



Para citar este artículo, le recomendamos el siguiente formato:

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). Consultado el día de mes de año en: <http://redie.uabc.mx/vol2no1/contenido-backhoff.html>

Revista Electrónica de Investigación Educativa

Vol. 2, No. 1, 2000

Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA)

The Difficulty Level and Discrimination Power of the Basic Knowledge and Skills Examination (EXHCOBA)

Eduardo Backhoff Escudero (1)
backhoff@uabc.mx

Norma Larrazolo Reyna (2)
normal@uabc.mx

y Martín Rosas Morales (3)
maromo@uabc.mx

(1), (2), (3) Instituto de Investigación y Desarrollo Educativo
Universidad Autónoma de Baja California

A. P. 453
C. P. 22830
Ensenada, Baja California, México

(Recibido: 28 de febrero del 2000; aceptado para su publicación: 28 de marzo del 2000)

Resumen

El Examen de Habilidades y Conocimientos Básicos (EXHCOBA) es uno de los pocos exámenes de gran escala en México que ha ido dando a conocer sus estándares de calidad. En este trabajo, se describen los resultados del análisis de reactivos, relativos al

nivel de dificultad y poder de discriminación del examen. Los resultados muestran que la mayoría de los reactivos del EXHCOBA tienen una dificultad media y un buen poder discriminativo. Asimismo, se observa que los reactivos que mejor discriminan pertenecen al área de matemáticas y los que poseen menor nivel de discriminación pertenecen al área de ciencias sociales.

Palabras clave: Exámenes de admisión, evaluación educativa, estándares psicométricos, educación superior.

Abstract

The Basic Knowledge and Skills Examination (EXHCOBA) is one of the few great-scale examinations in Mexico which has been publishing its psychometric parameters. In this paper we describe the item analysis results, regarding the exam's difficulty level and discrimination power. Results show that most of the items have a medium difficulty and a high discrimination power. They also reveal that the mathematics items have better discrimination power levels than the ones which belong to social science.

Key words: Admission tests, achievement tests, psychometric standars, higher education.

Introducción

La evaluación educativa es un proceso que involucra la elaboración, aplicación y análisis de los instrumentos de medición. La función principal de un instrumento de medición educativo, cuando se crea como medida para inferir las capacidades de las personas, es ofrecer información para la correcta toma de decisiones.

Cuando se utilizan instrumentos de gran escala y alto impacto, como en el caso de los exámenes de admisión, es necesario conocer los indicadores técnicos que definen la calidad del instrumento educativo que se emplea. Los exámenes de gran escala son aquellos diseñados para aplicarse en más de un plantel escolar, usualmente en un estado, región o país. Por su dimensión y por el poderoso impacto social que tienen, su elaboración debe ajustarse a rigurosos estándares de calidad (Aiken, 1996).

En 1985, la Asociación Americana de Psicología (APA), en colaboración con la Asociación Americana de Investigación Educativa (AERA) y el Consejo Nacional para la Medición en Educación (NCME), publicó el manual "Estándares para las pruebas educativas y psicológicas", documento que enfatiza la importancia de tomar en cuenta el bienestar de las personas que hacen una prueba y evitar el mal uso de los instrumentos de evaluación. El seguimiento de estos códigos ayuda a asegurar que los instrumentos psicométricos sean aplicados, calificados e interpretados por personas sensiblemente aptas.

Mientras que, en países como Estados Unidos, es obligatorio que estos criterios de

calidad se satisfagan, en México es inexistente esta normatividad.

Por otro lado, el mayor interés que deben tener las instituciones de educación superior (IES) al utilizar exámenes de admisión es asegurar que los lugares disponibles en la institución los ocupen quienes están más capacitados para estudiar una carrera profesional y tengan mayor probabilidad de éxito académico. Sin embargo, hasta hace muy poco tiempo, una práctica común de las IES mexicanas para seleccionar a sus estudiantes (cuando la matrícula se saturaba) era la utilización de exámenes de admisión informales, que eran elaborados por grupos de profesores que impartían clases en cada institución, pudiendo ser o no especialistas en el tema. Por desconocimiento, estos exámenes nunca se ponían en tela de juicio y, por lo tanto, no se validaban ni se estandarizaban. Una excepción a esta regla eran algunas IES privadas (y recientemente algunas públicas) que utilizaban la Prueba de Aptitud Académica (PAA); versión puertorriqueña del *Scholastic Aptitud Test (SAT)*, desarrollada por el *College Board* de los Estados Unidos de Norteamérica.

Este escenario nacional ha cambiado, recientemente, con la aparición del Centro Nacional de Evaluación de la Educación Superior (CENEVAL), el cual fue creado en 1994 con el propósito de elaborar y administrar exámenes para el ingreso a las instituciones de educación media superior y superior. Asimismo, otras instituciones han trabajado para contar con un examen de ingreso válido, como es el caso la Universidad Autónoma de Baja California (UABC), Universidad Nacional Autónoma de México (UNAM) y la Universidad Autónoma Metropolitana (UAM), por mencionar algunas. Sin embargo, aún existe un gran atraso en materia de evaluación educativa en México lo cual se evidencia por (1) la ausencia de organismos nacionales o regionales que normen sobre los criterios de calidad de las pruebas; (2) la falta de pruebas de calidad de gran escala desarrolladas, estandarizadas y validadas en nuestro país, y (3) la escasez de publicaciones de los indicadores psicométricos que poseen el reducido número de pruebas de gran escala desarrolladas en México.

