

Tests Informatizados. Nuevos desafíos prácticos y éticos para la Evaluación Psicológica ¹

Computerized Tests. New practical and ethical challenges for Psychological Assessment

Gabriela Susana Lozzia², Facundo Juan Pablo Abal, Germán Diego Blum, María Ester Aguerri, María Silvia Galibert y Horacio Félix Attorresi

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

(Recepción: Junio 2009 – Aceptación: Julio 2009)

Resumen

El objetivo de este trabajo es acercar al lector de nuestro medio a los problemas y soluciones que van surgiendo en el nuevo campo de trabajo que resulta de aplicar la informática a la Evaluación Psicológica. Se resaltan las cuestiones referidas a una adecuada práctica en evaluaciones asistidas por computadora y evaluaciones a través de Internet. Se caracterizan las distintas formas en que las nuevas tecnologías se aplican a la Evaluación Psicológica: aplicación de los tradicionales tests de lápiz y papel mediante computadora, elaboración automatizada de informes, tests adaptativos informatizados, construcción automatizada de tests y generación automática de ítems. Se examinan las normas y los reglamentos desarrollados específicamente para cada una de estas áreas. Se concluye indicando cuáles son las cuestiones candentes en el uso correcto de los Tests Informatizados. Finalmente, se incita a que los profesionales psicólogos no abandonen el debate y la reflexión sobre estos temas.

Palabras Clave: Evaluación psicológica, tests informatizados, evaluación por internet, tests adaptativos informatizados, normas éticas.

Abstract

The purpose of this study is to bring the readers in our field of knowledge closer to the new problems and solutions resulting from the application of computer systems to Psychological Assessment. Therefore, this work puts forward a suitable implementation of Computer-based and Internet-delivered Testing, includes a description of the new technologies that can be applied to Psychological Assessment: administration of traditional paper-and-pencil tests through computers, elaboration of automated reports, computerized adaptive tests, automated test construction and automatic generation of items, as well as the specific guidelines and regulations governing the development of each of these areas. This study provides an outline of the current issues connected with the appropriate use of Computerized Tests by way of conclusion and finally encourages psychologists to keep debating and reflecting on these topics.

Key Words: Psychological assessment, computerized tests, internet-delivered testing, computerized adaptive tests, ethical standards.

1 La investigación que se presenta en este artículo fue realizada con subsidios de la Universidad de Buenos Aires (UBACyT P043) y de la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT PICT 2004 N° 20909).

2 Correspondencia a: Gabriela Susana Lozzia. Dirección: Av. Lope de Vega 1507 Dpto. 2, (1407) Ciudad Autónoma de Buenos Aires, Argentina. Te: (011)4568-6172 Correo electrónico: glozzia@psi.uba.ar

Sin duda, la Evaluación Psicológica (EP) es una de las disciplinas de la psicología científica presente en cualquier actividad del psicólogo. Una de sus características salientes es la relevancia que los profesionales de la EP le han dado a la rigurosidad con que debe realizarse todo el proceso. Es numerosa la literatura en psicometría que destaca la importancia de la evaluación de la calidad métrica de los tests. Pero hay que tener en cuenta que además de tratar de conseguir instrumentos de buena calidad, es primordial que, posteriormente, se utilicen de forma correcta. En efecto, los errores en la práctica de la EP pueden surgir no sólo de la falta de adecuación de las propiedades técnicas y psicométricas (American Educational Research Association [AERA], American Psychological Association [APA] & National Council on Measurement in Education [NCME], 1999; Hambleton, 2001) sino también de su mala utilización (Muñiz & Fernández-Hermida, 2000).

Con el fin de proporcionar una serie de directrices que ayuden a mejorar el uso de los tests por parte de los profesionales se han creado varias comisiones internacionales entre las que merecen ser destacadas la *Task Force* sobre tests de la *Federación Europea de Asociaciones profesionales de Psicólogos* (EFPPA) y la *International Test Commission* (ITC). Las directrices dictadas por estos organismos internacionales no pretenden eliminar las diferencias que puedan existir en la práctica profesional de cada país, sino que tratan de brindar principios generales orientados a mejorar el uso de los tests. Asimismo, las asociaciones y colegios de psicólogos de distintas partes del mundo han redactado (o adaptado localmente) sus propios códigos de ética.

En los últimos años han acontecido grandes y vertiginosos desarrollos en el área de los Tests Informatizados (TIs). Estas nuevas formas de EP requirieron la actualización de las normativas vigentes. Si bien a ellos son aplicables los lineamientos correspondientes a la EP, presentan particularidades que deben ser tenidas en cuenta. Los nuevos desafíos planteados por esta modalidad de evaluación aún no se terminan de resolver. Es actual la discusión al respecto en el marco de la implementación tanto de los Tests Adaptativos Informatizados como de las evaluaciones realizadas a través de Internet.

En este trabajo se adoptará la expresión de TIs en su sentido más amplio. Los grado de informatización en los procesos de construcción y aplicación de estos tests pueden ser distintos; sin embargo, como señala Olea, Ponsoda y Prieto (1999), todos deben cumplir dos condiciones: tener adecuadas propiedades psicométricas (como todo test psicológico) y que sus ítems se presenten y respondan por medio de una computadora que también se emplea para estimar el nivel de rasgo o rendimiento del evaluado. Asimismo, puede tratarse de TIs realizados tanto de manera online como offline (i.e. a través de una conexión a Internet o no). Esta última versión puede estar contenida en un CD-ROM o ser un programa ejecutable.

El objetivo de este trabajo es acercar al lector los problemas y soluciones que van apareciendo en el nuevo campo de trabajo que resulta de aplicar la informática a la EP. Se resaltan las cuestiones referidas a una adecuada práctica en evaluaciones asistidas por computadora y evaluaciones a través de Internet. Tras caracterizar las distintas formas en que las nuevas tecnologías se aplican a la EP, se examinan las normas y los reglamentos desarrollados específicamente para esta área. Se concluye indicando cuáles son las cuestiones candentes en el uso correcto de las nuevas tecnologías de EP. Finalmente, se incita a que los profesionales psicólogos no abandonen el debate y la reflexión sobre estos temas.

La informatización de la Evaluación Psicológica

La revolución tecnológica ha introducido grandes cambios en la psicometría, no obstante, existen diferentes niveles de informatización de la EP. Sin la intención de describir la historia de este proceso ni los adelantos que implicó en el tratamiento estadístico de los datos, se puede destacar que la informatización de los tests se dio de cuatro maneras distintas (Bunderson, Inouye & Olsen, 1989): a) tests convencionales informatizados, b) elaboración automatizada de informes, c) tests adaptativos informatizados, y d) construcción automatizada de los tests.

