

Antillón, L. E., Larrazolo, N. y Backhoff, E. (2006). Igualación equipercen-til del examen de habilidades y conocimientos básicos (EXHCOBA). *RELIEVE*, v. 12, n. 2, p. 205-217. http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2_2.htm



**Revista ELectrónica de Investigación
y EValuación Educativa**

IGUALACIÓN EQUIPERCENTIL DEL EXAMEN DE HABILIDADES Y CONOCIMIENTOS BÁSICOS (EXH- COBA)

*[Equipercen-til equating of the Basic Ability and Knowledge Test
(EXHCOBA)]*

por

[Article record](#)

[About authors](#)

[HTML format](#)

Luz Elena Antillón

Norma Larrazolo

Eduardo Backhoff (backhoff@inee.edu.mx)

[Ficha del artículo](#)

[Sobre los autores](#)

[Formato HTML](#)

Abstract

Equipercen-til equating method is a statistical procedure where student raw scores of two different versions of the same test are considered equated if they correspond to the same percentile range. A graphic curve is presented to describe the difficult differences from version to version of a test. This work was aimed to estimate the equipercen-til equating values, by academic content area, of the Basic Ability and Knowledge Test (EXHCOBA, by its Spanish acronym) that is used by the University of Baja California (UABC) as a selection student test. This norm-referenced test has excellent quality standards, a high technological development, several reliability and validity support studies, and others good psychometric parameters. Estimation of equating parameters was done applying the analytic method described by Kolen and Brennan (1995), with the random group procedure utilized by UABC to collect data. Results shows that equating was effective to adjust four statistical moments (mean, standard deviation, bias, and kurtosis) of the frequency distributions of EXHCOBA's version 3 and 4 compared with version 2, by content area, producing equal score distributions. Nevertheless, irregularities appeared at the ends of the curves that suggests the need of a smoothing procedure.

Keywords

Test equating, equipercen-til equating method, Basic Abilities and Knowledge Test.

Resumen

La igualación equipercen-til es un método estadístico en el cual los puntajes crudos de dos versiones de una prueba se consideran igualados si ellos corresponden al mismo rango percentilar en un grupo de examinados. En la igualación equipercen-til se presenta una curva para describir las diferencias de dificultad de versión a versión. Este trabajo tuvo como objetivo estimar la igualación equipercen-til sin suavizado de las versiones 3 y 4, con la versión 2, por área temática del Examen de Habilidades y Conocimientos Básicos (EXHCOBA) que utiliza la Universidad Autónoma de Baja California (UABC) para la selección de aspirantes, examen que posee un nivel de calidad excelente, un desarrollo tecnológico elevado y cuenta con estudios constantes de fiabilidad, validez y otros indicadores psicométricos. La estimación de la igualación equipercen-til se llevó a cabo con el método analítico que describen Kolen y Brennan (1995) y con el diseño de grupos aleatorios que utiliza la UABC para la recolección de datos. Los resultados de este estudio indicaron que la igualación resultó efectiva al ajustar cuatro momentos (media, desviación estándar, sesgo y curtosis) de las distribuciones de las versiones 3 y 4 con la versión 2 por área temática, produciendo distribuciones iguales en las puntuaciones. Sin embargo, algunas irregularidades se encontraron en los extremos de las curvas que hacen suponer la falta de un proceso de suavizado.

Descriptores

Igualación de pruebas, método equipercen-til, Examen de Habilidades y Conocimientos Básicos.

INTRODUCCIÓN

El objetivo fundamental de toda evaluación del aprendizaje de gran escala es proporcionar un conocimiento general del rendimiento académico de los estudiantes, ya sea con propósitos de selección, certificación, diagnóstico o ubicación. Estas evaluaciones, por lo general, se realizan a través de pruebas estandarizadas diseñadas expresamente para los propósitos que se persiguen.