Una excepción a la regla es el Examen de Habilidades y Conocimientos Básicos (EXHCOBA), el cual se desarrolló, en 1992, por un grupo de investigadores de la UABC y la UNAM, con el propósito de contar con un instrumento para seleccionar a los mejores estudiantes que desean ingresar a la UABC y, hasta donde fuera posible, a otras instituciones de educación superior del país (Backhoff y Tirado, 1992).

Desde su inicio, el EXHCOBA ha venido acumulando evidencias acerca de los diversos indicadores que soportan su calidad técnica; algunos se han publicado en diferentes revistas, tales como la descripción del contenido (Backhoff y Tirado, *op. cit.*; 1993;1994), estandarización (Larrazolo, 1997a y 1997b), validez predictiva y concurrente (Tirado, Backhoff, Larrazolo y Rosas, 1997), y datos comparativos (Backhoff y Tirado, 1993;1994; y Backhoff, Tirado, Larrazolo y Antillón, 1997). Se han publicado otros indicadores en reportes técnicos o están en proceso como proyectos de tesis de posgrado.

Sin embargo, aún no se realizan o publican todos los estudios que den cuenta de la calidad técnica del EXHCOBA, tales como la validez de contenidos, la validez de constructo, la ausencia de sesgo, la confiabilidad y el análisis de reactivos. Son tres los indicadores que se utilizan para describir las características psicométricas de los reactivos de un test objetivo: su nivel de dificultad, su poder de discriminación y el funcionamiento de sus distractores.

El propósito de este estudio es dar a conocer los resultados de los indicadores psicométricos de los reactivos del EXHCOBA, los relacionados con su nivel de dificultad y poder de discriminación. Se excluye de este trabajo el funcionamiento de sus distractores por lo extenso que representaría ese trabajo.

Para lograr nuestro objetivo, en primer lugar, describiremos lo que en la literatura especializada se entiende técnicamente por dificultad y discriminación de un ítem, así como las formas de calcularlos. Posteriormente, describiremos la metodología empleada en nuestro estudio y los resultados obtenidos, para discutir los resultados encontrados.

Análisis de reactivos

Dificultad. La dificultad de un ítem se entiende como la proporción de personas que responden correctamente un reactivo de una prueba. Entre mayor sea esta proporción, menor será su dificultad. Lo que quiere decir que se trata de una relación inversa: a mayor dificultad del ítem, menor será su índice (Wood, 1960). Para calcular la dificultad de un ítem, se divide simplemente el número de personas que contestó correctamente el ítem entre el número total de personas que contestó el ítem. Usualmente, a esta proporción se le denota con una p , e indica la dificultad del ítem (Crocker & Algina, 1986). Se calcula con la siguiente fórmula:

$$p_i = \frac{A_i}{N_i} \dots\dots\dots(1)$$

- donde: p_i = Índice de dificultad del reactivo i
 A_i = Número de aciertos en el reactivo i
 N_i = Número de aciertos más número de errores en el reactivo i

De acuerdo al manual del EXHCOBA, el nivel medio de dificultad del examen debe oscilar entre 0.5 y 0.6, distribuyéndose los valores de p de la manera siguiente: 5% de reactivos fáciles, 20% medianamente fáciles, 50% con una dificultad media, 20% medianamente difíciles y 5% difíciles.

Discriminación. Si la prueba y un ítem miden la misma habilidad o competencia,

podemos esperar que quien tuvo una puntuación alta en todo el test deberá tener altas probabilidades de contestar correctamente el ítem. También debemos esperar lo contrario, es decir, que quien tuvo bajas puntuaciones en el test, deberá tener pocas probabilidades de contestar correctamente el reactivo. Así, un buen ítem debe discriminar entre aquellos que obtuvieron buenas calificaciones en la prueba y aquellos que obtuvieron bajas calificaciones.

Usualmente, se utilizan dos formas para determinar el poder discriminativo de un ítem: el índice de discriminación y el coeficiente de discriminación. Aunque hay varias maneras equivalentes de calcular el índice de discriminación, en este trabajo utilizaremos la siguiente fórmula:

$$D_i = \frac{GA_{aciertos} - GB_{aciertos}}{N_{grupomayor}} \dots\dots\dots(2)$$

- donde:
- D_i = Índice de discriminación del reactivo i
 - $GA_{aciertos}$ = Número de aciertos en el reactivo i del 27% de personas con las puntuaciones más altas en el test.
 - $GB_{aciertos}$ = Número de aciertos en el reactivo i del 27% de personas con las puntuaciones más bajas en el test.
 - $N_{grupomayor}$ = Número de personas en el grupo más numeroso (GA o GB).

Entre más alto es el índice de discriminación, el reactivo diferenciará mejor a las personas con altas y bajas calificaciones. Si todas las personas del GA contestan correctamente un reactivo y todas las personas del GB contestan incorrectamente, entonces $D = 1$ (valor máximo de este indicador); si sucede lo contrario, $D = -1$ (valor máximo negativo); si ambos grupos contestan por igual, $D = 0$ (valor mínimo de discriminación).