Lo primero fue la aplicación de los tradicionales test de lápiz y papel mediante computadora. Aunque lo único que se había informatizado era la forma de administración, este hecho introdujo

cambios beneficiosos: a) control y precisión en la presentación de los ítems a través del monitor, b) administración de ítems complejos o interactivos, c) registro del tiempo o del procesos de respuesta, d) corrección y almacenamiento de las respuestas, e) devolución rápida y eficaz de los resultados.

En los últimos treinta años ha tenido lugar un enorme desarrollo de la tecnología psicométrica con la aparición de los modelos englobados bajo la denominación de *Teoría de Respuesta al Ítem* (TRI). Este nuevo enfoque permite evaluar a las personas en un determinado rasgo sin necesidad de utilizar los mismos ítems para todas ellas, gracias a que los modelos de la TRI expresan en la misma métrica las puntuaciones obtenidas por medio de ítems distintos, lo cual era imposible en el marco de la Teoría Clásica de Tests (Muñiz, 1997). En pocas palabras, los modelos de TRI generan mediciones invariantes respecto de los tests utilizados para obtenerlas. Esta propiedad, en conjunción con las computadoras personales, va a dar lugar a la revolución tecnológica más importante de los últimos años en el campo de la EP: los *Tests Adaptativos Informatizados* (TAIs). A partir de un banco de ítems se selecciona el conjunto de ítems más apropiados para la medición de cada persona en función de su nivel de habilidad (Wainer et al., 2000). Conlleva grandes ventajas en la precisión de las mediciones y en el tiempo invertido, incluso en la satisfacción de los evaluados, ya que al enfrentarse a pruebas acordes con su nivel se minimizan los aspectos frustrantes que lleva aparejada toda evaluación. El uso de los TAIs plantea todo un conjunto de problemas técnicos y éticos nuevos sobre los que trabajan intensamente los especialistas en la actualidad. Entre los técnicos, cabría citar, por ejemplo, la nueva forma de estimar la fiabilidad mediante la *Función de Información*, que permite calcular la precisión del test para los distintos niveles de la variable medida, en vez de ofrecer un valor global como lo hacía el coeficiente de fiabilidad clásico; o los nuevos métodos para evaluar el sesgo de los instrumentos de medida, es decir, el *Funcionamiento Diferencial del Ítem*. Para una ampliación de estos puntos se puede consultar el libro de Olea y Ponsoda (2003).

Muchos son los estudios en el área de los TAIs en general, sin embargo pocos son los dedicados a la validez de los mismos en particular (Muñiz & Hambleton, 1999). Parecería que la validez se da por supuesta, como consecuencia de la gran precisión alcanzada con esos pocos ítems y garantizada por el buen ajuste a los datos. Sin embargo, la validez debe ser demostrada explícitamente. Para ello es necesario apelar a nuevas estrategias ya que las clásicas no son aplicables (Muñiz, 2005). En este caso no hay un test fijo que brinda un puntaje total para ser correlacionado con un criterio. Además, se complica el análisis estadístico ya que hay muy pocos evaluados que responden a los mismos ítems.

Un problema actual está relacionado con los modelos psicométricos utilizados habitualmente con los TAIs. Estos son unidimensionales, mientras que son muchos los constructos de interés psicológico que requieren un tratamiento multidimensional. Otro aspecto a tener en cuenta es que los TAIs pueden incrementar la ansiedad del evaluado ya que, a diferencia de los tests convencionales, no se pueden ni repasar ni corregir las respuestas. Los nuevos abordajes al respecto pueden consultarse en Huff y Sireci (2001).

Por otro lado, la ayuda que significa para los psicólogos la *elaboración automatizada de informes* puede ser también una fuente de abusos sino se utiliza adecuadamente. No se trata de sustituir la tarea del psicólogo sino de tener un punto de partida para responder a los objetivos de la evaluación. Las declaraciones del informe automatizado pueden ser generales y no dirigidas a responder a la demanda que motivó la EP. Por ello, sería conveniente utilizar el informe automatizado junto con la información obtenida a partir de otras fuentes y así garantizar un amplio tratamiento del entorno, comportamiento, capacidades, aptitudes y personalidad del evaluado, ajustando el lenguaje usado a las características del receptor. Sólo el profesional es capaz de reunir información de una variedad de fuentes, contextualizarlas e interactuar con la persona a evaluar siendo los resultados en los tests sólo una parte del proceso global de evaluación (Fernández-Ballesteros & Calero, 2007). Obviamente, si sólo se requiere de un tratamiento estadístico de los datos, los programas diseñados específicamente para esta función ofrecen resultados sin los errores que podrían cometerse con el cálculo manual. El profesional enriquece estos informes estadístico cuando aporta nueva información relevante no

contemplada por el programa (Godoy, 1996). En este sentido la *American Psychological Association* (APA, 1986, p. 12) indica:

Los informes generados por ordenador deben utilizarse únicamente en conjunción con el juicio profesional. El usuario debe juzgar la validez del informe automatizado para cada persona evaluada, basándose en su conocimiento profesional de todo el contexto de evaluación y en el rendimiento y características de la persona evaluada.

Otras aplicaciones de la informática a la psicometría son la *construcción automatizada de tests* y la *generación automática de ítems* (Hambleton, Slater, Narayanan & Setiadi, 1996). Asimismo ya se han dado los primeros pasos en la corrección informatizada de pruebas de carácter abierto, como ensayos y redacciones (Bejar & Bennet, 1999; Clauser, 2000). A pesar de los prometedores resultados, es necesario estar alerta a errores introducidos por la extrapolación del algoritmo de corrección de un campo de contenido a otro, o por su incapacidad para recoger contenidos atípicos o innovadores ya que sólo se detecta lo que se ha programado (Muñiz, 2005).

Directrices para la Evaluación Informatizada

Frente a este gran desarrollo tecnológico, es necesario prevenir contra el mal uso de las nuevas tecnologías. Si bien las recomendaciones generales mantienen su vigencia, en la actualidad se han escrito nuevas directrices con respecto a la evaluación informatizada. Los beneficios que aporta la informatización a la construcción, aplicación y corrección de los tests, no sólo no exime de comprobar que se cumplan los requisitos que definen a los instrumentos de medida científicos, sino que además ha agregado nuevos problemas. Estos están relacionados con las normas de administración, la seguridad de las pruebas y de los resultados y el control del proceso de evaluación.

