

Principios éticos en la evaluación psicológica

Los principios éticos en la evaluación psicológica propuestos por la *American Psychological Association* (1992) son de carácter general para cualquier administración de un test.

1. Evaluación, diagnóstico e intervenciones en un contexto profesional

- a) Los psicólogos únicamente llevan a cabo evaluaciones, diagnósticos o intervenciones dentro del contexto de una relación profesional definida.
- b) Las evaluaciones de los psicólogos, recomendaciones, informes y diagnósticos o apreciaciones valorativas se basan en información y técnicas (incluyendo entrevistas personales cuando sea conveniente) suficientes para proporcionar una fundamentación adecuada de sus hallazgos.

2. Competencia y uso apropiado de evaluaciones e intervenciones

- a) Los psicólogos que elaboran, aplican, puntúan, interpretan o usan técnicas de evaluación psicológica, entrevistas, tests u otros instrumentos lo hacen de modo que tanto el proceder como los fines perseguidos son apropiados de acuerdo con la investigación y los datos existentes acerca de la utilidad y pertinencia de las técnicas.
- b) Los psicólogos se abstienen del uso inapropiado de técnicas de evaluación, intervenciones, resultados e interpretaciones, y toman medidas razonables para evitar que otros usen inadecuadamente la información proporcionada por estas técnicas. Esto incluye el abstenerse de facilitar las puntuaciones de los tests u otros datos a personas, ajenas a los pacientes o clientes, no calificadas para utilizar dicha información.

3. Construcción de tests

Los psicólogos que construyen y realizan investigación con tests y otras técnicas de evaluación utilizan procedimientos científicos y conocimientos profesionales actualizados para el diseño de los tests, estandarización, validación, reducción o eliminación del sesgo, y recomendaciones de uso.

4. Uso de la evaluación en general y con poblaciones especiales

- a) Los psicólogos que llevan a cabo intervenciones o aplican, puntúan, interpretan o utilizan técnicas de evaluación conocen la fiabilidad, validez, los estudios sobre la estandarización y los resultados, aplicaciones adecuadas y usos de las, técnicas que manejan.
- b) Los psicólogos son conscientes de los límites de la certidumbre con la que pueden hacerse los diagnósticos, juicios y predicciones acerca de las personas.
- c) Los psicólogos intentan identificar las situaciones en las cuales determinadas intervenciones, técnicas de evaluación o normativas pueden no ser aplicables o pueden requerir un ajuste en la aplicación o interpretación, debido a características de la persona tales como el género, edad, raza, etnia, nacionalidad, religión, orientación sexual, discapacidad, idioma o estatus socioeconómico.

5. Interpretación de los resultados de la evaluación

Al interpretar los resultados de la evaluación, incluidas las interpretaciones automatizadas, los psicólogos tienen en cuenta los distintos aspectos del test y características de la persona evaluada que pueden afectar a los juicios del psicólogo o reducir la precisión de sus interpretaciones. Han de indicar cualquier reserva significativa que tengan acerca de la precisión o limitaciones de sus interpretaciones.

6. Personas no cualificadas

Los psicólogos no promueven el uso de técnicas de evaluación psicológica por parte de personas no cualificadas.

7. Tests obsoletos y resultados desfasados de los tests

- a) Los psicólogos no basan sus evaluaciones, decisiones sobre la intervención o recomendaciones en datos o resultados de los tests que están desfasados en relación con los objetivos perseguidos en la actualidad.
- b) Del mismo modo, los psicólogos no basan sus decisiones o recomendaciones en medidas y tests obsoletos, inapropiados para los objetivos actuales.

8. Servicios de calificación e interpretación

- a) Los psicólogos que ofrecen procedimientos de calificación o evaluación a otros profesionales describen con precisión la finalidad, normas, validez, fiabilidad y aplicaciones de los procedimientos, así como cualquier calificación especial aplicable a su uso.
- b) Los psicólogos 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.
- c) Los psicólogos mantienen la responsabilidad de la aplicación adecuada, la interpretación y el uso de los instrumentos de evaluación, tanto si son ellos mismos los que puntúan e interpretan los tests como si utilizan servicios automatizados u otros para hacerlo.