Ebel y Frisbie (1986) nos dan la siguiente regla de “dedo” para determinar la calidad de los reactivos, en términos del índice de discriminación. La tabla I, muestra los valores D y su correspondiente interpretación. Asimismo, en la tabla se señalan las recomendaciones para cada uno de estos valores.

Tabla I. Poder de discriminación de los reactivos según su valor D

D =	Calidad	Recomendaciones
> 0.39	Excelente	Conservar
0.30 - 0.39	Buena	Posibilidades de mejorar
0.20 - 0.29	Regular	Necesidad de revisar

0.00 - 0.20	Pobre	Descartar o revisar a profundidad
< -0.01	Pésima	Descartar definitivamente

$$r_{pbis} = \frac{\bar{X}_1 - \bar{X}_0}{S_x} * \sqrt{\frac{n_1 n_0}{n(n-1)}}$$
 Dos indicadores más de la efectividad discriminativa de un reactivo son el punto de correlación biserial y el coeficiente de correlación biserial, a los cuales se les llama coeficientes de discriminación. La ventaja de utilizar el coeficiente de discriminación en lugar del índice de discriminación (D) es que con el primer método se toman en cuenta todas y cada una de las personas evaluadas, mientras que con el segundo, sólo se toma 54% de ellas (27% más alto y 27% más bajo).

El coeficiente de correlación biserial (r_{bis}) se calcula para determinar el grado en que las competencias que mide el test también las mide el reactivo. El r_{bis} proporciona una estimación de la correlación producto-momento de Pearson entre la calificación total de la prueba y el continuo hipotético del reactivo, cuando éste se dicotomiza en respuestas correctas e incorrectas (Henrysson, 1971).

La correlación del punto biserial (r_{pbis}) se utiliza para saber si las personas “adecuadas” son las que obtienen las respuestas correctas, qué tanto poder predictivo tiene el reactivo, y cómo puede contribuir a las predicciones. Henrysson (1971) sugiere que el r_{pbis} nos dice más sobre la validez predictiva del test que el coeficiente de correlación biserial, ya que éste tiende a favorecer los reactivos de dificultad media. También se sugiere que el r_{pbis} es una medida que combina la relación entre el criterio del reactivo y el nivel de dificultad.

Ebel y Frisbie (1986) afirman que el r_{pbis} simplemente describe la relación entre las respuestas a un reactivo (0 o 1) y las calificaciones en el test de todas las personas. La ecuación para obtener este indicador, de acuerdo a Glass y Stanley (1986), es la siguiente.

.....(3)

- donde: \bar{X}_1 = Media de las puntuaciones totales de aquellos que respondieron correctamente el ítem.
 \bar{X}_0 = Media de las puntuaciones totales de aquellos que respondieron incorrectamente el ítem.
 S_x = Desviación estándar de las puntuaciones totales.
 n_1 = Número de casos que respondieron correctamente el ítem.
 n_0 = Número de casos que respondieron incorrectamente el ítem.
 n = $n_1 + n_0$

Método

Instrumentos. El EXHCOBA es un examen normativo de opción múltiple, diseñado para seleccionar a los mejores aspirantes que desean ingresar al nivel superior (aunque también se puede utilizar para el nivel medio superior haciendo ciertos ajustes). Algunas de sus características importantes para el diseño de este estudio, son las siguientes:

- El examen está conformado por tres secciones: habilidades básicas (con dos áreas temáticas y 60 preguntas), conocimientos básicos (con cuatro áreas temáticas y 70 preguntas) y conocimientos básicos de especialidad (con nueve áreas temáticas y 180 preguntas). La estructura del examen se presenta en la tabla II. La primera corresponde al nivel de primaria, la segunda, al nivel de secundaria y la tercera, al nivel de bachillerato.
- Todos los estudiantes contestan las primeras dos secciones del examen, que suman un total de 130 preguntas, y sólo contestan 60 preguntas de la tercera sección; es decir, tres áreas temáticas que se seleccionan de acuerdo a la carrera a la que se desee ingresar. Con las áreas temáticas de esta tercera sección del examen, se configuran siete distintos tipos de exámenes que corresponden a las áreas de conocimiento que utiliza la ANUIES. La tabla III muestra la configuración de éstos para las carreras que se ofrecen en la UABC.
- Adicionalmente, se cuenta con cinco versiones de preguntas clon para cada uno de los 310 ítems del examen. En este estudio se utilizó solamente una de ellas (la versión dos).