Los primeros intentos por determinar las directrices técnicas para los TAI fueron resumidas en un artículo de Green, Bock, Humphreys, Linn y Reckase (1984). Los autores trataron los temas: dimensionalidad, error de medición, validez, estimación de parámetros de los ítems, características del pool de ítems, y factores humanos. Otro artículo pionero fue escrito por Hofer en 1985. Él dio importancia a la redacción de normas en las siguientes categorías: equivalencia de puntuaciones, administración, formato y contenido, equipo y procedimientos.

La última versión de *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 1999) hace sólo tres referencias específicas a los requerimientos exigibles a los TIs: a) documentación que acredite su efectividad y explique los procedimientos a seguir para su puesta en funcionamiento (Standard 3.12), b) instrucciones claras de cómo responder y oportunidad de tener una práctica previa que garantice el adecuado manejo del equipo involucrado (Standard 5.5), y c) si el test admite diferentes formas de aplicación (e.g. individual, grupal, informatizada, papel y lápiz) se debe informar la medida en que los resultados derivados de estos métodos son comparables (Standard 6.11).

Como el uso de estos productos ha ido en aumento fue imperioso asegurar que el desarrollo, la distribución y la administración de estos instrumentos de evaluación siguieran las normas de lo que se considera una buena práctica. En respuesta a esto, la ITC (2005) desarrolló directrices internacionalmente consensuadas con respecto a la evaluación informatizada a través de computadoras personales o de Internet. Esta publicación recoge las contribuciones de especialistas en evaluación informatizada tanto psicológica como educativa, incluyendo a constructores, editores y usuarios de tests de distintos países.

En términos generales las directrices de la ITC buscan llamar la atención sobre cuatro aspectos relacionados con la evaluación informatizada: 1) la Tecnología - asegurar que los aspectos técnicos de los TIs sean tenidos en cuenta, sobre todo en relación con el hardware y el software requerido para llevar adelante las pruebas, 2) la Calidad - garantizar la calidad del TI y de los materiales usados y asegurar un manejo adecuado del mismo a través de todo el proceso de evaluación, 3) el Control - control de la autenticación del evaluado y de su práctica previa, y 4) la Seguridad - resguardar

tanto los materiales del test como los datos de los evaluados (privacidad, protección de datos y confidencialidad).

Estas directrices tienen recomendaciones particulares para los constructores de TIs. Les sugiere utilizar el software o el hardware que sea esencial para medir el constructo y que sea fácilmente incorporado a los sistemas usados por los usuarios de los TIs. Ellos deben demostrar que el sistema informático que está detrás del test funciona correctamente y deben proporcionar una descripción clara de los requerimientos mínimos de hardware y software necesarios para su aplicación. El sistema informático que sostiene el test debe ser lo suficientemente robusto y capaz de superar los posibles errores del sistema y del usuario. De esta manera se reduciría al mínimo el surgimiento de problemas mientras el evaluado responde. Por ejemplo, se pueden tomar como equivalentes el ingreso de fuentes en mayúsculas y minúsculas, deshabilitar teclas o comandos que no tienen ninguna función en la prueba o impedir que el evaluado salga de la prueba por accidente. Es conveniente proporcionar a los usuarios orientación sobre qué hacer en caso de que virus, errores o cualquier otro problema interfieran el proceso de prueba. También se recomienda ofrecer servicios de soporte técnico y prácticas para el manejo del programa.

Los constructores deben elegir el formato del test informatizado que sea adecuado para la evaluación tanto del constructo como de la población a la que está destinado. Al crear los ítems deben asegurarse de que pueden ser mostrados correctamente de manera informatizada. El diseño de la imagen en pantalla debe ser coherente mostrando sólo la información pertinente y facilitando la diferenciación de ensayos e instrucciones. El diseño puede contemplar la posibilidad de ajustes razonables para adaptar las características de la prueba de manera que faciliten la participación de personas con necesidades especiales (e.g. el aumento de tamaño de la fuente) mientras no afecten la estandarización del instrumento ni el resultado de la evaluación.

Con respecto a los informes automatizados, se pueden diseñar distintas plantillas de informes para todos los interesados en el proceso de evaluación. El constructor tendrá que describir la manera de obtenerlos y lo que cada uno informa. Es conveniente que se brinde una guía sobre cómo interpretar los informes y de qué manera comunicar los resultados.

Los constructores tienen la obligación de: describir las bases teóricas y prácticas de los algoritmos usados en la determinación de la puntuación, la selección de ítems y/o el control del orden en los TAIs y el modelo de medida; verificar que hubo un adecuado ajuste al modelo psicométrico propuesto; documentar las propiedades psicométricas (validez y fiabilidad); y evaluar el Funcionamiento Diferencial de los Ítems. Si éste existe para uno o más grupos, se debe identificar dónde se produce y modificar la prueba para superarlo. Los constructores y editores de tests sólo deben publicar y ofrecer de manera online aquellos tests con evidencias psicométricas apropiadas que justifiquen su uso.

Si el test informatizado fue desarrollado a partir de una versión de papel y lápiz, es necesario dar a conocer que hay pruebas de equivalencia documentadas para demostrar que las dos versiones: tienen fiabilidades comparables, correlacionan entre sí en el nivel esperado, correlacionan de manera similar con otras pruebas y con criterios externos, y obtienen medias y desviaciones estándares comparables o que han sido debidamente calibrados para hacer equivalentes a las puntuaciones. También se debe asegurar que el evaluado tenga el mismo grado de control (capacidad de adelantar o retroceder en el protocolo) y que las formas de responder sean equivalentes. Todos estos requisitos deberán ser cumplidos para que las normas de la versión tradicional puedan ser usadas para interpretar los puntajes obtenidos en la versión informatizada.

Cuando el test informatizado se utilice a nivel internacional, habrá que tener en cuenta: a) que los países difieren en su acceso a la tecnología, b) evitar el uso de contenidos específicos de un país o cultura, y c) si es necesaria una adaptación se deben seguir las normas desarrolladas al respecto. Si hay evidencias de una posible desigualdad en el acceso, se recomienda hacer formas alternativas del test.

Siempre es importante llevar adelante controles sobre la práctica previa y la exposición de los ítems. Por otro lado, el evaluador debe informarse sobre como reducir el riesgo de trampas por parte del evaluado y minimizar el peligro de que ítems, claves de respuesta, y algoritmos de interpretación puedan ser ilegítimamente impresos, descargados, copiados o enviados electrónicamente a otra computadora. El software debe impedir el acceso a las opciones de menú (como copiar, pegar, etc.) y estar protegido contra hackers y virus.