9. Explicación de los resultados de la evaluación

A menos que se informe claramente con antelación a la persona evaluada que la evaluación excluye cualquier tipo de explicación de los resultados (como ocurre con algunas situaciones de consejo organizacional o ciertas evaluaciones de

preempleo, seguridad o del ámbito jurídico), los psicólogos han de garantizar que se proporcione una explicación de los resultados, utilizando un lenguaje que pueda entender la persona evaluada u otra persona legalmente autorizada por el cliente. Independientemente de que la puntuación e interpretación sean hechas por el propio psicólogo por ayudantes o por medio de servicios automatizados u otros servicios externos, los psicólogos toman las medidas necesarias para asegurar que se proporcionan explicaciones apropiadas de los resultados.

10. Mantenimiento de la seguridad de los tests

Los psicólogos se esfuerzan para mantener la integridad y seguridad de los tests y otras técnicas de evaluación, de acuerdo con la ley y las obligaciones contractuales, y de tal modo que les permitan cumplir con los requerimientos del código ético.

Directrices para los tests informatizados

La *American Psychological Association* (1986) propuso unas directrices para guiar el desarrollo, el uso y la evaluación técnica de los tests informatizados. Estas directrices pretenden garantizar que los tests informatizados cumplen los requisitos esperables en todo instrumento científico.

1. Los efectos sobre las puntuaciones de los tests debidos a la aplicación por ordenador, y que no tienen que ver con los objetivos de la evaluación, deberían eliminarse o ser tenidos en cuenta para la interpretación de las puntuaciones.
2. Si se introduce algún cambio en el equipamiento estándar, condiciones o procedimientos, respecto a los que se describen en el manual de test o instrucciones de aplicación, debe demostrarse que no afecta de forma apreciable a las puntuaciones de test. De lo contrario, debería llevarse a cabo una calibración adecuada y documentada (véase directriz 16).
3. El entorno en el que se encuentra situada la terminal para realizar el test debe ser tranquilo, confortable y libre de distracciones.
4. Los ítems que se presentan en la pantalla deben ser legibles y estar libres de reflejos luminosos.
5. El equipamiento debe comprobarse sistemáticamente y mantenerse en condiciones adecuadas. No se debe aplicar el test en un equipamiento defectuoso. Si el equipamiento falla cuando se está aplicando el test, puede haber que volver a aplicar el test completo o parte de él.
6. Hay que supervisar la ejecución del test y procurar una asistencia adecuada al examinado si la necesita. Si es técnicamente posible, hay que avisar inmediatamente al responsable cuando ocurra alguna irregularidad.
7. Debe entrenarse de forma adecuada en el uso del ordenador a las personas examinadas, así como establecer procedimientos para eliminar cualquier posible

efecto sobre las puntuaciones del test que sea debido a la falta de familiaridad del examinado con el ordenador.

8. Deben llevarse a cabo ajustes razonables para aquellas personas que pudiesen tener desventajas para realizar el test informatizado. En los casos en los que la desventaja no pueda ser completamente solventada, las puntuaciones obtenidas deben interpretarse con suma prudencia.
9. 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 dicha persona.
10. La aplicación informatizada de los tests debería proporcionar a los examinados al menos el mismo grado de información y control editorial sobre sus respuestas que tendrían si se aplicase el test de forma tradicional.
11. Las personas a las que se aplica el test deberían ser claramente informadas de todos los aspectos de la ejecución que sean importantes para el resultado del test
12. El sistema de test informatizado debe presentar el test y la forma de respuesta de modo que no cause una frustración innecesaria o limite la ejecución de los examinados.
13. El sistema de test informatizado debe estar diseñado de modo que su mantenimiento y verificación resulten sencillos.
14. El equipo informática, el procedimiento y las condiciones bajo las cuales se obtuvieron los datos para las normas, la fiabilidad y la validez deben describirse con claridad para permitir la replicación exacta de las condiciones.
15. Los servicios de informatización de los tests deben establecer los procedimientos oportunos para asegurar la confidencialidad de la información y la privacidad de los examinados.
16. Cuando se interpreten puntuaciones de versiones informatizadas de tests convencionales, debe establecerse y documentarse la equivalencia de las

