

ASPECTOS INDIVIDUALES QUE INFLUENCIAN EN EL DESEMPEÑO DURANTE EL EXAMEN PSICOLÓGICAS

Thomas Oakland; César Merino Soto

RESUMEN – Las conductas de examen tienen un importante impacto sobre el rendimiento en las pruebas de desempeño máximo, como ocurre en las pruebas cognitivas para niños; de este modo, la observación y registro de las conductas de examen es una acción necesaria del psicólogo evaluador. Aún cuando ha prevalecido un enfoque clínico, una evaluación más formal y normativa de las conductas de examen solo recientemente ha emergido. Actualmente se ha elaborado y marco de referencia teórico y técnico para hacer que la observación de las conductas de examen permita sea normativa y más informativa que la sola observación informal. El presente artículo expondrá brevemente los hallazgos en esta área de evaluación.

PALABRAS CLAVE: Pruebas psicológicas. Niño. Conducta y mecanismos de conducta.

ASPECTOS INDIVIDUALES QUE INFLUENCIAN EL DESEMPEÑO EN LAS PRUEBAS

La evaluación psicológica es un proceso en que psicólogos capacitados observan cuidadosamente y registran el desempeño actual de las personas bajo condiciones estandarizadas¹. Este proceso hace uso del arte y de la ciencia. La ciencia da las reglas que generalmente gobiernan la obtención, registro e interpretación de los datos. El arte de evaluar surge de la sensibilidad de los psicólogos hacia importantes conductas relacionadas con la prueba, que junto con su experiencia de evaluación a sus clientes, son cualidades que les permiten hacer impor-

tantes decisiones cuando examinan e interpretan sus datos.

Prácticamente, no hay psicólogo egresado que no tenga en su preparación experiencias de evaluación individual, y ensayos y errores en sus observaciones clínicas. Su preparación formal y su sensibilidad convergen en su habilidad para hacer conclusiones precisas, de tal modo que integra el arte y la ciencia.

Abordaremos algunos estrechamente asociados al tema central del presente artículo, sin agotar todo lo investigado. Haremos referencia a la evaluación de los niños, aunque los principios expuestos podrían ser interpolados a mayores de 16 años.

Thomas Oakland - Ph.D. Professor of Educational Psychology Universidad de Florida.

César Merino Soto - Psicólogo licenciado; Profesor en la Universidad Privada San Juan Bautista y Universidad San Martín de Porrer.

Correspondência

Cesar A. Merino Soto

Calle Enrique Palacios 430, Chorrillos – Lima 9 – Peru

Tel.: 51-(1) 251-8583.

E-mail: sikayax@yahoo.com.ar

LA EVALUACIÓN COMO CIENCIA Y ARTE

Las observaciones clínicas son críticas para la ciencia y el arte de evaluar psicológicamente. Las observaciones de los psicólogos les permiten registrar con precisión las conductas, entender la manera que los niños llegan a sus respuestas e identificar las fuerzas y debilidades cognitivas, temperamentales y otras personales, de tal modo que les facilitan sus interpretaciones.

LA EVALUACIÓN DEL DESEMPEÑO TÍPICO O DESEMPEÑO MÁXIMO

El proceso de evaluación, como es llevado en la práctica, se ajusta a dos metas que puede tener en mente el examinador: la meta de evaluar el desempeño máximo o el desempeño típico². En la primera situación, generalmente el psicólogo trabaja para crear las condiciones que permiten a los niños hacer lo mejor que puedan cuando son evaluados en su rendimiento escolar, inteligencia o habilidades perceptuales; aquí se trata de determinar las máximas capacidades y habilidades, no meramente su desempeño típico, usual o como normalmente ocurre. En contraste, en la segunda situación, tenemos que el estándar de desempeño típico enfatiza la importancia de adquirir una comprensión de las cualidades tal como ellas aparecen normalmente, no como su mejor desempeño posible. Este tipo de estándar de evaluación se emplea comúnmente en la evaluación de atributos sociales y de personalidad. En resumen, dos tipos de estándares de evaluación del desempeño deben figurar en las metas de evaluación; y estas metas deben dirigir la observación del psicólogo durante el proceso de evaluación.