Las puntuaciones de estas pruebas se utilizan como piezas de información para tomar decisiones que pueden impactar a un individuo, a una institución o, incluso, a un sistema educativo. Este es el caso de las pruebas de admisión que se utilizan en las instituciones de educación con el objetivo de seleccionar a los mejores estudiantes.

Las puntuaciones de estas pruebas se utilizan como piezas de información para tomar decisiones que pueden impactar a un individuo, a una institución o, incluso, a un sistema educativo. Este es el caso de las pruebas de admisión que se utilizan en las instituciones de educación con el objetivo de seleccionar a los mejores estudiantes.

Cuando un examen se administra regularmente en repetidas ocasiones, como es el caso de los exámenes de admisión, es necesario contar con diversas versiones de la prueba, con el fin de garantizar la seguridad e imparcialidad de sus resultados. Sin embargo, cuando se aplican formas diferentes de una misma prueba a una población de estudiantes, las decisiones que se tomen con base en los resultados pueden llegar a ser injustas. De no implementarse una estrategia de ajuste, algunos estudiantes se verán afectados; es decir, los estudiantes examinados mediante una versión más difícil tendrán desventajas sobre aquellos examinados con una versión más fácil, debido a la variación en la dificultad de los reactivos de cada versión (Kolen y Brennan, 1995).

Por esta razón, las versiones de los exámenes se construyen tratando de que sean equivalentes, o paralelas, para lo cual se debe garantizar que evalúen los mismos contenidos y que tengan la misma estructura, formato, extensión y tipo de reactivos. Sin embargo, a pesar de los cuidados que se tengan al diseñar y construir las distintas versiones de una prueba, es inevitable que éstas varíen en cuanto a su dificultad y otras propiedades psicométricas (distribución, varianza, sesgo, etc.). Dos versiones de una prueba rara vez se pueden construir con una equivalencia exacta o precisa, lo que hace necesaria un proceso de igualación de las pruebas (Holland y Rubin, 1982).

La igualación de pruebas consiste en hacer ajustes numéricos de los puntajes obtenidos en cada versión para compensar las variaciones en dificultad de versión a versión. En otras palabras, la igualación convierte el sistema de unidades de una versión al sistema de unidades de otra versión; así los puntajes derivados de las dos versiones después de la conversión serán equivalentes (Holland y Rubin, *op. cit.*).

Con el enfoque de la Teoría Clásica de Test se tienen diferentes métodos de igualación y diseños de recolección de datos que permiten establecer puntajes equivalentes entre versiones de un mismo examen. Los métodos de igualación se pueden clasificar en dos tipos: lineal y equipercen-til. El procedimiento para el ajuste de la función lineal se lleva a cabo a partir de la estimación de las medias aritméticas y las desviaciones estándar de las distribuciones de calificaciones de las dos versiones del examen. Por su parte, el proceso de ajuste de las funciones de igualación equipercen-til se lleva a cabo mediante la construcción de las funciones de rangos percentilares para cada versión de la prueba, utilizando la definición de igualación equipercen-til (Arce-Ferrer y Backhoff, 2002). Así, en el método de igualación equipercen-til

los puntajes crudos de dos versiones de una prueba se consideran igualados si ellos corresponden al mismo rango percentilar en un grupo de examinados (Linn, 1993).

Ambos tipos de métodos (lineal y equipercantil) se utilizan para igualar los resultados de los exámenes de gran escala. Por ejemplo, el *American College Testing* (ACT) utiliza diferentes versiones de pruebas que son igualadas con el método equipercantil; los exámenes del *Scholastic Aptitude Test* (SAT) son igualados con los métodos de igualación equipercantil y lineal, así como con los de la Teoría de Respuesta al Ítem (Walberg y Haertel, 1992); y los EXANI-I y EXANI-II (del CENEVAL) son igualados con el método de equivalencia lineal (Tristán, 1996).