puntuaciones de las versiones informatizadas antes de proceder a utilizar las normas o puntos de corte obtenidos a partir de los tests convencionales. Las puntuaciones provenientes de aplicaciones convencionales y de tests informatizados pueden considerarse equivalentes cuando: a) el rango de las puntuaciones de las personas en ambas formas es muy similar; y b) las medias, variabilidad y forma de las distribuciones de las puntuaciones son aproximadamente las mismas o pueden hacerse similares mediante un reescalamiento de las puntuaciones de la versión informatizada.

17. Quienes realizan una versión informatizada de un test convencional deben aportar los datos sobre su validez.
18. Los servicios de tests deberían advertir a los usuarios de los problemas potenciales que pueden existir cuando las puntuaciones de una versión del test no son equivalentes a las de la versión a partir la cual se han establecido las normas.
19. El constructor de un test debería proporcionar estudios comparativos de las versiones convencional e informatizada para establecer la fiabilidad relativa de la aplicación informatizada.
20. No se puede asumir sin más la precisión de la puntuación y de la interpretación informatizadas. Los proveedores de servicios de tests informatizados deben comprobar y controlar la calidad del software y del hardware, incluyendo la puntuación, los algoritmos y el resto de los procedimientos descritos en el manual.
21. Los servicios de tests informatizados deben proporcionar un manual en el que se muestren los fundamentos y evidencia que justifican la interpretación de las puntuaciones de la versión informatizada.
22. El sistema de clasificación utilizado para elaborar informes de evaluación debe ser suficientemente consistente para el objetivo perseguido. Por ejemplo, en algunos casos es importante que la mayoría de los examinados sean clasificados en los mismos grupos si se les vuelve a pasar el test (asumiendo que la conducta en cuestión no ha cambiado).
23. Hay que proporcionar información a los usuarios de los servicios de elaboración de informes automatizados en relación con la consistencia de las clasificaciones,

incluyendo, por ejemplo, el número de clasificaciones y el significado que tiene para la interpretación el cambio de una determinada clasificación a las adyacentes.

24. Debe facilitarse a los usuarios de los tests las puntuaciones originales utilizadas para elaborar interpretaciones informatizadas. La matriz de respuestas originales debe proporcionarse o estar disponible para los usuarios que la soliciten, con la debida consideración para la seguridad del test y la privacidad de los examinados.
25. El manual, o en algunos casos el informe evaluativo, debe describir cómo se han derivado de las puntuaciones originales las distintas partes del informe automatizado.
26. Los informes evaluativos automatizados deben incluir información sobre la consistencia de las interpretaciones, así como advertir sobre los errores de interpretación más habituales.
27. Debe dejarse claro en qué medida las afirmaciones que se hacen en un informe evaluativo automatizado se basan en investigaciones cuantitativas o en una opinión clínica experta.
28. Cuando las afirmaciones del informe evaluativo se basan en una opinión clínica experta, hay que proporcionar a los usuarios la información que les permita sopesar la credibilidad de dicha opinión.
29. Cuando las predicciones de ciertos resultados o recomendaciones específicas se basan en investigaciones cuantitativas, debe proporcionarse información que muestre las relaciones empíricas entre la clasificación y la probabilidad de la conducta criterio en el grupo de validación.
30. Los servicios de tests informatizados deben garantizar que los informes tanto para los usuarios como para los examinados son comprensibles y establecen claramente los límites dentro de los cuales pueden extraerse conclusiones precisas, teniendo en cuenta variables tales como la edad o sexo, las cuales pueden modular las interpretaciones.
31. Debe facilitarse información adecuada a los profesionales cualificados, implicados en la revisión especializada de los servicios automatizados, sobre el sistema, así como un acceso razonable a la forma de evaluación de las respuestas.