ATRIBUTOS QUE IMPACTAN EN EL DESEMPEÑO DURANTE LA EVALUACIÓN

La investigación clínica y psicométrica^{3,4} ha señalado que los siguientes ocho atributos pueden impactar sobre el desempeño de un niño en las pruebas:

- *Condiciones ambientales de la evaluación* -

La oficina debe estar libre de distracciones, apropiados muebles y espacio, etc.

- *Habilidades del lenguaje* - El lenguaje receptivo y expresivo deben estar lo suficientemente desarrolladas como para servir de un vínculo confiable entre el examinador y el niño. Las diferencias en estas habilidades de lenguaje sobre el desempeño de las pruebas pueden ser un factor mediador del éxito en las pruebas incluso en preescolares, durante la administración de pruebas grupales.
- *Condiciones relacionadas con la salud* - El desempeño del niño no debería ser atenuado por alguna enfermedad, dolor o malestar.
- *Rapport* - La relación interpersonal entre el examinador y el niño debe contener confianza de cara al desempeño en la prueba.
- *Disposición para la evaluación*- Autoconfianza, voluntad para ser evaluado durante el tiempo destinado para ello.
- *Motivación* - Es decir, estar comprometido con el desempeño en las tareas y mantener la motivación durante la actividad de examen.
- *Temperamento* - El tipo temperamental predominante del niño, como ocurre con los niños que son extrovertidos, quienes tienden más a expresar sus ideas y hablar sobre ellos mismos mientras están siendo evaluados. Las diferencias entre los estilos temperamentales es, por lo tanto, un área no menos importante de observación.
- *Severidad de alguna condición discapacitante* - Dado una discapacidad presente, por ejemplo, en niños con discapacidades mentales (profundas o severas), se necesitarán más modificaciones del proceso de evaluación.

ESTÁNDARES PARA LA EVALUACIÓN DEL DESEMPEÑO EN LAS PRUEBAS: PERFECCIÓN, POTENCIAL Y PARES

Dado el deseo de crear las condiciones que permiten a los niños hacer lo mejor que puedan durante el examen, los psicólogos juzgan si las condiciones presentadas son suficientemente apropiadas para que los niños puedan desempeñarse óptimamente. Así, los psicólogos usan

uno de los siguientes estándares para hacer este juicio: perfección, *potencial o normatividad*.

La *perfección* se refiere a las condiciones de prueba y las conductas del niño en la prueba que ocurrieron impecablemente; sin embargo, esto raramente ocurre. Más aún, los psicólogos más probablemente difieren en sus descripciones sobre este estándar. Así que, las decisiones basadas sobre este estándar son probablemente no confiables. *Potencial* se refiere a si las condiciones de prueba fueron tan buenas como las que podían ser esperadas. Este estándar también es difícil de describir y aplicar confiablemente. *Normatividad*, o el uso de un estándar normativo, descansa sobre la información derivada desde medidas bien validadas y estandarizadas nacionalmente, diseñadas para evaluar la conducta relacionada a las pruebas en relación a un numeroso grupo de niños. Cabe señalar que el uso de este último estándar no nos excluye del uso de los dos estándares anteriores.