Los usuarios de TIs deben asegurarse de tener un buen manejo de las exigencias técnicas y operativas del test, así como del hardware, del software y de los recursos humanos necesarios para usar y mantener el TI. Es importante verificar si el evaluado sabe cómo interactuar con el sistema informático. También hay que constatar que la prueba no requiera conocimientos, capacidades o habilidades (e.g. conocimientos informáticos) que sean irrelevantes para los fines de la evaluación. Recoger datos sobre las reacciones hacia la evaluación informatizada es útil para proporcionar retroalimentación a los constructores de la prueba a fin de introducir mejoras en el proceso de evaluación.

Cuando la evaluación se realiza en un centro diseñado para tal fin, el evaluado debe estar cómodo con el entorno de trabajo. Si la evaluación es a través de Internet, se puede orientar al evaluado sobre los mejores métodos para realizar la prueba. Los evaluados deben recibir información sobre las reglas de puntuación utilizadas (e.g. qué sucede si dejan ítems sin responder, si hay sanciones por adivinar) y contar con un medio de contacto ya sea telefónicamente o por correo electrónico.

Directrices para la Evaluación a través de Internet

En los últimos años, se ha producido un incremento sustancial en la oferta de servicios psicológicos a distancia. El rápido desarrollo de Internet ha acentuado la necesidad de revisar la práctica profesional con respecto a los principios éticos de la evaluación a través de este medio. En varios países de Europa se están llevando a cabo trabajos sobre evaluación a distancia vía Internet en las áreas de trabajo y educación (Colegio Oficial de Psicólogos & ITC, 2000). Surgen nuevos problemas: el de la validación de las pruebas en esas circunstancias, la autenticación (asegurarse que la persona evaluada es la que responde), el control del proceso de evaluación y la seguridad de los resultados (Muñiz, 2005). Por ello, también se han desarrollado recomendaciones específicas para la evaluación por Internet. En el año 2004, la Comisión Deontológica Estatal del *Colegio Oficial de Psicólogos* de España ha editado en castellano el documento elaborado en relación a este tema por la EFPPA.

Una evaluación a través de Internet puede ser llevada adelante con distintos niveles de control. De este aspecto dependerá la calidad de los resultados obtenidos. En el modo Abierto no hay ningún tipo de control. En el modo Controlado, aunque no hay control personal directo, la prueba sólo está disponible para evaluados que disponen de una clave de acceso. En el modo Supervisado, el evaluador tiene la obligación de registrar al candidato y confirmar que la evaluación fue administrada y completada correctamente. Finalmente, para el modo Gestionado, se requiere un alto nivel de supervisión y control sobre las condiciones de evaluación (como en un centro dedicado a la EP). Si la evaluación se realiza en los modos Abierto o Controlado, se pueden evitar las trampas indicando al evaluado que podría requerírsele una evaluación posterior de validación bajo la condición Gestionada para una confirmación de los resultados.

Para los tests por Internet los estudios de validación deben llevarse a cabo a través de Internet con participantes que completen la prueba en las mismas condiciones que lo haría la población objetivo (e.g. sin supervisión o sin estandarización de las condiciones de evaluación). El test debe ser usado sólo en los modos de administración para los cuales ha sido diseñado (e.g. no administrar el test de forma no supervisada cuando fue diseñado para ser usado bajo supervisión).

Un tema primordial es el de la identificación del evaluado (autenticación). Deben establecerse mecanismos de control por asegurar que la comunicación se hace con el que ha de ser examinado y no con otra persona (e.g. a través de una identificación numérica). Se tienen que tomar medidas que dificulten la acción de los impostores. El anonimato del usuario puede ser apropiado bajo

algunas circunstancias, pero en tales casos el psicólogo es el encargado de tomar los recaudos extra y reconocer la dificultad aumentada en la determinación de la vulnerabilidad del usuario y su personería jurídica (e.g. si es un niño).

Es necesario explicar y alertar a los examinados de las limitaciones de la tecnología informática, como son las limitaciones atinentes a la relación entre evaluador - evaluado (e.g. Internet es un medio impersonal y el evaluador sólo puede proporcionar un asesoramiento limitado) o que la interpretación se basa sólo en los resultados del test cuyos datos fueron utilizados como insumos, otros datos auxiliares que puedan ser importantes no son tenidos en cuenta. También se presentan dificultades en asegurar de manera absoluta la privacidad y la confidencialidad a través de Internet. Si el evaluado no acepta estas limitaciones expuestas se lo tendrá que remitir a otro tipo de intervención. También se le tiene que informar a los evaluados: qué registros serán guardados, los derechos del usuario en cuanto a estos, el tiempo que se mantendrá la información y cuáles son los medios para acceder a la misma. La información almacenada deberá mantenerse de manera segura (incluyendo copias de seguridad). Se deben tomar los recaudos pertinentes en relación a la protección de la base de datos que contengan los historiales y los datos personales del evaluado. En este punto se siguen las mismas directrices que regulan el manejo de la información confidencial para todo trabajo en psicología.

La interacción psicológica a través de Internet no escapa a las normas propias del ciberespacio en cuanto a la protección de las transacciones (e.g. las transacciones financieras). Por ello, se debería garantizar seguridad máxima sobre cualquier sitio de Internet, línea telefónica u otro medio usado para la interacción, incluyendo el empleo de servicios encriptados. Los sistemas deben tener revisiones regulares que garanticen su adecuado funcionamiento y mejoren los niveles de seguridad.

Si la aplicación del test se realiza vía Internet, se deberán tener en cuenta: a) las fluctuaciones de la demanda en diferentes momentos, b) las diferencias en las velocidades de conexión que se usan en el mundo, c) las fallas por pérdidas de la conexión a Internet o por lentitud en las descargas (e.g. el sistema debería asegurar que no se pierde información en estos casos).

También es importante informar si los servicios se ofrecen en *tiempo real*, a través de programas interactivos o por correo electrónico. En este último caso se deben aclarar los tiempos de espera. Todos los servicios, tanto los de comunicación directa como mediante respuestas automatizadas, tienen que estar bajo la supervisión de un psicólogo.

Cuando los informes automatizados se obtienen directamente de Internet, se plantean nuevos problemas como la dificultad de saber el efecto que causará una retroalimentación negativa en el evaluado, la falta de conocimiento de su estado emocional, o la dificultad de prestar apoyo inmediato al examinado cuando la información tiene un impacto negativo. Por ello, la retroalimentación debería incluir instrucciones sobre la forma de acceder a una consulta u otra información pertinente.

