EVIDENCIAS DE VALIDEZ DE CRITERIO CONCURRENTE DEL EXAMEN PROGRAMA DE DESARROLLO INTEGRAL DE LA PERSONA (PRODIP-R)

María del Rayo López Contreras
rayolopezc@gmail.com

Resumen

El examen Programa de Desarrollo Integral de la Persona es un instrumento de medición a gran escala. En la Universidad Autónoma de Baja California (UABC) es aplicado a los aspirantes de ingreso (aproximadamente 30, 000 al año) para identificar el nivel de dominio de 12 habilidades del pensamiento; con base en sus resultados se da atención psicopedagógica a los estudiantes.

No obstante la relevancia de la utilidad del PRODIP-R, son pocos los estudios que aportan evidencias de su validez. En consecuencia de lo anterior y debido a que se aplica a una cantidad importante de aspirantes, el presente trabajo tiene como objetivo informar los resultados que se obtuvieron de la investigación realizada para aportar evidencias de validez de criterio del PRODIP-R

La muestra estuvo integrada por 827 aspirantes a ingresar a la UABC en el periodo 2009-2 que contestaron ambos instrumentos. El personal autorizado por la Universidad y encargado de los resultados de cada examen integró la base de datos utilizada para la investigación realizada. Tras su entrega se hicieron los cambios pertinentes para su análisis, los cuales consistieron en obtener los indicadores psicométricos (dificultad, discriminación y confiabilidad) del PRODIP-R, además de estimar las correlaciones entre éste, el EXHCOBA (Examen de Habilidades y Conocimientos Básicos) y el promedio de las calificaciones del bachillerato.

En conclusión, se obtuvieron las correlaciones entre ambos instrumentos y el promedio de calificaciones del bachillerato, identificando así que, de manera general, existe una correlación positiva y elevada (r = .57) entre el EXHCOBA y el PRODIP-R; y Media Baja (r = .20) con el promedio de bachillerato, todas con un nivel de significancia a .01

.

Palabras claves: validez criterio concurrente, habilidades pensamiento.

Planteamiento del problema y justificación

El proceso de la evaluación educativa involucra el diseño, aplicación y análisis de los instrumentos de medición y una de sus funciones en educación es evaluar la capacidad de las personas y brindar información pertinente para la toma de decisiones (Backhoff, Larrazolo y Rosas, 2000). Messick (1993, como se citó en Velasco 2008) mencionó que es necesario que el instrumento se diseñe, construya y valide de tal manera que los resultados que se obtengan puedan ser interpretados como evidencia de que el sustentante tiene o no las habilidades que se evalúan. Así mismo, debido al impacto social que tienen los instrumentos de medición educativa a gran escala, deben ser validados por medio de rigurosos estándares de calidad establecidos por organismos internacionales y nacionales como: la American Educational Research Association (AERA); la American Psychological Association (APA); el National Council on Measurement in Education (NCME) y el Centro Nacional de Evaluación para la Educación Superior (CENEVAL) (AERA, APA y NCME, 2004; Linn, 2006; Martínez et al., 2000), entre otros.

En el ámbito estatal, cada año la Universidad Autónoma de Baja California (UABC) lleva a cabo dos convocatorias de ingreso a esta institución y para cada convocatoria realiza un proceso de admisión. En dicho proceso se obtiene información de los aspirantes (30, 000 al año aproximadamente) a través de diversas fuentes, entre ellas el examen psicométrico que está conformado por tres instrumentos: el Cuestionario de Rasgos de la Personalidad (CRP), el Inventario de Intereses Vocacionales (CIV) y el examen Programa de Desarrollo Integral de la Persona (PRODIP-R).

En particular, el PRODIP-R es un instrumento de medición a gran escala que identifica el nivel de dominio de 12 habilidades del pensamiento. Fue diseñado, elaborado, estandarizado y computarizado entre los años 1997 y 2001 por personal de la misma universidad (Cervantes, 2002). Entre el 2008 y 2009, debido a que no se contaba con información acerca del grado de validez y confiabilidad de esta prueba, un grupo de investigadores del Instituto de Investigación y Desarrollo Educativo (IIDE) realizó un diagnóstico y reingeniería del instrumento, con el objetivo de mejorarlo (Caso, 2010, comunicación personal).