Tabla II. Estructura del ESCOBA

Niveles	Secciones	Áreas temáticas	k
Primaria	Habilidades básicas	Habilidades verbales	30
		Habilidades cuantitativas	30
Secundaria	Conocimientos básicos	Lengua española	15
		Matemáticas	15
		Cs. naturales	20
		Cs. sociales	20

Niveles	Secciones	Áreas temáticas	k
Bachillerato	Conocimientos básicos para especialidad	Estadística	20
		Cs. sociales	20
		Cs. económico-administrativa	20
		Cálculo	20
		Biología	20
		Química	20
		Física	20
		Lenguaje	20
		Humanidades	20

Tabla III. Áreas temáticas de la sección de Conocimientos Básicos de Especialidad que se evalúan según el tipo de examen

Tipo de examen	Áreas temáticas (bachillerato)
Económica-Administrativa	Matemáticas-Estadística Ciencias Sociales Económico-administrativa
Químico-Biológicas	Matemáticas-Cálculo Biología Química
Salud	Física Biología Química
Ingeniería	Matemáticas-Cálculo Física Química
Físico-Matemáticas	Matemáticas-Cálculo Física Lenguaje
Humanidades	Humanidades Ciencias Sociales Lenguaje
Ciencias Sociales	Matemáticas-Estadística Ciencias Sociales Lenguaje

Este examen se puede administrar de dos formas: en lápiz y papel, y en computadora. En este estudio, se utilizó el formato computarizado. Este formato implica que al

estudiante se le presentan las preguntas y las opciones de respuesta en la pantalla de una terminal de computadora (una pregunta a la vez). Independientemente del conocimiento y práctica que se tenga sobre el uso de computadoras, cualquier estudiante puede responder el examen sin que esta condición merme los resultados. Una descripción más detallada de la interfaz, y de su equivalencia con el formato de lápiz y papel, se encuentra publicada en varios trabajos (véase, por ejemplo, Backhoff, Ibarra y Rosas, 1995).

Población. Participaron en este estudio 876 aspirantes que deseaban ingresar a la UABC, en el verano de 1998. Estos representan la décima parte del total de estudiantes que realizaron el EXHCOBA para ser seleccionados a las distintas carreras universitarias que ofrece la institución. La población estudiada se seleccionó aleatoriamente entre aquellos estudiantes cuya ficha de examen terminó en el dígito dos.

Procedimiento. Los pasos que se siguieron para administrar y calificar el examen fueron los mismos que define la universidad en su proceso de admisión: (1) el estudiante pagó y obtuvo una ficha, en la cual se le indicó la fecha, lugar y hora de su examen; (2) según la carrera a la que aspiró ingresar, el sistema automatizado le presentó uno de los siete tipos de exámenes, correspondientes a las áreas de conocimiento en que se agrupan las carreras de la UABC; (3) el estudiante respondió el examen sin ningún tipo de ayuda (calculadoras, diccionarios, etc.), salvo lápiz y papel; y (4) el sistema calificó las respuestas en forma automática, dándole al estudiante un reporte por escrito de su ejecución. Durante este proceso, se encontró siempre presente una persona capacitada que resolvió cualquier problema o duda sobre el manejo de la parte computarizada del examen.

Conforme los estudiantes fueron terminando de contestar el examen, los resultados se capturaron en una base de datos. Esta se preparó posteriormente para realizar los análisis estadísticos correspondientes. Básicamente, la preparación consistió en transformar los resultados crudos de los estudiantes en un formato binario (0 y 1), así como depurar los casos con patrones de respuesta anormales (ej. registros vacíos). Utilizando el *software* de hoja de cálculo Excel y el paquete estadístico SPSS, se calcularon los valores p (dificultad), D (índice de discriminación) y r_{pbis} para todos los reactivos del examen.

El índice de dificultad se calculó con la ecuación (1), el índice de discriminación con la ecuación (2) y el coeficiente de discriminación con la ecuación (3).

Resultados

La tabla IV presenta los promedios de las medias y desviaciones estándar de los valores p , D y r_{pbis} de los reactivos de las 15 áreas temáticas del examen. Aquí, podemos observar que el promedio general de dificultad de los reactivos del EXHCOBA fue de 0.56, con una desviación estándar de 0.18, para los 876 estudiantes de este estudio.

Analizando los niveles de dificultad de acuerdo a sus temáticas generales y a su ubicación en los tres niveles escolares, podemos señalar que los valores promedio de p , para la sección de habilidades básicas (primaria) fue de 0.65; para la sección de conocimientos básicos (secundaria) fue de 0.57, y para la sección de conocimientos básicos de especialidad (bachillerato), fue de 0.54. Estos valores, en principio, concuerdan con los niveles de escolaridad que representan.

Tabla IV. Medias y desviaciones estándar del índice de dificultad (p), índice de discriminación (D) y coeficiente de discriminación (r_{pbis}) por área temática del ESCOBA

Secciones	Área temática	N=	p		D		r_{pbis}	
			Media	D.E.	Media	D.E.	Media	D.E.
Habilidades básicas	Hab. verbales	876	0.65	0.15	0.36	0.11	0.26	0.10
	Hab. cuantitativas	876	0.65	0.14	0.43	0.10	0.34	0.10
Conocimientos básicos	Español	872	0.61	0.18	0.35	0.09	0.25	0.10
	Matemáticas	872	0.47	0.17	0.44	0.14	0.39	0.12
	Cs. naturales	872	0.62	0.19	0.36	0.12	0.27	0.15
	Cs. sociales	872	0.58	0.21	0.35	0.14	0.28	0.12
Conocimientos básicos de especialidad	Estadística	362	0.50	0.13	0.39	0.13	0.31	0.11
	Cs. sociales	475	0.59	0.20	0.18	0.10	0.16	0.09
	Eco-administrativas	249	0.56	0.13	0.30	0.15	0.25	0.11
	Cálculo	170	0.52	0.17	0.35	0.22	0.34	0.17
	Biología	146	0.51	0.22	0.35	0.16	0.32	0.14
	Química	302	0.45	0.22	0.29	0.15	0.25	0.13
	Física	288	0.50	0.13	0.31	0.17	0.21	0.17
	Lenguaje	113	0.62	0.20	0.30	0.13	0.27	0.13
	Humanidades	226	0.64	0.22	0.31	0.16	0.25	0.12
Total / promedio		876	0.56	0.18	0.34	0.14	0.28	0.12