La rapidez, los bajos costos y la gran accesibilidad de Internet son un beneficio para la EP. Un nuevo test puede estar disponible en todo el mundo en cuestión de instantes sólo por contar con una conexión a Internet. También se agiliza la devolución de los resultados. El tiempo invertido tanto en cargar o escanear las respuestas de los evaluados para su análisis como en realizar e imprimir los informes se acorta ya que el evaluado ingresa directamente sus respuestas y puede recibir un informe automatizado segundos después de haber contestado al último ítem del test. Además, la evaluación a través de Internet implica menores costos ya que se evita la impresión de todo el material y los gastos de distribución y corrección propios de los tests de lápiz y papel. En un futuro se podría pensar en sistemas de evaluación a través de Internet sin costos para los evaluados. Por ejemplo, los investigadores pueden poner los tests en sus páginas Web esperando que la gente complete la evaluación con el fin de recibir un informe. A cambio, los constructores obtienen los datos aportados por los encuestados. Actualmente algunos editores de tests presentan pruebas de forma gratuita en sus sitios Web como medio de atraer a potenciales clientes.

Los TIs proporcionan algunas ventajas en comparación con los de papel y lápiz. De hecho, muchas investigaciones así lo demuestran (ver Drasgow & Olson-Buchanan, 1999; Sands, Waters, & McBride, 1997). Los TIs ofrecen puntuaciones más precisas ya que hay menos errores en la carga

de los datos dado que los evaluados responden directamente en la computadora y pueden verificar si han ingresado la respuesta en forma correcta. También permiten administrar TAI's o evaluar habilidades y destrezas imposibles de plasmar en un test convencional. Los recursos multimedia hacen más vistosos e interesantes los tests para el evaluado, lo que puede redundar en una mayor motivación para realizar la prueba. Gracias a los TIs, los psicólogos aprovechan más eficientemente su tiempo. Por ejemplo, las actividades de evaluación rutinarias se realizarían a través de Internet, dejando a los psicólogos libres para dedicarse a la interpretación y devolución de los resultados.

Como contrapartida, la informatización de la EP puede generar datos distorsionados a causa de la distinta familiaridad de los evaluados con las computadoras (Taylor, Kirsch, Eignor & Jamieson, 1999) o con el sistema operativo utilizado (Windows o Macintosh).

Obviamente para todas estas aplicaciones de la informática a la EP siguen vigentes las normas tradicionales que son recogidas por la totalidad de los códigos de ética (Colegio Oficial de Psicólogos & ITC, 2000). A saber, que el psicólogo hará uso de los tests de forma adecuada, profesional y ética, prestando la debida atención a las necesidades y derechos de las personas implicadas en el proceso de evaluación, y teniendo muy en cuenta las razones para utilizar los tests, así como el contexto en el cual se lleva a cabo su aplicación. Los profesionales son los que eligen los servicios de calificación e interpretación (incluidos los servicios automatizados) basándose en los datos sobre la validez del programa y procedimientos, así como en otras consideraciones pertinentes. Además los psicólogos tienen el compromiso de proporcionar una explicación de los resultados, utilizando un lenguaje que pueda entender la persona evaluada, independientemente de que la puntuación e interpretación sean hechas por el propio psicólogo o por medio de servicios automatizados u otros servicios externos (APA, 1992).

Comentarios finales

Son muchos los temas que en la actualidad ocupan las energías de los investigadores y profesionales del área. Pero existe una discusión sobre la ética y deontología que pugnan por abrirse paso e imponerse. Es que de poco vale una tecnología de evaluación poderosa sin las riendas ético-deontológicas (Rodríguez-Neira, 2001). No hay que olvidar que se evalúa a personas y sobre ellas se toman decisiones, de modo que toda precaución en este sentido nunca es suficiente. No sólo los psicólogos sino también todos los organismos que requieren de los resultados de las EP deberían ser conscientes de la necesidad de una buena preparación en la tecnología de la evaluación, lo que redundaría en beneficios tanto para el proceso de evaluación como para los evaluados. En este sentido, a modo de ejemplo de su repercusión en otras áreas, son muy interesantes las directrices elaboradas por la Federación de Profesores Americanos sobre las competencias requeridas por los profesores para evaluar a los alumnos, así como las directrices para el uso adecuado de los tests en ámbitos educacionales elaboradas por varias organizaciones educativas (Rodríguez-Neira, 2001).

Como señaló Fernández-Ballesteros (1998), la EP involucra a distintos colectivos (e.g. constructores, editores). Por ello, se han redactado *Estándares para la construcción de tests* (APA, 1985) que tratan de normalizar qué requisitos deben reunir los instrumentos para ser considerados herramientas científicas. A su vez, los tests son también productos comerciales que han de ser publicados y distribuidos con ciertas garantías. Por esta razón, en algunos países se han promulgado normas para la venta y distribución de tests.

Otra práctica habitual es la traducción y adaptación de los tests de una cultura a otra, proceso que también ha sido regulado por la ITC (Hambleton, 2001). En nuestro medio, tomando las palabras de Buela-Casal, Sierra, Carretero-Dios y De los Santos-Roig (2002) con respecto a la realidad española, se adolece de un "nacionalismo psicológico". Se utilizan instrumentos extranjeros, generalmente con inadecuadas adaptaciones y carentes de baremos locales.

De gran importancia es el problema de la regulación de la utilización de los tests donde se involucran el intrusismo profesional (ejercicio de actividades profesionales por una persona no

autorizada legalmente para ello) y la necesidad de cualificación para aplicar los distintos tipos de instrumentos de medida. Estas herramientas tienen diversos grados de dificultad y, por tanto, requieren distintos niveles de formación. Actualmente la ITC está intentando regular este ámbito ya que la causa más importante del uso inapropiado de los tests son las deficiencias en los conocimientos técnicos (Muñiz & Fernández-Hermida, 2000). Sin embargo, no es fácil describir de forma general los conocimientos necesarios para la utilización adecuada de los tests en todas las situaciones. Los manuales de cada test deberían indicar con precisión los conocimientos requeridos para su aplicación. Según el Colegio Oficial de Psicólogos de España y la ITC (2000), esta descripción ha de cubrir las distintas áreas de conocimiento implicadas en la situación de que se trate, como son los conocimientos sobre: a) Teoría de los Tests y propiedades técnicas de los tests (fiabilidad, validez, estandarización, sesgo, análisis de ítems, etc.), b) principios de la medición para entender adecuadamente los resultados, c) Teoría de cada uno de los constructos medidos que garantice la pertinente elección de las pruebas y la interpretación de los resultados, d) tests disponibles, así como los editores correspondientes, e) avances técnicos recientes, f) consecuencia del uso de los tests, entre otros posibles.