En un estudio donde se comparan ambos métodos, Arce-Ferrer y Backhoff (2002) demostraron que: 1) el método de igualación lineal ignora la variación de la dificultad de un examen a lo largo de la escala de calificaciones, cosa que no ocurre con la igualación equipercantil, 2) la forma de la relación entre las calificaciones crudas observadas y las igualadas es de naturaleza curvilínea, 3) el análisis de la cantidad de error de igualación muestra regiones en la escala en donde las puntuaciones igualadas con el método lineal difirieren de aquéllas obtenidas con el método equipercantil y 4) las discrepancias entre las calificaciones igualadas con los dos métodos resultan mayores en los extremos de la escala de calificaciones.

Por otro lado, la Universidad Autónoma de Baja California (UABC) desarrolló el Examen de Habilidades y Conocimientos Básicos (EXHCOBA), el cual se utiliza desde 1994 con propósitos de selección tanto en la UABC como en otras instituciones públicas de educación superior mexicanas (Backhoff, 2001), entre las que se encuentran: la Universidad de Guanajuato (UG), la Universidad de Sonora (UNISON), la Universidad Autónoma de Nayarit (UAN) y la Universidad Autónoma de Querétaro (UAQ). La impor-

tancia de este examen se puede medir por el número de estudiantes que lo responden anualmente, que es del orden de 50 000.

A pesar de que los estándares para la medición psicológica y educativa recomiendan realizar estudios de igualación de las distribuciones de calificaciones de las versiones de un examen (AERA, APA, NCME, 1999; Martínez-Rizo *et al.*, 2000), hasta la fecha el EXHCOBA no ha definido un método de igualación para ajustar sus puntuaciones; en gran medida, debido a que entre las organizaciones que desarrollan exámenes, no hay un acuerdo claro con respecto a los métodos que se deben de utilizar.

Considerando lo anterior, se decidió probar el método equipercantil para igualar tres versiones del examen, con el fin de conocer sus equivalencias estadísticas, así como su efectividad; metas que constituyen el propósito del presente trabajo.

Este estudio centra su atención en un tema muy poco analizado y documentado por los diversos organismos iberoamericanos que desarrollan y aplican exámenes de gran escala. Estudios de esta naturaleza no sólo contribuirán a profesionalizar el campo de la medición y evaluación educativa; sino también, pondrán en marcha el proceso de rendición de cuentas basado en evidencias que demanda la sociedad actual.

MÉTODO

Sujetos

Como parte regular del proceso de selección de los estudiantes que ingresan a la UABC, en 1998 se administró el EXHCOBA a una población total de 8,925 aspirantes que fueron examinados con la versión computarizada de este examen. Del total de estudiantes, aproximadamente, 10% respondió a una de las tres versiones del examen. Así, a 883 estudiantes se les administró la versión 2, la versión 3 la respondieron 887 alumnos y, finalmente, a 888 jóvenes se les aplicó la

versión 4. El resto de los estudiantes (70%) respondieron a versiones mezcladas aleatoriamente, razón por la que no participaron en este estudio.

Instrumento

El EXHCOBA es una prueba de aprovechamiento escolar referida a una norma, diseñada expresamente para seleccionar aspirantes a la instituciones de educación superior mexicanas. Este examen evalúa tanto las habilidades como los conocimientos básicos que se adquieren en la escuela, desde los últimos grados de primaria hasta los últimos del bachillerato (Backhoff y Tirado, 1993).

El EXHCOBA está estructurado en tres niveles: 1) Habilidades Básicas del nivel de primaria, 2) Conocimientos Básicos del nivel de secundaria, y 3) Conocimientos Básicos de Especialidad de bachillerato. Todos los estudiantes responden las preguntas de los dos primeros niveles, mientras que solo responden a tres áreas de contenido del tercer nivel, dependiendo de la carrera que elija (Backhoff y Tirado, 1992). En total, los estudiantes responden 190 reactivos (de 310) de opción múltiple: 60 del primer nivel, 70 del segundo y 60 del tercero. La Tabla I muestra la estructura del examen.