Haciendo una comparación de las 15 áreas temáticas del examen, podemos observar que su dificultad oscila entre 0.45 y 0.65, y las más fáciles son las relacionadas con las habilidades verbales y cuantitativas, así como las humanidades, y las más difíciles las relacionadas con las ciencias naturales del nivel bachillerato (química, física y biología)

así como las tres de matemáticas (matemáticas, cálculo y estadística). Sin embargo, hay que aclarar que no todos los estudiantes respondieron las mismas áreas temáticas de la sección de conocimientos básicos de especialidad (ver tabla III), por lo que esta comparación no es del todo exacta, aunque sí muy aproximada.

Ahora bien, para hacer un análisis del comportamiento de la dificultad de los 310 reactivos, se presenta la figura 1, donde se muestra la distribución de las frecuencias de los valores p . Aquí se podrá observar que: (1) un gran número de reactivos se agrupan en la frecuencia con una media de 0.65, (2) la distribución de estos valores tiene un sesgo negativo y (3) hay ligeramente mayor número de reactivos difíciles que fáciles.

Clasificando los reactivos según su nivel de dificultad, podríamos agruparlos de la siguiente manera: 11.6% altamente difíciles ($p < 0.32$); 24.2% medianamente difíciles (de 0.33 a 0.52); 45.2% de dificultad media (de 0.53 a 0.73); 14.2 % medianamente fáciles (de 0.74 a 0.86); y 4.8% altamente fáciles ($p > 0.86$).

Por otro lado, en cuanto al poder discriminativo (D), podemos observar en la tabla IV que el índice de discriminación promedio de todo el examen es de 0.34, mientras que el coeficiente de discriminación promedio es de 0.28.

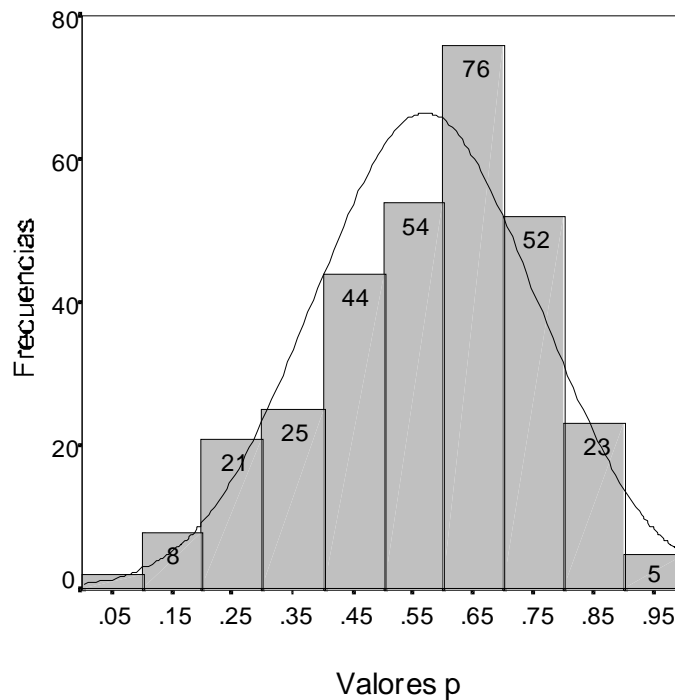


Figura 1. Distribución de la dificultad de los reactivos del EXHCOBA

Analizando los índices de discriminación (**D**) de acuerdo a sus temáticas generales, podemos señalar que los valores promedio de **D**, para la sección de habilidades básicas (primaria), fue de 0.40; para la sección de conocimientos básicos (secundaria), fue de 0.38, y para la sección de conocimientos básicos de especialidad (bachillerato), fue de 0.31. De la misma manera, encontramos que los valores de los coeficientes de discriminación (r_{pbis}) para las secciones de habilidades y conocimientos básicos, fue de 0.30, mientras que, para la sección de conocimientos básicos de especialidad, fue de 0.26.

Es interesante hacer notar que, siendo esta prueba para el nivel superior, los reactivos que mejor discriminan son los que se relacionan con las habilidades que se adquieren en la primaria, seguidos por los conocimientos básicos que se adquieren en la secundaria y, por último, los conocimientos que se adquieren en el bachillerato; argumento que fortalece la idea central del EXHCOBA de evaluar competencias básicas que se adquieren durante toda la vida escolar.