PROBLEMAS ASOCIADOS CON EL REGISTRO DE LAS CONDUCTAS BASADAS EN LA OBSERVACIÓN CLÍNICA

Los psicólogos típicamente confían en sus observaciones clínicas para hacer juicios sobre la calidad de las conductas del niño durante la prueba. Esta es solo una aparente seguridad pues tiene varias desventajas. Identificaremos algunas de ellas:

Las cualidades identificadas pueden ser irrelevantes - Clínicos difieren en lo que ellos creen sobre las conductas que importantemente impactan sobre el desempeño en las pruebas. Las cualidades pueden emanar del folclor popular. Los clínicos pueden aceptar las creencias comúnmente aceptadas e incorporarlas en su práctica evaluativa. Por ejemplo, hay poco apoyo científico para lo que frecuentemente se ha sostenido que las conductas relacionadas con la personalidad observada durante la ejecución de la prueba reflejan atributos de personalidad perdurables^{5,6}.

Diferencias relacionadas con la edad pueden ser pasadas por alto - Los psicólogos pueden

ser insensibles a sutiles pero importantes diferencias relacionadas con la edad en la conducta de examen⁶.

Las observaciones no son estructuradas - Los psicólogos generalmente no son confiables en lo referente a qué conductas de prueba registrar y dónde registrarlas. Es usual encontrar psicólogos que no pueden justificar sus conclusiones sobre las conductas de prueba, basados sobre las evidencias registradas. Es curioso que exista un pobre énfasis en documentar completamente el comportamiento durante la prueba, considerando las exigencias éticas y legales que sobre el proceso evaluativo se establecen actualmente. Esta situación pone en un compromiso la habilidad del psicólogo para justificar sus conclusiones y encarar el examen externo de otras personas.

Los psicólogos no tienen el beneficio de usar interpretaciones referidas a normas - Esto constituye la más grande limitación cuando se cuenta solamente con observaciones clínicas. Las conductas que son consideradas impactantes sobre el desempeño en la prueba raramente son derivadas de las evidencias de la investigación². En efecto, solo hasta hace poco, la investigación sobre las conductas de prueba asociadas con instrumentos administrados individualmente han sido escasos. Por ejemplo, solo debido a su inclusión como parte del registro de prueba de la Escala de Inteligencia Stanford-Binet^{7,8}, el Protocolo de Observación Stanford-Binet (SBOS⁹), es quizás la medida más popular para registrar la conducta durante la sesión de prueba. Aún con esto y no obstante sus 40 años de historia en el uso y aceptación clínica, solamente un estudio⁶ ha dado alguna evidencia de su validez y confiabilidad.

INVESTIGACIÓN SOBRE EL COMPORTAMIENTO DE PRUEBA: VALIDEZ DE CONSTRUCTO

Tres aspectos de los factores de conducta de prueba influyen aparentemente el desempeño en la prueba¹⁰: *la evitación*, *el ánimo no cooperativo* y *la desatención*.

La evitación se refiere a la tendencia de un evaluado a evitar las tareas y expresar temor;

ejemplos de ello incluyen mostrar un marcado desinterés en las actividades de la prueba y dubitación al dar las respuestas.

El *ánimo no cooperativo* se refiere al irregular ajuste del niño a la situación de prueba, falta de cooperación y necesidad de elogios e incentivos. Algunos ejemplos pueden incluir el deterioro del desempeño al final de la prueba y preguntar cuándo terminará la prueba.

La *desatención* se refiere al inadecuado control de impulsos y de la atención, y que pueden incluir la dificultad para escuchar atentamente las instrucciones e ítems de la prueba, y los intentos de responder antes que las preguntas sean terminadas.

Estos tres factores están correlacionados y en consecuencia forman un factor común sobre la conducta de prueba¹⁰.

GUÍA DE EVALUACIÓN DE LA CONDUCTA DURANTE LA SESIÓN DE PRUEBA (GATSB, GUIDE TO THE ASSESSMENT OF TEST-SESSION BEHAVIOR)

Esta es la primera medida referida a normas que fue nacionalmente estandarizada en los Estados Unidos, respecto a la habilidad del evaluado para desempeñarse en las pruebas. El GATSB fue co-normado con el WISC-III y el WIAT, y por lo tanto esta orientada hacia los niños entre 6 y 16 años. Algo de la investigación con el GATSB se resume a continuación.

