estimación del rasgo latente (θ). Además, el resultado usualmente es transformado (total de respuestas correctas en un test lineal estándar, percentiles, nivel de logro) para hacerlo más interpretable.

En cuanto a las experiencias existentes referidas al uso de los test adaptativos computarizados, podemos, entre otras, señalar las siguientes:

- School Graduation Exams (SGEs): desarrollado en Georgia por el National Assessment and Examinations Center (NAEC). Evalúa las áreas de: Lengua y literatura, Inglés, Historia y geografía, Química, Biología, Física y Matemáticas (Bakker, 2014).
- Graduate Management Admission Test (GMAT) implementado por el Graduate Management Admission Council, busca ser una herramienta que ayude al proceso de admisión a las escuelas de administración en Estados Unidos (Rudner, 2010).
- MATHCAT: implementado en Holanda para evaluar capacidades matemáticas en cursos orientados hacia la educación de adultos (Verschoor & Straetmans, 2010).
- Uniform CPA Exam: aplicado por la Association of International Certified Professional Accountants (AICPA) para obtener la licencia profesional (2010).
- National Council Licensure Examination (NCLEX): desarrollado por el National Council of State Boards of Nursing para otorgar la licencia profesional en enfermería, tanto en Estados Unidos como Canadá (<https://www.ncsbn.org/1216.htm>).
- Patient-Reported Outcomes Measurement Information System (PROMIS): patrocinado por el U.S. National Institutes of Health (NIH) (Walter, 2010).
- ESQUIZO-Q: está basado en el Cuestionario de Oviedo para la evaluación de la esquizotipia (Fonseca-Pedrero, Menéndez, Paino, Lemos-Giráldez & Muñíz, 2013).

Además de las experiencias de uso antes señaladas, es importante saber que desde el año 2010 existe la International Association for Computerized Adaptive Testing [IACAT]¹⁰, Desde su conformación, dicha organización se encarga, entre otras cosas, de organizar eventos académicos, siendo el último convocado la 2019 IACAT Conference, a realizarse en la ciudad de Minneapolis. También tiene a su cargo el Journal of Computerized Adaptive Testing que empezó a publicarse en el año 2012¹¹.

Por último, debe señalarse que, si bien existen diferentes alternativas comerciales para implementar un test adaptativo computarizado, es posible considerar alternativas de código abierto, como Concerto, que ha sido desarrollado por el Centro de Psicometría de la Universidad de Cambridge¹². Esta plataforma combina el lenguaje de programación R, las bases de datos en MySQL y el uso de HTML. Además, ponen a disposición información adicional, incluyendo toda la documentación y tutoriales para su uso¹³.

Conclusiones

En este artículo se ha señalado que uno de los aportes más importantes que el avance tecnológico le ha dado a la evaluación psicométrica es el uso de las computadoras en los procesos de evaluación. Específicamente se han destacado tres aspectos referidos a ello: el uso de ítems mejorados con tecnología, los bancos de ítems y los test adaptativos computarizados.

En cuanto a los primeros se destacan las nuevas posibilidades que se abren al permitir evaluar capacidades que de otro modo no serían factibles de cuantificar. Por ello, este tipo de ítems se han aplicado mucho en contextos de evaluación orientada a la acreditación profesional (Bryant, 2017): licencia médica, contabilidad, diseño arquitectónico, manejo de pacientes, ciencias de la computación, etc.). También son bastante útiles en la evaluación formativa, trabajándose en forma de simulaciones, mundos virtuales, juegos, redes sociales, realidad aumentada (Sireci & Zenisky, 2016).

Al hablar sobre los bancos de ítems se puso énfasis en las grandes posibilidades que brindan en la actualidad, no solo en términos de ser un almacén organizado, sino en cuanto a las diversas funcionalidades que presentan ahora, vinculadas a la elaboración, revisión y aplicación

de ítems (Muckle, 2016). Esto contribuirá a elaborar de manera más eficiente los ítems que conforman los diferentes instrumentos de medición.

Es precisamente la disponibilidad de un buen banco de ítems calibrados, lo que hace posible que existan los test adaptativos informatizados. Mediante ellos, se puede realizar una aplicación más eficiente de las pruebas psicométricas, logrando una reducción del error estándar de medición (van der Linden & Pashley, 2010). Esto es posible gracias a que de forma iterativa van reestimando la medida de un rasgo latente y aplicando nuevos ítems, adaptados a dicha medida.

Finalmente, aunque este no ha sido un tema tratado en el presente artículo, se considera que otro aporte que trae consigo el uso de la tecnología es que se pueden aplicar nuevos (y más complejos) modelos psicométricos (véase el trabajo de Kyllonen & Zu, 2016), gracias a que las computadoras facilitan entre otras cosas, registrar el tiempo de reacción, además de la respuesta emitida.