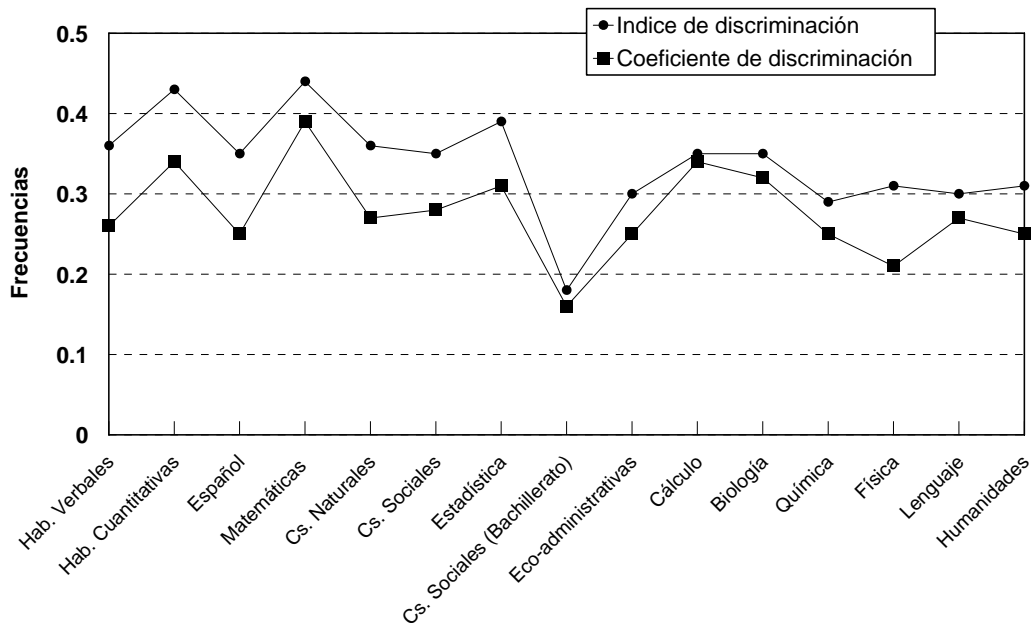


Figura 2. Índices y coeficientes de discriminación de las 15 áreas temáticas del EXHCOBA

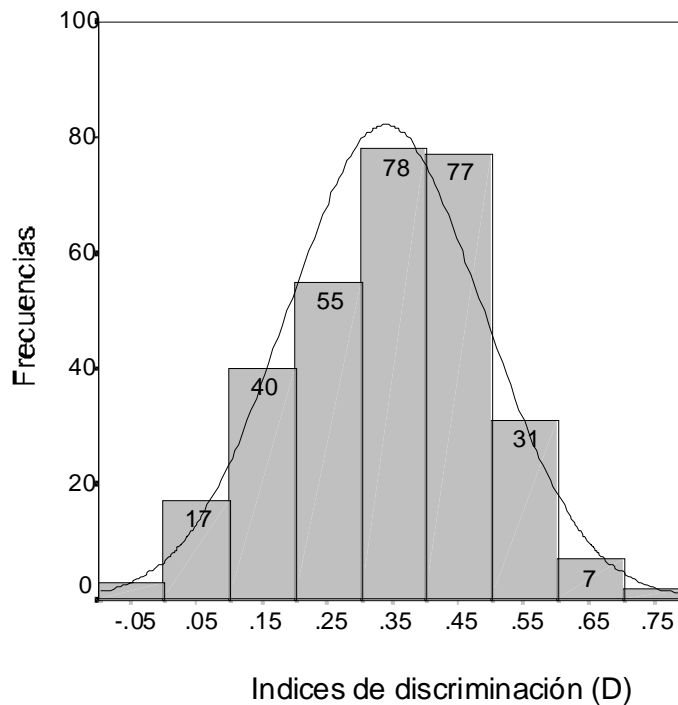
Analizando los índices de discriminación de las 15 áreas temáticas del examen, apreciamos que éstos oscilan entre 0.18 y 0.44, mientras que los coeficientes de discriminación oscilan entre 0.16 y 0.39. En la figura 2 se aprecia el comportamiento discriminativo de las áreas temáticas del EXHCOBA con estos dos indicadores. Como

se podrá observar, existe un paralelismo entre ambas medidas de discriminación, y son en promedio más altos por 0.5 puntos los valores D (índices de discriminación) que los valores r_{pbis} (coeficientes de discriminación).

Utilizando los valores D, podemos agrupar las áreas temáticas de acuerdo a su poder discriminativo en tres grupos: dos áreas con ítems que discriminan en forma excelente con valores mayores a 0.40 (matemáticas y habilidades cuantitativas); un área cuyos reactivos discriminan deficientemente con valores menores a 0.20 (ciencias sociales, del nivel bachillerato) y 12 áreas que tienen buen poder discriminativo, con valores entre 0.29 y 0.39 (áreas restantes). De la misma manera, con los valores r_{pbis} podemos agrupar las áreas temáticas de la siguiente manera: cuatro áreas cuyos ítems discriminan en forma excelente con valores mayores a 0.30 (habilidades cuantitativas, matemáticas, cálculo y biología); un área con un débil poder discriminativo con un valor menor a 0.20 (ciencias sociales, de bachillerato) y diez áreas con un buen poder discriminativo (áreas restantes).