Asimismo, todos los códigos subrayan la importancia de la actualización permanente en el conocimiento y uso de instrumentos y/o procedimientos de evaluación, para asegurar la adecuada competencia (Comité Coordinador de Psicólogos del Mercosur & Países Asociados, 1997). Como sólo se pueden proporcionar los servicios para los que se está calificado por formación y experiencia, esta premisa de la actualización cobra mayor relevancia en el área de la evaluación informatizada por las novedades que no cesan de surgir. Es recomendable mantenerse al día de los cambios y avances en relación con el uso y construcción de los tests, incluyendo los cambios de normas y legislación. Sin embargo, estar actualizado no significa necesariamente administrar técnicas basadas en las nuevas tecnología. Significa saber, por un lado, cuándo es pertinente su aplicación por los datos que aportará y, por otro lado, cómo utilizarlos adecuadamente. Siempre se debe tener una justificación para el empleo de software de alta complejidad. No hay que caer en la novedad por la novedad misma.

Para Prieto y Muñiz (2000), mejorar la formación de los usuarios es la más viable de las tres vías que podrían seguirse para mejorar el uso de los tests. En relación a la opción de restringir el acceso de los usuarios, permitiendo su empleo solamente a profesionales debidamente capacitados y acreditados, las asociaciones profesionales no disponen de las armas legales para introducir estas restricciones. La otra vía, fomentar la investigación para desarrollar pruebas eficientes (construcción de nuevas pruebas, adaptación adecuada de las desarrolladas en otros países y revisión periódica de la calidad técnica de las ya disponibles), es un camino que avanza lentamente en la mayoría de los países de América Latina.

Un medio eficaz para mejorar la EP es ofrecer a los usuarios toda la información posible acerca de la calidad y características de los tests disponibles. El Colegio Oficial de Psicólogos de España llevó adelante un trabajo de evaluación de los tests más utilizados en dicho país para proporcionar así una información técnica precisa a los usuarios de las pruebas (Prieto & Muñiz, 2000). El cuestionario que se diseñó para tal fin podría ser de utilidad en otros países. Un modelo similar funciona satisfactoriamente en Inglaterra y Holanda, donde se publican periódicamente monografías en las que se aporta información técnica de los principales tests usados en esos países (Evers, 1996). Otra de las medidas actualmente en debate en la EFPPA es la posibilidad de utilizar un sello de acreditación o control de calidad que garantice la adecuación técnica de las pruebas (Muñiz & Fernández-Hermida, 2000). Mientras tanto, en nuestro medio la información sobre las características de cada tests no es exhaustiva ni fácilmente accesible o, al ser brindada por el editor del test, carece de imparcialidad. Entonces, los profesionales no poseen una orientación científica que les permita seleccionar el instrumento más apropiado para sus objetivos.

Otro aspecto de la EP con escasa regulación en nuestro entorno es el atinente a los derechos de los evaluados. Internacionalmente, hay dos tipos de regulación sobre este tema: 1) los Códigos Deontológicos profesionales que agrupan normas éticas para regular el auto-control del profesional

que en su relación con el examinado está comprometido a una serie de actuaciones, y 2) las asociaciones de ciudadanos que han planteado sus derechos como destinatarios de los tests (Fremer, 1997). La EFPPA solicita a los colegios y asociaciones vinculados a ella que presenten listados de sus miembros a través de Internet, preferiblemente con acceso a las páginas Web de los miembros. También sugiere el desarrollo de guías para el público sobre los servicios psicológicos.

Más allá de todos estos aspectos parciales que requieren regulación (construcción, distribución, adaptación y utilización de tests), la consideración de la EP como un proceso de toma de decisiones merece ser destacado. El psicólogo solicita información, se plantea hipótesis sobre el caso, selecciona los instrumentos de evaluación a utilizar y llega a unas conclusiones de las que se derivan orientaciones y/o intervenciones relativas al caso de evaluación. Este proceso ha sido investigado desde múltiples perspectivas (e.g. resolución de problemas, toma de decisiones, sistemas de experto) y en múltiples contextos aplicados (e.g. clínico, orientación vocacional, selección). Fernández-Ballesteros et al. (2001) aportó una propuesta de normalización de este proceso.

Con este recorrido se observa la gran preocupación ética existente en el área de la EP. Según Fernández-Ballesteros (1998) se comienza el Siglo XXI, con un bagaje normativo aparentemente sólido. No obstante, como señalan el Colegio Oficial de Psicólogos de España y la ITC (2000), es necesario que las directrices sean revisadas periódicamente, actualizándolas de acuerdo con los avances producidos en los tests o en su utilización.

Por otra parte, el cambio más revolucionario de la tecnología evaluativa no lo constituye solamente la informatización de la evaluación. Como Fernández-Ballesteros (1998) destacó, también son importantes los aportes de las técnicas de neuroimagen, que han permitido explorar funcionalmente lo que acontece a nivel neural cuando se administran determinadas tareas, tests u otros estímulos relevantes; los microcircuitos y los sistemas integrados, que han supuesto una revolución en el auto-registro y en la recogida de información en situaciones naturales; los sistemas expertos, que son utilizados en la simulación de procesos mentales; y la realidad virtual, que está permitiendo replicar la situación natural a través de una análoga en el laboratorio. Asimismo, todos estos avances tecnológicos plantean nuevas cuestiones y preocupaciones que habrán de ser reguladas.

El avance de Internet en tareas relacionadas con la práctica de la psicología es inexorable, aunque con una forma y estructura que todavía es imposible de prever. A pesar de que es posible la EP por este medio, se debe tener en cuenta la pérdida de información aportada por la comunicación no verbal y la observación del comportamiento del evaluado. Por ejemplo, podría haber dificultades para evaluar la afectividad, las características del lenguaje, etc. La interacción cara a cara hace posible el feedback instantáneo que permite un cambio rápido de la estrategia de evaluación. Es cierto que el desarrollo de las nuevas tecnologías (e.g. la videoconferencia) puede paliar algunos de estos déficits, pero también es verdad que el contacto personal es una parte importante de la exploración psicológica, y éste está seriamente afectado cuando la evaluación se realiza vía Internet. Es preciso seguir de cerca la evolución futura de estas técnicas para encontrar buenas soluciones a estos problemas.