Notas

¹ <http://www.scip.ws/>

² <http://www.teaconference.org/en/presentation/>

³ <https://webapps.ph.ed.ac.uk/qtiworks/anonymous/simples/>

⁴ <http://www.corestandards.org/about-the-standards/frequently-asked-questions/>

⁵ http://ccsstoolbox.agilemind.com/parcc/PARCCPrototype_main.html/

⁶ <https://edcite.com/ccssquestionstree/>

⁷ <http://nextgen.apps.sparcc.org/home/>

⁸ <http://pages.uoregon.edu/kscalise/taxonomy/taxonomy.html/>

⁹ <https://www.imsglobal.org/>

¹⁰ <http://www.iacat.org/>

¹¹ <http://iacat.org/jcat/index.php/jcat/index/>

¹² <http://concertplatform.com/>

¹³ <https://www.psychometrics.cam.ac.uk/newconcerto/>

Referencias

- Bakker, S. (2014). *The introduction of large scale computer adaptive testing in Georgia. Political context, capacity building, implementation, and lessons learned*. Recuperado de http://siteresources.worldbank.org/INTREAD/Resources/Bakker_Introduction_to_CAT_Georgia_for_READ.pdf
- Beker, K. A., & Bergstrom, B. A. (2013). Test administration models. *Practical Assessment, Research & Evaluation*, 18(14), 1-7. Recuperado de <https://pareonline.net/pdf/v18n14.pdf>
- Benov, D. M. (2014). The Manhattan Project, the first electronic computer and the Monte Carlo method. *Monte Carlo Methods and Applications*, 22(1), 73-79. doi: <https://doi.org/10.1515/mcma-2016-0102>
- Bryant, W. (2017). Developing a strategy for using technology-enhanced items in large-scale standardized tests. *Practical Assessment Research and Evaluation*, 22(1), 1-10. Recuperado de <http://pareonline.net/getvn.asp?v=22&n=1>
- Cohen, R. J., & Swerdlik, M. (2009). *Psychological testing and assessment: an introduction to tests and measurement*. Columbus, OH: McGraw-Hill.
- Fonseca-Pedrero, E., Menéndez, L. F., Paino, M., Lemos-Giráldez, S., & Muñiz, J. (2013). Development of a Computerized Adaptive Test for schizotypy assessment. *PLOS ONE*, 8(9), 1-9. doi: <https://doi.org/10.1371/journal.pone.0073201>
- Kingston, N. M., Scheuring, S. T., & Kramer, L. B. (2013). Test development strategies. En K. F. Geisinger (Ed.), *APA Handbook of testing and assessment in psychology. Volume 1: Test theory and testing and assessment in industrial and organizational psychology* (pp. 165-184). Washington, D. C.: American Psychological Association.

- Kyllonen, P. C., & Zu, J. (2016). Use of response time for measuring cognitive ability. *Journal of Intelligence*, 4(4), 1-29. doi: <https://doi.org/10.3390/jintelligence4040014>
- Melican, G. J., Breithaupt, K., & Zhang, Y. (2010). Designing and implementing a Multistage Adaptive Test: The Uniform CPA Exam. En W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 167-189). New York, NY: Springer.
- Molina, J. G., Pareja, I., & Sanmartín, J. (2008). Modeling item banking: analysis and design of a computerized system. *Revista Electrónica de Metodología Aplicada*, 13(2), 1-14.
- Muckle, T. J. (2016). Web-based item development and banking. En S. Lane, M. R. Raymond & T. M. Haladyna, *Handbook of test development* (2ª ed, pp. 241-258). New York, NY: Routledge.
- Parshall, C. G. & Harmes, J. C., (2009). Improving the quality of innovative item types: four tasks for design and development. *Journal of Applied Testing Technology*, 10(1), 1-20. Recuperado de <http://www.jattjournal.com/index.php/atp/article/view/48349/39219>
- Parshall, C. G., Harmes, J. C., Davey, T., & Pashley, P. (2010). Innovative items for computerized testing. En W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 215-230). Norwell, MA: Kluwer Academic Publishers.
- Piton-Gonçalves, J., & Aluísio, S. M. (2015). Teste adaptativo computadorizado multidimensional com propósitos educacionais: princípios e métodos. *Ensaio: Avaliação e Políticas Públicas em Educação*, 23(87), 389-414. doi: <https://doi.org/10.1590/S0104-40362015000100016>
- Rudner, L. M. (2010). Implementing the Graduate Management Admission Test Computerized Adaptive Test. En W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 150-165). New York, NY: Springer.
- Sireci, S. G., & Zenisky, A. L. (2006). Innovative item formats in computer-based testing: in pursuit of improved construct representation. En S. M. Downing, & T. M. Haladyna, *Handbook of test development* (pp. 329-348). NJ: Lawrence Erlbaum Associates.
- Sireci, S. G., & Zenisky, A. (2016). Computerized innovative ítem formats: achievement and credentialing. En S. Lane, M. R. Raymond & T. M. Haladyna, *Handbook of test development* (2ª ed, pp. 313-334). New York, NY: Routledge.
- Strain-Seymour, E., Way, W., & Dolan, R. P. (2009). *Strategies and processes for developing innovative items in large-scale assessments*. Recuperado de <http://images.pearsonassessments.com/images/tmrs/StrategiesandProcessesforDevelopingInnovativeItems.pdf>
- Vale, D. C. (2006). Computerized item banking. En S. M. Downing, & T. M. Haladyna, *Handbook of test development* (pp. 261-286). New Jersey, NJ: Lawrence Erlbaum Associates.
- Van der Linden, W. J. (1998). Bayesian item-selection criteria for adaptive testing. *Psychometrika*, 62, 201-216. doi: <https://doi.org/10.1007/BF02294775>
- Van der Linden, W. J., & Pashley, P. J. (2010). Item selection and ability estimation in adaptive testing. En W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 3-30). New York, NY: Springer.
- Verschoor, A. J., & Straetmans, G. J. J. M. (2010). MATHCAT: A flexible testing system in mathematics education for adults. En W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 137-149). New York, NY: Springer.
- Walter, O. B. (2010). Adaptive tests for measuring anxiety and depression. En W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 123-136). New York, NY: Springer.
- Weiss, J. (2013). Item banking, test development, and test delivery. En K. F. Geisinger (Ed.), *APA Handbook of testing and assessment in psychology. Volume 1: Test theory and testing and assessment in industrial and organizational psychology* (pp. 185-200). Washington, D. C.: American Psychological Association.