El objetivo de aplicar el PRODIP-R es que con base en sus resultados el Departamento de Orientación Educativa y Psicológica (DOEP) brinde asesorías a los alumnos y docentes que lo soliciten sobre las habilidades del pensamiento del estudiante, con la finalidad de mejorar su rendimiento académico. Pese a que la utilidad del examen es relevante, es escasa la información sobre sus orígenes e insuficientes los estudios que aportan evidencias de su validez; por lo tanto, cabe la posibilidad de que: (a) la interpretación de los resultados que arroja este instrumento no corresponda al nivel de las habilidades del pensamiento que posee el aspirante; (b) las decisiones que se están tomando sean inadecuadas, y (c) no se están cumpliendo con los estándares de calidad que establecen los organismos nacionales e internacionales en materia de pruebas psicológicas y educativas. En consecuencia y debido a que el PRODIP-R se aplica a una cantidad considerable de aspirantes, y en virtud de que esta investigación brinda evidencias de validez del PRODIP-R, se logrará aportar resultados que ayuden a mejorar el instrumento, obteniendo con esto la posibilidad de mejorar las decisiones que de él se deriven. De igual manera será posible apoyar a que la universidad cuente con mejores instrumentos psicológicos propios ya que a la vez ayudará a cumplir con los estándares de calidad establecidos.

Fundamentación teórica

La presente investigación está basada en la Teoría Clásica de los Test (TCT), así como en la sexta versión de los estándares establecidos para las evaluaciones psicológicas y educativas, establecidos en 1999 por la American Educational Research Asociation (AERA), American Psychological Association (APA) y la National Council on Measurement in Education (NCME).

La Teoría Clásica de los Tests (TCT) consiste esencialmente en asumir que el puntaje que un individuo obtiene en un test (puntuación empírica: X) se conforma por dos componentes: la puntuación verdadera (V) y el error (e), este último puede deberse a muchas causas que no es posible controlar (Muñiz, 2010). Lo anterior se expresa formalmente como:

$$X = V + e$$

Estándares para las evaluaciones psicológicas y educativas

Existen diferentes asociaciones que establecen estándares para regular la adecuada elaboración y práctica de las pruebas psicológicas y educativas. Dichos estándares consisten en una serie de normas que se crean de acuerdo con el propósito de la prueba, con el fin de obtener resultados confiables y justos. Entre las asociaciones internacionales se pueden mencionar la AERA, APA y la NCME, y a nivel nacional el CENEVAL.

A través de los estándares se promueve el uso racional y ético de las pruebas, y se proporciona una base para evaluar la calidad de sus prácticas y los efectos del uso de las mismas. Según, la AERA, APA y NCME (1999) estas normas no poseen niveles jerárquicos entre ellas, es decir, son igualmente importantes; sin embargo, dependiendo del contexto, propósito del desarrollo o uso de la prueba algunas normas sobresaldrán más que otras.

En el contexto nacional, Martínez et al. (2000) establecieron las normas mexicanas de las evaluaciones educativas basándose en los estándares de las instituciones citadas en el párrafo anterior, para lo cual realizaron algunas adecuaciones aplicables al contexto nacional. Estos autores mencionan que las pruebas psicológicas y educativas necesitan ser elaboradas con los más altos estándares de calidad, y deben ser corregidas con base en sus resultados.

Validez de las pruebas psicológicas y educativas

La AERA, APA y NCME (1999) señalaron que la validez de las pruebas se refiere al grado en que la evidencia y la teoría apoyan la interpretación de los resultados, de acuerdo con el propósito para el cual fueron creadas. Por su parte, Messick (1995) la identificó como un concepto unificado y un proceso que consiste en la acumulación de evidencias para proporcionar una base científica para las interpretaciones de los resultados.

Fuentes de evidencias de validez

Acorde con el concepto unificado de validez, la AERA, APA y NCME (1999) refieren las evidencias de validez, más que a los tipos de validez. En la figura 1 se esquematiza la división de dichas evidencias