Validez de Constructo

El posible sesgo del constructo en el GATSB se investigó¹¹ considerando el género, la edad, etnicidad – raza (anglos, afroamericanos e hispanos) y nivel socioeconómico (alto, medio y bajo) con la muestra de estandarización de la escala. Los coeficientes y congruencia factorial fueron lo suficientemente altos para concluir que los tres factores del GATSB son comparables a través del género, edad, raza y nivel socioeconómico.

Validez Concurrente

La correlación entre el Puntaje Total GATSB y

el CI Total del WISC-III es sustancial: $-.36$. Es decir, el 13% de la varianza asociada con los puntajes de la Escala Total del WISC-III pueden ser atribuidos a las conductas de prueba durante el examen. Entre estos tres factores, Evitación absorbió la pluralidad de la varianza. Los niños que muestran deficientes conductas de prueba generalmente obtienen puntajes de CI Total de 7 a 10 puntos más abajo que los niños con conductas de prueba más apropiadas. Una magnitud del efecto así de grande (es decir, más de la mitad de una desviación estándar) representa una sustancial diferencia en los CIs. Entonces, estamos advertidos que dentro de los resultados de una prueba de inteligencia, el WISC-III por ejemplo, el efecto de las conductas no asociadas con la habilidad medida no es algo que debemos ignorar.

La correlación entre el GATSB Total y el puntaje compuesto total del WIAT es también sustancial: $-.38$. Es decir, aproximadamente el 14% de la varianza asociada con el rendimiento puede ser atribuido a las conductas durante la situación de prueba. Las correlaciones entre el GATSB Total y los cuatro puntajes compuestos del WIAT son también similares.

Correlaciones con el Origen Étnico

Como se anotó arriba, la correlación entre el GATSB Total y el puntaje total del WISC-III es $-.36$ para el grupo total. Esta correlación es $-.28$ para los anglos, $-.37$ para afroamericanos y $-.55$ para hispanos. Por lo tanto, el monto de varianza explicada atribuida al comportamiento durante la prueba sobre el WISC-III es del 14% para los niños afroamericanos y 30% para los niños hispanos. Estas diferencias en la varianza nos sugieren que las conductas de examen impactan diferentemente dado la procedencia étnica del niño¹². Sin embargo, estos hallazgos deben ser juzgados juiciosamente, pues en el marco de las investigaciones de origen (contexto norteamericano) la minoría de los niños "no blancos" es la mayoría de niños en sus países de origen; adicionalmente, la magnitud de la relación puede hacer sugerir su importancia teórica, pero de menor efecto considerando su importancia práctica.

Posibles Diferencias de Edad y Género

Como se puede esperar, niños de mayor edad muestran mejores comportamiento de examen que los niños más jóvenes. Entonces, se deberían utilizar normas de edad separadas cuando se evalúan las conductas de prueba a los niños. En cambio, los niños y niñas no difieren significativamente¹³.

Confiabilidad

Las puntuaciones obtenidas del GATSB son estables. De este modo, una estimación de la confiabilidad test-retest en una muestra de 300 niños y niñas, alcanzó una correlación de 0.87 para el puntaje total. Por otro lado, una estimación de la confiabilidad por consistencia interna alcanzó un coeficiente de 0.92 para el puntaje total. Estos niveles de confiabilidad son considerados generalmente óptimos para usar un instrumento a nivel clínico¹⁴, y por lo tanto, para extraer conclusiones sobre el desempeño individual del niño evaluado.

CONCLUSIONES

Los resultados subrayan la importancia de observar rutinariamente las conductas de examen, juzgándolos por estándares normativos y determinando si el desempeño del niño negativamente impactar la medición de su inteligencia y de su desempeño cognitivo en

general durante la evaluación individual. Los psicólogos pueden estar particularmente atentos a los efectos de las diferencias socioeconómicas, pero la investigación empírica aún no ha respondido a esta alerta. Las diferencias étnicas se han observado, pero se debe hacer un escrutinio en países de habla hispana.