Otro inconveniente está relacionado con la falta de control sobre los contenidos que se encuentran en la Web. Hay disponible en Internet gran cantidad de material de dudosa calidad, entre ellos muchos tests. Esto desvirtúa una herramienta tan importante para la EP, ya que estos mal llamados “tests” no cuentan con evidencia de fiabilidad y validez ni responden a lo que se considera una buena práctica al respecto. Termina siendo el usuario el que debe evaluar la calidad de la información desde su escaso o nulo conocimiento acerca del tema. La libre accesibilidad permite que cualquiera pueda modificar, cambiar o traducir un test sin la debida autorización o validación. Por otra parte, es posible que los tests de Internet no sean apropiados para determinados grupos de personas, tengan defectos en sus propiedades psicométricas, o estén mal traducidos a otros idiomas. Todo esto refuerza la idea de que la evaluación por Internet tiene que estar debidamente regulada para garantizar calidad y utilidad. Las directrices sugieren supervisar la Web para detectar si hay versiones ilegales o desactualizadas y tomar las medidas (e.g. la aplicación de leyes de derechos de autor) para eliminar estas violaciones (Comisión Deontológica Estatal, 2004).

Frente a lo atractivo de su disponibilidad y su bajo costo, se encuentra la dificultad de comprobar la fiabilidad y la validez de los resultados obtenidos por Internet. Todas las ventajas que presenta la evaluación a través de la Web son irrelevantes si la información que brinda no es fiable y válida en función de los fines para los cuales se la requirió. De todas formas, no alcanza con que los tests gocen de adecuadas propiedades psicométricas; la administración debe también seguir normas que garanticen que el proceso completo de evaluación y no solo el test sea válido. Si la evaluación es de alto impacto, es decir que aporta información para la toma de decisiones con respecto a aspectos importantes de las personas, será conveniente utilizar un modo supervisado de administración del test, ya sea en forma directa (un profesional supervisa la ejecución del evaluado en la computadora) o indirecta (filmación). Asimismo, no hay que olvidar que una buena evaluación requiere la integración de información obtenida a través de diferentes medios. Por ejemplo, se pueden llevar adelante las primeras etapas de una selección de personal mediante una evaluación por Internet y luego realizar evaluación cara a cara con los preseleccionados.

Del mismo modo, los constructores de tests y las editoriales que realizan programas de tests a través de Internet deben velar por que estos sigan las mismas normas psicométricas que las pruebas tradicionales. Esto incluiría, por ejemplo, la documentación que resume las muestras de normalización, la confiabilidad y la validez como así también evidencias adicionales, tales como la equivalencia con la versión en papel, la uniformidad de la calidad del estímulo en diferentes pantallas, y así sucesivamente, para asegurar la calidad del proceso de administración del test.

En los países más desarrollados, la mayoría de las personas están familiarizadas con el uso de Internet, sin embargo, ésta no es una realidad generalizable a toda América Latina. Y si bien se pueden usar los tests por Internet para evaluar a personas que se encuentran lejos de los centros urbanos o con dificultades de movilidad para acercarse a los lugares de evaluación, esto requeriría una infraestructura tecnológica disponible excepcionalmente en nuestro medio. En consecuencia, es importante identificar las situaciones en las cuales la evaluación informatizada puede requerir algún ajuste en su aplicación o ser directamente inaplicable debido a las características de la persona (edad, familiaridad con el soporte, discapacidad, idioma, nivel educativo, etc.).

Entre los nuevos interrogantes, surgen los relacionados con los aspectos legales: ¿la evaluación informatizada aumentará la probabilidad de que el psicólogo tenga problemas legales en comparación con el uso de los tests convencionales? Para países como Estados Unidos, las cuestiones candentes están relacionadas con los derechos de autor, el sesgo (desventajas ante la prueba para determinados grupos sociales) y la práctica del *disclosure* (dar a conocer los tests y las respuestas luego de ser aplicado). Este último punto, que surgió para controlar el proceso evaluativo llevado adelante con los tests convencionales, no es directamente aplicable a los TAIs. La publicación de los ítems administrados tornaría inútil esta tecnología (Muñiz & Hambleton, 1999).

Internet no es más que un medio para realizar el test, lejos está de reemplazar el amplio proceso que el profesional realiza cuando encara una EP. Aunque en el futuro los métodos de prueba a través de Internet puedan mejorarse, los requisitos para la adecuada EP superan la capacidad actual de Internet.

Frente a estas nuevas posibilidades, el psicólogo debe pensar de forma creativa sobre cómo sus actividades de investigación y sus prácticas se pueden mejorar por medio de la evaluación por Internet. Los tiempos han cambiado, Internet ha sacado a la EP del entorno seguro y controlado por el psicólogo. El equilibrio entre la accesibilidad generalizada y buenas prácticas de evaluación presenta un desafío para los psicólogos del nuevo milenio. Hay muchas cuestiones que esperan resolución. Durante los próximos años, gran parte de la investigación y el pensamiento crítico de los especialistas del área será necesaria para abordar estas cuestiones.

A lo largo de esta presentación se han mencionado las ventajas de la informatización de la EP, no obstante, su implementación en las prácticas cotidianas no es aún masiva. La sola informatización del test no garantiza la bondad técnica y psicométrica del test (Muñiz & Hambleton, 1999). El uso de las computadoras conlleva nuevos costos en cuanto a la adquisición y mantenimiento de los equipos.

Por ello, su introducción es mayor en el contexto de evaluaciones individuales que en aplicaciones colectivas. Además, podría haber un perjuicio contra los sectores de menores recursos. Ante estas dificultades para su implementación masiva, Muñiz y Hambleton (1999) suponen una prolongada convivencia de las formas de administración convencional e informatizada:

La sustitución completa de los tests convencionales de papel y lápiz por los ordenadores parece tan lejana como la de los libros y revistas científicas por otros soportes alternativos; y es que, aunque las ventajas son claras en determinadas circunstancias, en otras no ocurre lo mismo. Debemos, pues estar preparados para un largo período de convivencia y complementariedad más que rivalidad; cada tecnología tiene sus pros y sus contras, y lo sensato es utilizar la más apropiada en cada caso (p. 25).

Evidentemente, las nuevas tendencias van a convivir con los modos más clásicos y tradicionales de evaluar, pero los psicólogos de nuestro medio deberían estar atentos a los cambios rápidos que se producen en el ámbito de la EP si no quieren perder competitividad.

El rápido desarrollo de la informatización ha acentuado la necesidad de repasar la práctica profesional en lo que concierne a los principios éticos. La pregunta principal es si estos acontecimientos introducen cuestiones éticas fundamentalmente diferentes o si los principios éticos y normas existentes, como los formulados en los códigos de ética de las distintas asociaciones de psicólogos, son apropiados para orientar el comportamiento de los profesionales. Lo expuesto en este trabajo indica que la dimensión ética de la práctica profesional como psicólogo es la misma si tal práctica es en persona o a distancia, si vía Internet o algún otro medio no directo. En efecto, la evaluación informatizada es alcanzada por los requerimientos de confidencialidad y experticia. Sin embargo, las nuevas tecnologías hacen necesaria la adaptación de estas directrices a las características propias de este tipo de evaluación.