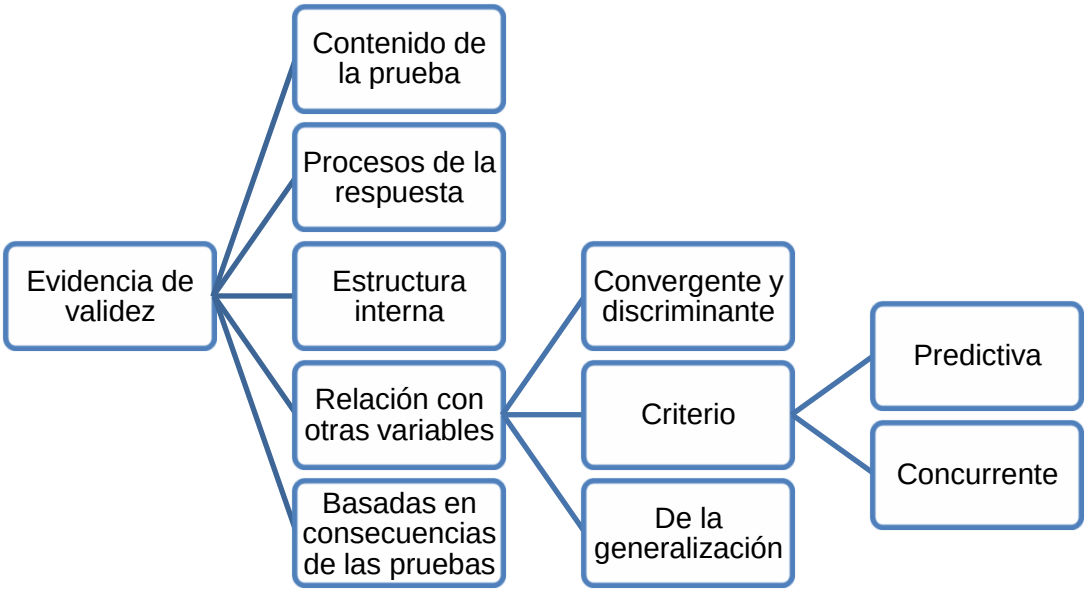

Figura 1. *Esquematización de las evidencias de validez de la AERA, APA y NCME (1999).*

Con base en lo anterior y debido a que el enfoque central del presente trabajo es la evidencia de validez de criterio concurrente a continuación se explica la misma: Criterio de la prueba: concebida como evidencias de validez vinculadas con criterios externos; son evidencias enlazadas con la significación de la relación entre los puntajes de una prueba y los puntajes de otro criterio. Es decir, determina en qué medida los puntajes en un test se relaciona con una variable criterio, la cual se asume que es una medida del mismo atributo que mide el test. Dentro de éstas evidencias de validez existen dos tipos: la predictiva y la concurrente. La predictiva es aquella en que la variable criterio con la que se contrasta la prueba ocurre después; mientras que en la concurrente la prueba y el criterio ocurren simultáneamente.

Objetivo

Aportar evidencia de validez de criterio concurrente del examen Programa de Desarrollo Integral de la Persona (PRODIP-R).

Metodología

En el periodo 2009-2, 1242 aspirantes contestaron el EXHCOBA y 857 el PRODIP-R, sin embargo, 30 de éstos últimos no respondieron el examen de ingreso. Por lo cual, 827 aspirantes cumplen con el criterio de contestar ambos exámenes para formar la muestra de este trabajo. De dicha muestra 504 (61%) fueron hombres y 323 (39%) mujeres. El rango de edad fue de 17 a 57 años, con una media de 21.47 y una desviación estándar de 5.10.

El PRODIP-R está basado en la teoría del intelecto propuesta por J. P.Guilford en 1959 (Sánchez & Caso, 2010). Éste instrumento identifica el nivel de dominio de 12 habilidades del pensamiento: ocho de lectura (mecánica, de comprensión e integración) y cuatro analíticopráctica (cálculo mental y razonamiento lógico). Esta conformado por 12 subescalas (una para cada habilidad) con un total de 258 reactivos, con cuatro opciones de respuesta (sólo una correcta). El PRODIP-R se aplica de manera individual a través de un sistema computarizado. Se considera importante indicar que para dar respuesta a algunos reactivos es necesario el uso de auriculares, a través de éstos se da la instrucción y el examinado elige, de las opciones presentadas en pantalla, la respuesta que considere correcta.

El EXHCOBA se administra de manera computarizada. Consta de 310 reactivos, de los cuales el estudiante tiene que responder 190, de acuerdo a la carrera que desee ingresar. Los primeros 130 reactivos los contestan todos los aspirantes, ya que miden las habilidades y conocimientos básicos que los estudiantes deben de adquirir durante la primaria y la secundaria. Este examen es de opción múltiple con una respuesta correcta, tres incorrectas y una más para indicar que el alumno no conoce la respuesta (Backhoff, 2003).

En cuanto al procedimiento de esta investigación. Al ser ambos instrumentos parte de la selección de ingreso a la UABC, su aplicación estuvo a cargo del personal encargado del proceso de admisión de la Universidad. En ambas aplicaciones los aspirantes se presentaron al laboratorio de cómputo que le correspondía, en donde se les asignó una computadora en la cual respondieron el examen.

Las respuestas de los reactivos se obtuvieron mediante dos archivos en Excel, uno correspondía al PRODIP-R y otro al EXHCOBA. Los dos archivos se exportaron al programa IBM SPSS Statistics 19, mediante la opción ¨*merge files*¨ originándose así una sola base de datos. Ésta estuvo conformada por: (a) las características de los estudiantes (edad, sexo, carrera y facultad), (b) las respuestas de cada reactivo tanto del PRODIP-R como del EXHCOBA; (c) la calificación estándar del EXHCOBA (sin tomar en cuenta el promedio del bachillerato) y (d) la calificación final de ambos exámenes. Adicionalmente se ingresó la calificación promedio del bachillerato de cada estudiante.