Tabla I. Estructura del EXHCOBA

Áreas de contenido y nivel educativo	Áreas temáticas	Número de Reactivos	Proporción en el examen	
Habilidades (primaria)	Básicas	Habilidades Verbales	30	0.16
		Habilidades Cuantitativas	30	0.16
Conocimientos (secundaria)	Básicos	Lengua Española	15	0.08
		Matemáticas	15	0.08
		Ciencias Naturales	20	0.10
		Ciencias Sociales	20	0.10
		Lenguaje	20	
Conocimientos de Especialidad (bachillerato)	Básicos	Matemáticas para el Cálculo	20	
		Matemáticas para la Estadística	20	
		Física	20	0.32*
		Química	20	
		Biología	20	
		Ciencias Sociales	20	
		Humanidades	20	
		Económico-Administrativa	20	

* Calculado con base en 60 reactivos que cada alumno responde de esta área

Este examen cuenta con una interfaz computarizada la cual permite que el alumno visualice los reactivos y los responda en pantalla. Tanto los reactivos como sus cuatro opciones de respuesta se presentan en forma aleatoria, de acuerdo a las áreas de contenido (Backhoff, Ibarra & Rosas, 1995). La interfaz le permite al alumno regresar a las preguntas y corregir las respuestas tantas veces como lo desee.

El EXHCOBA cuenta con diversas versiones las cuales son idénticas en cuanto a su estructura, contenido, extensión y tipo de reactivos, por lo que se puede considerar que

son versiones paralelas. En esta investigación se utilizaron las versiones 2, 3 y 4 (la primera versión se había liberado para utilizarla como guía de estudio).

Procedimiento

Se elaboraron tres bases de datos, en las cuales las respuestas fueron transformadas a un código binario (respuestas correctas e incorrectas). El proceso de igualación equipercantil se basó en el procedimiento descrito por Kolen y Brennan (1995). Para realizar los cálculos respectivos se requirió desarro-

llar un programa de cómputo (en lenguaje CLIPPER, versión 5.2).

El procedimiento de igualación equipercen-
 tilar requirió de la estimación de las funciones
 de los rangos percentilares de cada una de las
 15 áreas temáticas de cada versión. La ecua-
 ción (1.1) se utilizó para la obtención de los
 rangos percentilares en puntajes discretos.

$$P(x) = 100[F(x-1) + f(x)/2] \dots\dots\dots(1.1)$$

Ahora bien, se utilizaron tres ecuaciones para
 estimar la función de la equivalencia equipercen-
 tilar $e_Y(x)$ de los puntajes x en las 15 áreas temá-
 ticas de la versión 3 y 4 (x) a la versión 2 (y), o
 lo que es igual la función inversa del rango per-
 centilar para la versión 2, que se simboliza co-
 mo $Q^{-1}[P(x)]$, o lo que es igual a estimar
 $y_U[P(x)]$ o $y_L[P(x)]$. Estas últimas hacen refe-
 rencia a obtener el rango percentilar de puntajes
 x , simbolizado como $P(x)$, y la proporción acu-
 mulada de puntajes y simbolizada como $G(y_U^*)$
 y $G(y_L^*)$ (donde y_U^* es el puntaje entero con un
 porcentaje acumulado $[100G(y)]$ que es más
 grande que el rango percentil $P(x)$ próximo supe-
 rior, y donde y_L^* es el puntaje entero con un por-
 centaje acumulado $[100G(y)]$ que es más peque-
 ño que el rango percentil $P(x)$ próximo anterior).