Cuando sea necesario proporcionar secretos comerciales, se establecerá un acuerdo escrito para evitar su difusión.

Principios para el desarrollo de tests informatizados

En la experiencia adquirida después de la elaboración de unos 1000 tests informatizados y cuatro versiones de su batería CAM (del inglés, *Cognitive Abilities Measurement*) en seis años de dedicación, Kyllonen ha elaborado ciertos principios aplicables para la elaboración de tests informatizados (Kyllonen, 1991, 1994 y 1996). Los principios directamente relacionados con la evaluación informatizada son los siguientes:

1. Es conveniente considerar la modalidad de presentación de los estímulos más adecuada en función del test administrado. Cuando se administra un test informatizado, la presentación de los estímulos más obvia es mediante el uso del monitor del ordenador. No obstante, resulta interesante utilizar modalidades alternativas de presentación de los estímulos (como la presentación auditiva mediante los dispositivos multimedia del ordenador).
2. Es conveniente utilizar diferentes modalidades de registro de las respuestas del sujeto en función del test administrado o de los sujetos evaluados. Habitualmente, el registro de las respuestas de los sujetos se realiza mediante el teclado del ordenador. Otras alternativas para el registro de respuestas de los sujetos son la implementación de teclados especialmente diseñados para el test administrado.
3. Se ha de considerar si las características de los ítems son las más adecuadas para el rasgo latente medido. La utilización del ordenador permite la presentación de ítems o estímulos de un modo mucho menos rígido que cuando el test se administra en una versión clásica de papel y lápiz.
4. La administración informatizada de cuestionarios permite alterar el orden de los ítems o de los diferentes subtests que componen una determinada batería de un modo mucho más simple que en las versiones de papel y lápiz.

5. Mientras que los ítems de un cuestionario de papel y lápiz son estáticos, los cuestionarios informatizados permiten generar ítems durante la administración del test cuya dificultad es incrementada de forma progresiva y sistemática.
6. Los tests informatizados suelen diferir bastante respecto a sus versiones clásicas de papel y lápiz. La consecuencia más directa es que las instrucciones de los primeros no deberían basarse de forma directa en las instrucciones que fueron escritas para los segundos. Los siguientes aspectos han de ser considerados:
 - Las instrucciones deberán ser reelaboradas y especialmente diseñadas en la administración informatizada de un test.
 - Mientras que las instrucciones de un test de papel y lápiz se presentan en una sola página, las instrucciones de un test se debería presentar utilizando una pantalla por cada idea que se le quiere comunicar al sujeto evaluado.
 - Es preferible iniciar las instrucciones del cuestionario informatizado con el ejemplo de un ítem y el modo de responderlo. El sujeto evaluado parece distraerse cuando se le facilita información referente a, por ejemplo, la finalidad del test que se le administra. Si se cree conveniente facilitar información complementaria, Kyllonen (1991, 1996) aconseja facilitarla después de administrar el cuestionario. Un ejemplo es la utilización de una ventana con un ítem de ejemplo en la zona superior de la pantalla y de una segunda ventana en la zona inferior de la pantalla con comentarios de cómo solucionar o responder al ítem. Por otra parte, se recomienda dar las mínimas instrucciones que sean necesarias.
 - La utilización sistemática de colores ayuda minimizar la confusión del sujeto y a separar los diferentes componentes del test informatizado. Kyllonen (1991, 1996) propone una estandarización del uso de los colores: el color blanco sobre negro para el título del test; blanco sobre *cyan* (azul turquesa) para las instrucciones del test; amarillo sobre azul para los ítems; amarillo sobre rojo para instrucciones de velocidad y precisión de las respuestas, que deberían ser presentadas inmediatamente antes del primer ítem; y blanco sobre verde para la motivación del sujeto al finalizar el test. Por otra parte, la ventana de

instrucciones informa de cómo responder, mientras que sucesivas pantallas presentadas a continuación instruirán al sujeto de cómo pasar a la siguiente pregunta, de cómo revisar las respuestas dadas, o de cómo finalizar el cuestionario.