Es interesante notar que el área de conocimientos con mejor nivel discriminativo pertenece a las matemáticas y el área con menor poder discriminativo, a las ciencias sociales.

Para observar el comportamiento de los reactivos en cuanto a su poder discriminativo, se presentan las figuras 3 y 4, que muestran la distribución de las frecuencias de los valores D y r_{pbis} , respectivamente. Aquí se puede observar que la mayor parte de los valores D se agrupan en el rango de 0.30 a 0.50, y que hay tres reactivos que discriminan negativamente. Por su parte, la mayoría de los valores r_{pbis} se agrupan en los intervalos de 0.20 a 0.40, y se identifican siete reactivos con discriminación negativa.



Según la información presentada en las figuras 2, 3 y 4, podemos clasificar los reactivos del EXHCOBA de acuerdo a sus:

- Niveles de dificultad que permite agruparlos de la siguiente manera: 31 (10%) son muy difíciles ($p < 0.3$); 69 (22.2%) son medianamente difíciles ($p = 0.31 - 0.50$); 130 (42%) son de dificultad media ($p = 0.51 - 0.70$); 75 (24.2%) son medianamente fáciles ($p = 0.71 - 0.90$); y 5 (1.6%) son muy fáciles ($p > 0.80$).
- Indices de discriminación permite agruparlos de la siguiente manera: 3 (1%) discriminan negativamente ($D < 0$); 57 (18.4%) discriminan pobremente ($D = 0 - 0.19$); 55 (17.7%) poseen un poder discriminativo regular ($D = 0.30 - 0.39$); 78 (25.2%) discriminan bien; y 117 (37.7%) poseen un excelente nivel discriminativo ($D > 0.39$).

- Coeficientes de discriminación permite agruparlos de la siguiente manera: 7 (2.3%) discriminan negativamente ($r_{pbis} < 0$); 45 (14.5%) discriminan pobremente ($r_{pbis} = 0 - 0.14$); 72 (23.2%) poseen un poder discriminativo regular ($r_{pbis} = 0.15 - 0.25$); 88 (28.4%) poseen un buen poder discriminativo ($r_{pbis} = 0.26 - 0.35$); y 98 (31.6%) poseen un excelente nivel discriminativo ($r_{pbis} > 0.35$).

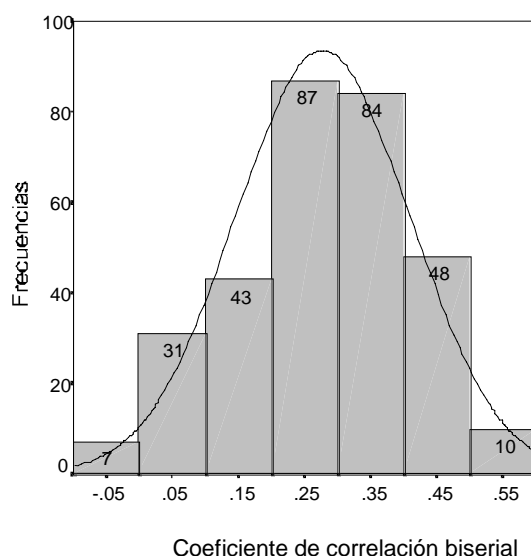


Figura 4. Distribución de los coeficientes de discriminación del EXHCOBA

Finalmente, para identificar los 310 reactivos del EXHCOBA que necesitaban alguna modificación, utilizamos el siguiente criterio óptimo de calidad para aceptar o rechazar reactivos: aquellos con valores p menores de 0.20 y mayores de 0.90; con valores D menores de 0.20; y, con valores r_{pbis} menores de 0.15. Utilizando estos parámetros, pudimos observar que los reactivos que no cumplían las mínimas normas de calidad se distribuyen en casi todas las áreas temáticas, pero especialmente fueron más frecuentes en el área de ciencias sociales (nivel de bachillerato) y, prácticamente, fueron inexistentes en las áreas de habilidades cuantitativas y matemáticas.

Discusión

Una forma fundamental de conocer la calidad de los reactivos de un test es poniéndolos a prueba y analizando su comportamiento empíricamente. Dos indicadores fundamentales para realizar este análisis son el nivel de dificultad y el poder discriminativo.

En cuanto al primero de ellos, podemos decir que el promedio de dificultad de los reactivos del EXHCOBA (0.56) es ligeramente más difícil que el nivel de dificultad esperado para un examen normativo con cuatro opciones de respuesta (0.62). Igualmente, la distribución de la dificultad de sus reactivos se acerca mucho a lo deseado, cubriendo prácticamente toda la gama de valores de p y concentrando 42% de sus reactivos con un valor medio de dificultad. Sin embargo, se nota una proporción ligeramente mayor de reactivos difíciles y medianamente difíciles que fáciles y medianamente fáciles (5% aproximadamente). Por último, es importante señalar que solamente 12% de los reactivos presentó niveles inapropiados de dificultad (muy fáciles o muy difíciles); valor muy cercano al 10% esperado teóricamente.