Hay varias rutas de investigación que no han sido abiertas, o al menos mínimamente publicado. Nos podemos referir a la evaluación psicológica en casos de maltrato a niño. Por ejemplo, nos preguntaríamos sobre el impacto de las experiencias de maltrato recibidas o atestiguadas por los niños(as) en su vida familiar o escolar, sobre las conductas de prueba tal como son medidas por el GATSB. Aunque podemos anticipar desde la experiencia clínica un patrón consistente de evitación y ánimo no cooperativo, estas pueden estar moderadas por el género del investigador; o quizás del efecto proveniente del maltrato puede ser mínimamente útil para la práctica evaluativa.

Esperamos que el presente artículo aguijonee a los lectores, para dirigir su mirada hacia lo simplemente importante: las conductas de examen de los niños(as) evaluados, tal como son registradas por el GATSB o por incluso instrumento similarmente útil dentro del marco de su evaluación formal y estandarizada.

SUMMARY

Individual aspects impact on the performance during psychological tests

The test-taking behaviors have an important impact on the performance in the maximum performance test, like it happens in the cognitive tests for children; this way, the observation and recording these behaviors is the necessary action of the psychologist assessor. Still when a clinical focus, justly it has emerged a more formal and more normative assessment of the taking-test behaviors. Currently, technical and theoretical framework it has been elaborated make that the observation of the test-taking behaviors allows it is normative and more informative that the single informal observation. The present article will briefly present the findings in this assessment topic.

KEY WORDS: Psychological tests. Child. Behavior and behavior mechanisms.

REFERENCIAS

1. Sattler JM. Evaluación de la inteligencia infantil y habilidades especiales. 2^{da} ed. México: El Manual Moderno;1988.
2. Brown FG. Principios de la medición en psicología y educación. México:El Manual Moderno;1980.
3. Glutting J, Oakland T, McDermott P. Observing child behavior during testing: Constructs, validity, and situational generality. *J School Psych* 1989;27:155-64.
4. Green K, Forehand R, Beck S, Vosk B. An assessment of the relationship among measures of children's social competence and children's academic achievement. *Child Dev* 1980;51:1149-56.
5. Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psych Bull* 1987;101:213-32.
6. Glutting J, Oakland T. Guide to the assessment of test session behaviors for the WISC-III and WIAT. San Antonio:The Psychological Corporation;1993.
7. Terman LM, Merrill MA. Stanford-Binet intelligence scale: form L-M. Boston: Houghton Mifflin;1960.
8. Thorndike RL, Hagen EP, Sattler JJ. Stanford Binet intelligence scale. 4th ed. Chicago: Riverside;1986.
9. Thorndike RL, Hagen EP, Sattler JJ. Stanford Binet observation schedule. Chicago: Riverside;1986.
10. Oakland T, Glutting J. Assessment of test behaviors with the WISC-III. In: Prifitera A, Saklofske D, eds. WISC-III; a scientist-practitioner perspective. New York:Academic Press;1998.
11. Konold T, Glutting J, Oakland T, O'Donnell L. Congruence of test-behavior dimensions among child groups that vary in gender, race-ethnicity, and SES. *J Psychoed Assess* 1995;13:111-9.
12. Glutting J, Oakland T, Konold R. Criterion-related bias with the guide to the Assessment of test-session behavior for the WISC-III and WIAT: possible race, gender, and SES effects. *J School Psych* 1994; 32:355-69.
13. Oakland T, Glutting JJ. Examiner observations of children's WISC-R test-related behaviors: possible sex, race, and gender effects. *Psych Assess: Cons Clin Psych* 1990; 2(1):86-90.
14. Nunnally JC, Bernstein J. Teoría psicométrica. 3^{ra} ed. México:McGraw-Hill;1995.