- Las instrucciones deben garantizar que todos los sujetos responden al cuestionario bajo las mismas consignas respecto a cómo optimizar su rendimiento en el test. En el caso de test de velocidad, las confusiones más habituales ocurren cuando el sujeto no es capaz de discernir si se requieren respuestas especialmente rápidas o especialmente precisas. Kyllonen (1991, 1996) propone estandarizar tres instrucciones: a) *Sólo importa la precisión, el ordenador no registra cuanto tiempo utiliza usted*; b) *Responda tan rápido como pueda sin cometer demasiados errores*; y c) *Su puntuación depende principalmente de sus aciertos, pero se registra la velocidad de respuesta*. La primera instrucción estimula la precisión de la respuesta, la segunda la velocidad en las respuestas, y la tercera pretende potenciar la precisión y la velocidad simultáneamente. Otra solución es computar una única puntuación en función de la precisión y la velocidad de respuesta (donde los errores o respuestas muy lentas resten puntuación al sujeto), pudiendo ser el sujeto informado después de cada respuesta de cuál es su puntuación actual.
7. Un problema que se observa en la administración de tests informatizados es que los sujetos que emplean más tiempo en responder pueden obtener mejores puntuaciones, ya que su precisión aumenta considerablemente. Para corregirlo se han propuesto diferentes soluciones: fijar el tiempo de exposición de cada ítem o bien de ciertos bloques de ítems; elaborar un modelo estadístico que permita predecir la relación entre la precisión y la velocidad; combinar ambas soluciones previas utilizando diferentes tiempo de respuesta con un modelo estadístico final; informar después de cada ítem al sujeto de su precisión o de su velocidad de respuesta indicándole, si es necesario, que podría intentar responder más rápido o más lento.

8. La utilización del ordenador como administrador de tests permite informar al sujeto de su rendimiento durante la propia administración. Si bien informar al sujeto de si la última respuesta ha sido *correcta* o *incorrecta* parece mejorar su rendimiento, el informar sobre su velocidad de respuesta parece llevar a más errores o respuestas anticipadas. El informar del rendimiento promedio de la precisión y de la velocidad de respuesta después de un grupo de respuestas parece ser la opción más adecuada. Por otra parte, al finalizar el cuestionario es importante presentar una última pantalla de motivación, así como una breve descripción de la finalidad del test administrado.
9. Cuando se administran test informatizados es importante implementar al test con funciones que calculen el rendimiento en el test al finalizar el mismo. Kyllonen (1991, 1996) propone calcular los tantos por ciento de las respuestas correctas e incorrectas, la media y la mediana de las latencias para respuestas correctas, incorrectas y para el total del test. No obstante, es importante grabar todas las respuestas en un formato portable (por ejemplo, ASCII) de tal modo que puedan ser reanalizadas en el futuro.
10. La mayoría de test informatizados utilizan formatos de respuesta binarios o Likert simulando a los tests de papel y lápiz. No obstante, el ordenador permite utilizar otros formatos de respuesta, como son formatos de respuesta continua mediante la utilización de barras correderas. El tratamiento de respuestas abiertas introduce problemas más complejos, si bien se están planteando ciertas soluciones aun incipientes (por ejemplo, Bejar y Bennet, 1999).
11. Finalmente, es importante considerar los aspectos relacionados con la equivalencia del test informatizado que se administra con la versión (si existe) del test de papel y lápiz.

Directrices para los Tests Adaptativos Informatizados

Las directrices específicas para los Tests Adaptativos Informatizados fueron propuestas por Green, Bock, Humphreys, Linn y Reckase (1984) y finalmente estructuradas por Hambleton, Zaal y Pieters (1991).