En cuanto al índice de discriminación del examen, se puede decir que 81% de sus ítems presentaron niveles de aceptables a óptimos; 18% de ellos tuvo un nivel pobre y 1% de ellos discriminó negativamente. En cuanto a los coeficientes de discriminación, 83% de sus reactivos presentaron niveles de aceptables a óptimos; 14.5 % de ellos tuvo un nivel pobre y 2.5% discriminó negativamente. Como se puede apreciar, con ambos indicadores de poder discriminativo se obtienen resultados equivalentes, lo cual se muestra claramente en la figura 2.

Definitivamente, los ítems con mayor poder discriminativo fueron los relacionados con las matemáticas; inversamente, los que discriminaron más deficientemente fueron los relacionados con las ciencias sociales. Lo anterior se puede explicar si consideramos que las matemáticas evalúan competencias que tienen una estrecha relación lógica e inclusiva; es decir, sus conceptos fundamentales se enlazan y se construyen unos sobre otros, en forma progresiva, de tal manera que no es posible entender algún concepto o resolver algún problema sin entender los anteriores. Por ejemplo, para poder hacer divisiones es necesario dominar la suma, la resta y la multiplicación. En sentido opuesto, las disciplinas sociales se construyen sin esta estrecha relación entre sus conceptos básicos.

Finalmente, es interesante señalar que las áreas temáticas con mejor poder discriminativo se concentran en las dos primeras secciones del examen y que evalúan competencias que se adquieren en la primaria y la secundaria. Esto también tiene una explicación lógica, en el sentido de que estas dos secciones evalúan conocimientos muy básicos, necesarios para entender otros más complejos y que estas dos secciones las contestan la totalidad de los estudiantes. Este último aspecto hace que haya una mayor diversidad en las competencias del estudiantado, por lo cual serán más fácil de discriminar. Lo contrario es cierto para las áreas temáticas de especialidad.

Por todo lo anterior, es importante señalar que estos resultados, además de servir para mejorar la calidad de los reactivos, fortalecen una de las ideas centrales del EXHCOBA: evaluar habilidades y conocimientos básicos que se adquieren durante

todo el recorrido escolar, ya que con ellos se podrá discriminar y, por lo tanto, seleccionar a los mejores estudiantes que desean ingresar a la universidad.

Referencias

Aiken, L.R. (1996). *Tests psicológicos y evaluación*. México: Prentice Hall Hispanoamericana.

Backhoff, E., Ibarra, M.A. y Rosas, M. (1995). Sistema computarizado de exámenes (SICODEX). *Revista Mexicana de Psicología*, 10 (1), 55-62.

Backhoff, E. y Tirado, F. (1994). Estructura y lógica del Examen de Habilidades y Conocimientos Básicos. *Revista Sonorense de Psicología*, 8 (1), 21-33.

Backhoff, E. y Tirado, F. (1993). Habilidades y conocimientos básicos del estudiante

universitario: hacia los estándares nacionales. *Revista de la Educación Superior*, 88, 45-65.

Backhoff, E. y Tirado, F. (1992). Desarrollo del Examen de Habilidades y Conocimientos Básicos. *Revista de la Educación Superior*, 83, 95-117.

Backhoff, E., Tirado, F. Larrazolo, N. y Antillón, L. (1997). Desigualdad de la educación básica en México. *Revista Latinoamericana de Estudios Educativos*, 26 (3), 73-87.

CENEVAL (s/f). *Acerca del CENEVAL y los exámenes generales para el egreso de la licenciatura*. México: autor.

Crocker, L. y Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Holt, Rinehart and Winston.

Ebel, R.L. y Frisbie, D.A. (1986). *Essentials of Education Measurement*. Englewood Cliffs, NJ: Prentice Hall.

Glass, G. y Stanley, J. (1986). *Métodos estadísticos aplicados a las ciencias sociales*. (Traducción: E. Galvis y E. Guzmán). México: Prentice-Hall Hispanoamericana.

Henrysson, S. (1971). Gathering, Analysing, and Using Data on Test Items. En R.L. Thorndike (Ed.), *Educational Measurement*. Washington, DC: American Council on Education.

Larrazolo, N. (1997a). Escala de calificaciones del Examen de Habilidades y Conocimientos Básicos. Documento mimeografiado. Ensenada, B.C.: U.A.B.C.

Larrazolo, N. (1997b). Guía Oficial de Estudio. Documento mimeografiado. Ensenada, B.C.: Universidad Autónoma de Baja California.

Tirado, F., Backhoff, E., Larrazolo, N. y Rosas, M, (1997). Validez predictiva del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Mexicana de Investigación Educativa*. 11, (3), 67-84.

Wood, D.A. (1960). *Test Construction: Development and Interpretation of Achievement Tests*. Columbus, OH: Charles E. Merrill Books, Inc.

Direcciones electrónicas donde se puede obtener información relacionada con la temática del artículo:

Backhoff Escudero, *et al.*: Nivel de dificultad y poder de...

Evaluation Assistance Center – Western Region (EAC West). 1996. Testing and Test Scores: Background and Definitions. en:

<http://www.ncbe.gwu.edu/mispubs/eacwest/handouts/testing/backgrnd.htm>

(Consultado en enero 22 de 1999)

Kitao, S.K. & Kitao, K. (1997). Writing a Good Test.

<http://www.ling.lancs.ac.uk/staff/visitors/kenji/kitao/design2.htm>

(Consultado en enero 20 de 1999)