Si en la proporción de aciertos de puntajes y ,
 simbolizado como $g(y)$ no existieran ceros en
 todos los puntajes $0,1,\dots K_y$, (donde K_y es el
 número de ítem en la versión y) entonces cual-
 quiera de las dos expresiones siguientes (1.2 y
 1.3) podría utilizarse para la estimación de la fun-
 ción de la igualación equipercen-
 tilar. Sin embar-
 go, en este estudio la proporción de aciertos $g(y)$
 presentó ceros, por lo que $y_U \neq y_L$. Por tal moti-
 vo, se requirió de las dos ecuaciones anteriores y
 una tercera ecuación alterna, la ecuación (1.4)
 que es el promedio de y_U y y_L .

La ecuación (1.2) es la siguiente:

$$e_Y(x) = Q^{-1}[P(x)] = y_U[P(x)] \\
 = \frac{P(x)/100 - G(y_U^* - 1)}{G(y_U^*) - G(y_U^* - 1)} + (y_U^* - 5), \quad 0 \leq P(x) < 100, \\
 = K_Y + 5, \\
 P(x) = 100 \dots\dots\dots(1.2)$$

Asimismo, se utilizó una segunda ecuación
 (1.3) para estimar la igualación equipercen-
 tilar:

$$e_Y(x) = Q^{-1}[P(x)] = y_L[P(x)] \\
 = \frac{P(x)/100 - G(y_L^*)}{G(y_L^* + 1) - G(y_L^*)} + (y_L^* + .5), \quad 0 < P(x) \leq 100, \\
 = -.5, \\
 P(x) = 0 \dots\dots\dots(1.3)$$

Finalmente, se utilizó le ecuación (1.4) para la
 estimación de la función de la igualación equipercen-
 tilar, que es el promedio de y_U y y_L .

$$y = (y_U + y_L) / 2 \dots\dots\dots(1.4)$$

RESULTADOS

El criterio que se siguió para determinar la
 efectividad del método equipercen-
 tilar fue el
 propuesto por Kolen y Brennan (1995),
 quienes señalan que después de realizar la
 estimación equipercen-
 tilar, todos los momen-
 tos (medias, desviaciones estándar, sesgos y
 kurtosis) de las distribuciones de los puntajes
 deben haber sido ajustados; en este caso las
 versiones 3 y 4 con los momentos de la ver-
 sión 2.

La tabla II muestra la media, desviación es-
 tándar, sesgo y kurtosis de las tres versiones
 del EXHCOBA para cada una de sus 15
 áreas temáticas, antes de la equiparación.
 Asimismo, la tabla III muestra las diferencias
 entre los indicadores estadísticos de las ver-
 siones 3 y 4 con relación a la versión 2.

Tabla II. Indicadores estadísticos de las versiones 2, 3 y 4 del EXHCOBA

Áreas de K	Versión 2	Versión 3	Versión 4
------------	-----------	-----------	-----------