Resultados[1] y conclusiones

Con los resultados globales obtenidos se identifica que la dificultad del PRODIP-R es de .59, el Alpha de Cronbach de .94 y el promedio de discriminación de sus reactivos de .42 Además se ubica una correlación positiva, significativa y Alta (r = .57) con el EXHCOBA (y de Media a Alta con cada una de sus áreas de conocimiento, r = .37 a .48), y Media baja con el Promedio de bachillerato (r = .20), todas con un nivel de significancia a .01. Por lo tanto, con base en dichos estándares, es posible indicar que el PRODIP-R cuenta con algunas evidencias de validez de criterio concurrente. No obstante, en los resultados particulares de algunas de las subescalas se encontraron correlaciones no significativas, Malas o Bajas con ambos criterios externos, en específico con Memoria de Unidades Figurativas (MUF) y Evaluación de Sistemas Semánticos (ESM).

A través de los resultados es posible indicar que el PRODIP-R tuvo un nivel de dificultad Medio, una confiabilidad Elevada y una discriminación promedio de sus reactivos bastante Aceptable. En lo que a confiabilidad se refiere la AERA, APA y NCME (1999) mencionan que ésta es uno de los indicadores relevantes para la calidad de un instrumento. Por lo tanto, se concluye que la información que arroja el presente trabajo aporta una evidencia de confiabilidad del PRODIP-R.

Asimismo, en general éste instrumento presenta una correlación Alta con el EXHCOBA. Entre sus subescalas y las áreas de conocimiento, en su mayoría las correlaciones se presentaron de Media baja a

Alta con base en su fundamentación teórica, excepto en las subescalas MUF y ESM, ya que en algunos casos éstas correlacionaron Bajas, Malas o sin ningún nivel de significancia.

Respecto al promedio de bachillerato la puntuación fue Media baja (.20), lo que no es muy prometedor. Es importante encontrar una explicación aceptable de este hecho, pues sería esperable que un instrumento que mide habilidades intelectuales, se relacione estrechamente con el rendimiento académico de los estudiantes, que también depende, en gran medida, de las capacidades intelectuales.

Referencias bibliográficas

American Educational Research Asociation, American Psychological Association y National Council on Measurement in Education (1999). *Stardards for educational and psychological testing*.EstadosUnidos de América: American Educational Research Association.

Backhoff, E. (2003). *Desarrollo, validación e implementación de un sistema para la selección de estudiantes a la Universidad Autónoma de Baja California*. Tesis doctoral. Universidad Autónoma de Aguascalientes. Aguascalientes: México.

Backhoff, E., Larrazolo, N. y Rosas, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Electrónica de Investigación Educativa*, *2*(1). Recuperado de: http://redie.uabc.mx/vol2no1/contenido-backhoff.html

Cervantes, E. (2002). *Actualización del Examen Psicométrico de la Universidad Autónoma de Baja California*. Ponencia presentada en el tercer congreso de nacional y segundo internacional: retos y expectativas de la universidad, 6-9 de noviembre. Ixtapa de la Sal, Estado de México.

Linn, R. (2006). The standards for educational and psichological testing: guidance in test development. En S. Downing y T. Haladyna (Eds.). *Hadbook of test development* (pp. 27-38). Mahwah, NJ: Laurence Erlbaum Associates, Inc.

Martínez, F., Backhoff, E., De la Orden, A., Tristán, A., Shmelkes, S., Solano-Flores, et al. (2000). *Estándares de calidad para instrumentos de evaluación educativa*. México: Centro Nacional de Evaluación de Educación Superior.

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning *American psychologist 50*(9), 741-749.

Muñiz, J. (2010). La teoría de los tests: Teoría Clásica y Teoría de Respuesta a los Ítems. *Papeles del Psicólogo, Vol. 31* (1), pp 57-66. Recuperado de: http://www.papelesdelpsicologo.es/vernumero.asp?id=1796

Sánchez, C. y Caso, J. (2010a). *Actualización de las propiedades psicométricas y análisis factorial confirmatorio de los instrumentos que conforman el Examen Psicométrico de la UABC*. Documento no publicado: Universidad Autónoma de Baja California, Instituto de Investigación y Desarrollo Educativo. Mexicali, B. C.: México.

Velasco, V. (2008). *Evidencias de validez del Examen de Egreso del Idioma Inglés (EXEDII)* (Tesis de maestría inédita). Instituto de Investigación y Desarrollo Educativo, Ensenada, B. C.