Contenidos

1. Las especificaciones sobre el contenido de los ítems deberían ser las mismas para los Tests Adaptativos Informatizados que para los tests de papel y lápiz.
2. El contenido de los ítems seleccionados para el Banco de Ítems debe responder a las especificaciones de contenido hechas.
3. Los ítems deben diseñarse de modo que se ajusten a las características del equipamiento informático.

Dimensionalidad

4. Debe comprobarse el ajuste del modelo de Teoría de Respuesta a los Ítems utilizado.
5. Deberían seleccionarse ítems con un alto poder discriminativo.
6. Debería llevarse a cabo un análisis factorial de la matriz de correlaciones tetracóricas entre los ítems.
7. Hay que examinar la asunción de independencia local.
8. Cuando el test no es unidimensional, deberían formarse subtests.
9. Hay que balancear el test, para que refleje la heterogeneidad de los contenidos del dominio y los distintos tipos de formato.

Fiabilidad

10. Debe proporcionarse el error típico de medida de cada puntuación del test en función de la puntuación y en la métrica en la que se estén expresando las puntuaciones.
11. Hay que proporcionar también el error típico de medida de cada test en la métrica de la aptitud.

Validez

12. Hay que evaluar la similitud de las matrices de varianzas–covarianzas de los Tests Adaptativos Informatizados y los de papel y lápiz.
13. Debe compararse la estructura de las covarianzas de las dos versiones.
14. La versión adaptativa informatizada y la de papel y lápiz de un test deberían validarse frente al mismo criterio externo.
15. Hay que estudiar el posible sesgo predictivo para las subpoblaciones de interés.

Estimación de los parámetros de los ítems

16. El tamaño de la muestra para la calibración de los ítems debe ser el adecuado; en la actualidad, al menos, de 1.000 casos.
17. La muestra de calibración debe seleccionarse de tal modo que haya un número suficiente de personas en el rango de aptitud necesario para estimar la asíntota inferior y el punto de inflexión de la curva característica del ítem.
18. Hay que asegurarse de que el procedimiento para estimar los parámetros de los ítems sea empíricamente consistente (las muestras amplias deberían generar buenas estimaciones).
19. Ha de mostrarse que el procedimiento utilizado para la estimación de los parámetros de los ítems es insesgado.
20. Las curvas características de los ítems deben ajustarse a los datos empíricos.

21. Hay que comparar la dificultad de los ítems aplicados mediante la versión adaptativa informatizada y la correspondiente de papel y lápiz.

Métrica común para los ítems

22. El procedimiento utilizado para transformar los ítems en una escala común debe describirse con precisión.
23. Cuando para transformar los ítems en una misma escala se utiliza el procedimiento de grupos equivalentes, hay que demostrar que los grupos son realmente equivalentes.

Características del Banco de Ítems

24. Deben presentarse la distribución de las estimaciones de los parámetros de los ítems y los estadísticos descriptivos de las estimaciones.
25. Hay que ofrecer la función de información para todo el Banco de Ítems

Selección de los ítems y puntuación del test

26. El procedimiento para la selección de los ítems y la estimación de la aptitud debe documentarse con detalle.
27. El procedimiento debe incluir algún método para diversificar los ítems seleccionados, de modo que se evite el uso de unos pocos ítems exclusivamente.
28. El algoritmo utilizado debe ser capaz de aplicar determinados ítems y registrar las respuestas separadamente, sin interferir en el proceso adaptativo.
29. El ordenador debe ser capaz de basar la elección del primer ítem en información previa.

Factores humanos

30. El entorno de la terminal donde se realiza el test debe ser tranquilo, confortable y estar libre de distracciones.
31. La pantalla debe ubicarse de modo que permita una correcta visibilidad.
32. La legibilidad de la pantalla debe evaluarse empíricamente.
33. La pantalla debe permitir la inclusión de diagramas que exigen mucho detalle.