conocimiento del EXHCOBA	(reactivos)	N	m	s	Sesgo	Kurtosis	N	m	s	Sesgo	Kurtosis	N	m	s	Sesgo	Kurtosis
H. Verbales		883	19.72	4.31	-.34	-.29	887	18.59	4.11	-.20	-.45	888	19.19	4.34	-.34	-.28
H. Cuantitativas		883	19.39	5.72	-.34	-.65	887	18.42	5.98	-.24	-.69	888	17.49	6.03	-.10	-.87
Español	30	883	9.30	2.43	-.19	-.37	887	8.57	2.68	.01	-.34	888	8.79	2.40	-.07	-.38
Matemáticas	30	883	6.70	3.42	.47	-.43	887	7.02	3.16	.32	-.55	888	7.69	3.04	.28	-.53
Ciencias	15	883	12.33	3.15	-.10	-.28	887	13.49	2.97	-.33	-.22	888	12.73	2.91	-.47	.39
Naturales	15	883	11.79	3.47	-.02	-.47	887	12.40	3.43	-.17	-.47	888	10.48	3.57	.04	-.48
Ciencias Sociales	15	404	9.51	4.05	.19	-.63	422	9.84	4.05	.32	-.52	401	9.22	4.04	.38	-.40
Sociales	20	529	12.33	2.79	-.92	2.5	540	9.93	3.28	-.07	-.26	515	9.94	2.79	-.37	1.1
Mat. para Estadísticas	20	280	11.25	3.86	-.30	-.12	297	10.62	3.51	-.33	.32	273	12.07	3.82	-.88	1.5
Ciencias Sociales	20	213	10.15	4.25	.06	-.80	219	10.11	4.01	.23	-.77	244	9.39	3.94	.20	-.50
Económico-Admon.	20	155	9.86	3.51	.10	-.13	143	11.57	3.77	-.13	-.96	146	9.78	3.35	-.46	.63
Matemáticas Cálculo	20	326	8.56	3.30	.12	.57	316	8.03	3.36	-.30	.36	344	7.65	3.86	-.12	-.20
Biología	20	340	9.75	3.53	-.17	-.14	332	10.55	2.80	-.48	.55	356	8.00	3.11	-.16	.38
Química	20	277	11.84	3.88	-.67	.54	274	11.6	3.69	-1.3	1.8	271	10.72	3.65	-.76	1.0
Física	20	125	12.79	2.98	-.15	-.23	118	11.99	3.68	-.16	-.51	114	11.40	2.58	.00	-.43
Lenguaje																
Humanidades																

Tabla III. Diferencias en los momentos de las distribuciones de las versiones 2, 3 y 4

Áreas temáticas	Versión 2 vs 3				Versión 2 vs 4			
	m	s	Sesgo	kurtosis	m	s	Sesgo	kurtosis
H. Verbales	1.13	0.2	-0.14	0.16	0.53	-0.03	0	-0.01
H. Cuantitativas	0.97	-0.26	-0.1	0.04	1.9	-0.31	-0.24	0.22
Español	0.73	-0.25	-0.2	-0.03	0.51	0.03	-0.12	0.01
Matemáticas	-0.32	0.26	0.15	0.12	-0.99	0.38	0.19	0.1
Ciencias Naturales	-1.16	0.18	0.23	-0.06	-0.4	0.24	0.37	-0.67
Ciencias Sociales	-0.61	0.04	0.15	0	1.31	-0.1	-0.06	0.01
Mat. para Estadísticas	-0.33	0.00	-0.13	-0.11	0.29	0.01	-0.19	-0.23
Ciencias Sociales	2.4	-0.49	-0.85	2.76	2.39	0	-0.55	1.4
Económico-Admon.	0.63	0.35	0.03	-0.44	-0.82	0.04	0.58	-1.62
Matemáticas Cálculo	0.04	0.24	-0.17	-0.03	0.76	0.31	-0.14	-0.3
Biología	-1.71	-0.26	0.23	0.83	0.08	0.16	0.56	-0.76
Química	0.53	-0.06	0.42	0.21	0.91	-0.56	0.24	0.77
Física	-0.8	0.73	0.31	-0.69	1.75	0.42	-0.01	-0.52
Lenguaje	0.24	0.19	0.63	-1.26	1.12	0.23	0.09	-0.46
Humanidades	0.8	-0.7	0.01	0.28	1.39	0.4	-0.15	0.2

Nota: las diferencias positivas indican son a favor de la versión 2, mientras que las negativas son a favor de las versiones 3 o 4, según corresponda.

Respecto a la dificultad de las tres versiones de EXHCOBA, en general resultó más fácil la versión 2 que las versiones 3 y 4 en las áreas de habilidades verbales, habilidades cuantitativas, español, ciencias sociales (especialidad), matemáticas cálculo, lenguaje y humanidades. Asimismo, la versión 3 resultó más fácil que las versiones 2 y 4 en las áreas

de ciencias naturales, ciencias sociales, matemáticas estadística y biología. Finalmente, la versión 4 resultó más fácil que las versiones 2 y 3 en las áreas de matemáticas y económico-administrativo.