Tanto ahora como en los próximos años será una exigencia la reflexión cuidadosa sobre el comportamiento ético cuando el psicólogo esté involucrado en un nuevo medio de evaluación, sobre todo cuando se carece de la experiencia sobre las preguntas éticas específicas y los dilemas que pueden surgir con el uso de estas herramientas.

Finalmente, este trabajo planteó los nuevos desafíos prácticos y éticos de la informatización de la Evaluación Psicológica en nuestro medio. Se espera que los responsables de un proceso de evaluación con tests informatizados conozcan y tengan presente lo que se considera una adecuada implementación de estos instrumentos.

Referencias

- American Educational Research Association (AERA), American Psychological Association (APA) & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington: APA.
- American Psychological Association, APA. (1985). *Standards for test constructions*. Washington: Author.
- American Psychological Association, APA. (1986). *Guidelines for Computer-Based Tests and Interpretations*, Washington: Author.
- American Psychological Association, APA. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, 47, 1597-1611.
- Bejar, I. & Bennet, R. (1999). La puntuación de las respuestas como un parámetro del diseño de exámenes: implicaciones en la validez. En J. Olea, V. Ponsoda & G. Prieto (Eds.), *Tests informatizados: fundamentos y aplicaciones* (pp. 53-59). Madrid: Pirámide.
- Buela-Casal, G., Sierra, J. C., Carretero-Dios, H. & De los Santos-Roig, M. (2002). Situación actual de la evaluación psicológica en lengua castellana. *Papeles del Psicólogo*, 83, 27-33.
- Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 367-407). London: Collier Macmillan.
- Clauser, B. E. (2000). Recurrent issues and recent advances in scoring performance assessment. *Applied Psychological Measurement*, 24, 310-324.
- Colegio Oficial de Psicólogos & International Test Commission (ITC). (2000). Directrices internacionales para el uso de los tests. *Infocop*, 77, 21-32.
- Comisión Deontológica Estatal (2004). *Ética y Deontología para Psicólogos*. Madrid: Colegio Oficial de Psicólogos.
- Comité Coordinador de Psicólogos del Mercosur & Países Asociados. (1997). Protocolo de acuerdo marco de principios éticos para el ejercicio profesional de los psicólogos en el Mercosur y Países Asociados. En Conselho Federal de Psicologia, *A psicologia no Mercosul* (pp. 11-14). Brasilia: Autor.
- Drasgow, F. & Olson-Buchanan J. B. (Eds.). (1999). *Innovations Internet computerized assessment*. Mahwah, NJ: Erlbaum.
- Evers, A. (1996). Regulations concerning test qualifications and test use in The Netherlands. *European Journal of Psychological Assessment*, 12, 153-159.
- Fernández-Ballesteros, R. (1998). Evaluación psicológica. Algunas claves para el futuro. *Papeles del Psicólogo*, 70, 3-13.
- Fernández-Ballesteros, R., De Bruyn, E., Godoy, A., Ter Laak, J., Vizcarro, C., Westhoff, K., Westmeyer, H., & Zaccagnini, J. (2001). Guidelines for the Assessment Process: A proposal for discussion. *European Journal of Psychological Assessment*, 17, 187-200.
- Fernández-Ballesteros, R. & Calero, M. D. (2007). Garantías científicas y éticas de la evaluación psicológica. En R. Fernández-Ballesteros (Ed.): *Evaluación psicológica. Conceptos, métodos y estudio de casos* (pp. 121-160). Madrid: Pirámide.
- Fremer, J. (1997). *Elaboration of rights of test-takers*. Comunicación presentada en la Michigan School Testing Conference. Ann Arbor, Michigan.
- Godoy, A. (1996). *Toma de decisiones y juicio clínico: Una aproximación psicológica*. Madrid: Pirámide.

- Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. B., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement, 21*, 347-360.
- Hambleton, R. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal Psychological Assessment, 17*, 164-172.
- Hambleton, R., Slater, S., Narayanan, P. & Setiadi, H. (1996). Construcción automatizada de los tests: conceptos básicos, avances técnicos y aplicaciones. En J. Muñiz (Ed.). *Psicometría*. Madrid: Universitas.
- Hofer, P. J. (1985). Developing Standards for Computerized Psychological Testing. *Computers in Human Behavior, 1*, 301-315.
- Huff, K. L. & Sireci, S. G. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practice, 20*, 16-25.
- International Test Commission (ITC). (2005). Guidelines on computer-based and internet-delivered testing. Extraído el 01 de Julio, 2008, de <http://www.intestcom.org/>
- Muñiz, J. (1997). *Introducción a la Teoría de Respuesta a los Ítems*. Madrid: Pirámide.
- Muñiz, J. (2005). La validez desde una óptica psicométrica. *Acta Comportamentalia, 13*, 9-20.
- Muñiz, J. & Fernández-Hermida, J. R. (2000). La utilización de los tests en España. *Papeles del Psicólogo, 76*, 41-42.
- Muñiz, J. & Hambleton, R. K. (1999). Evaluación psicométrica de los tests informatizados. En J. Olea, V. Ponsoda & G. Prieto (Eds.), *Tests informatizados: fundamentos y aplicaciones* (pp. 23-52). Madrid: Pirámide.
- Olea, J. & Ponsoda, V. (2003). *Tests adaptativos informatizados*. Madrid: UNED.
- Olea, J., Ponsoda, V. & Prieto, G. (1999). *Tests informatizados: fundamentos y aplicaciones*. Madrid: Pirámide.
- Prieto, G. & Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo, 77*, 65-75.
- Rodríguez-Neira, T. (Ed.) (2001). *La Evaluación en el Aula*. (2a. Ed.). Oviedo: Nobel.
- Sands, W., Waters, B., & McBride, J. (Eds.). (1997). *Computerized adaptive testing: From inquiry to operation*. Washington: American Psychological Association.
- Taylor, C., Kirsch, I., Eignor, D. y Jamieson, J. (1999). Examining the relationship between computer familiarity and performance on computer-based language tasks. *Language Learning, 49*, 219-274.
- Wainer, H., Dorans, N., Eignor, D., Flaugher, R., Green, B., Mislevy, R., Steinberg, L. & Thissen, D. (2000). *Computerized Adaptive Testing: A Primer* (2a. Ed.). Mahwah, NJ: Erlbaum.