Respecto a los momentos de las distribuciones, se observan las siguientes diferencias promedio entre las versiones 2 y 3, y 2 y 4,

respectivamente: en las medias, de 0.83 y 1.01; en las desviaciones estándar, de 0.28 y 0.21; en los sesgos, de 0.25 y 0.23, y en las kurtosis, de 0.47 y 0.49.

Una vez equiparadas las versiones 3 y 4 con la versión 2, se obtuvieron los resultados que se muestran en la tabla IV.

**Tabla IV. Indicadores estadísticos de las versiones 2, 3 y 4 del EXHCOBA, después de la igualación equipercen-
 til**

Áreas de conocimiento del EXHCOBA	Total de reactivos	Versión 2					Equiparación Versión 3					Equiparación Versión 4				
		N	m	s	Sesgo	Kurtosis	N	m	s	Sesgo	Kurtosis	N	m	s	Sesgo	Kurtosis
H. Verbales																
H. Cuantitativas																
Español	30	883	19.72	4.31	-.34	-.29	887	19.72	4.29	-.34	-.31	888	19.72	4.29	-.34	-.30
Matemáticas	30	883	19.39	5.72	-.34	-.65	887	19.39	5.71	-.34	-.65	888	19.39	5.71	-.34	-.66
Ciencias Naturales	15	883	9.30	2.43	-.19	-.37	887	9.30	2.42	-.19	-.39	888	9.30	2.39	-.18	-.35
Ciencias Sociales	15	883	6.70	3.42	.47	-.43	887	6.70	3.39	.49	-.43	888	6.70	3.41	.48	-.44
Mat. Estadísticas	20	883	12.33	3.15	-.10	-.28	887	12.33	3.13	-.11	-.30	888	12.33	3.12	-.11	-.29
Ciencias Sociales Económico-Admon.	20	883	11.79	3.47	-.02	-.47	887	11.80	3.45	-.02	-.47	888	11.80	3.45	.01	-.55
Matemáticas Cálculo	20	404	9.51	4.05	.19	-.63	422	9.52	4.01	.20	-.65	401	9.53	4.03	.21	-.66
Biología	20	529	12.33	2.79	-.92	2.5	540	12.33	2.77	-.87	2.29	515	12.33	2.80	-.98	2.79
Química	20	280	11.25	3.86	-.30	-.12	297	11.26	3.83	-.29	-.13	273	11.26	3.81	-.25	-.32
Física	20	213	10.15	4.25	.06	-.80	219	10.15	4.24	.06	-.82	244	10.15	4.23	.06	-.81
Lenguaje	20	155	9.86	3.51	.10	-.13	143	9.86	3.47	.10	-.18	146	9.88	3.44	.15	-.31
Humanidades	20	326	8.56	3.30	.12	.57	316	8.52	3.33	.04	.54	344	8.60	3.17	.36	.17
	20	340	9.75	3.53	-.17	-.14	332	9.76	3.48	-.17	-.17	356	9.74	3.52	-.19	-.14
	20	277	11.84	3.88	-.67	.54	274	11.83	3.85	-.67	.57	271	11.84	3.90	-.73	.70
	20	125	12.79	2.98	-.15	-.23	118	12.77	2.97	-.17	-.28	114	12.80	2.90	-.10	-.26

En esta tabla se puede observar con facilidad que los indicadores estadísticos de las tres versiones del examen en cada una de sus 15

En esta tabla se puede observar con facilidad que los indicadores estadísticos de las tres versiones del examen en cada una de sus 15 áreas temáticas son equivalentes. Por lo anterior, podemos decir que la igualación equipercen-til resultó efectiva al ajustar los cuatro momentos de las distribuciones de las tres versiones del EXHCOBA.

Ahora bien, un análisis gráfico de los resultados nos permite conocer el comportamiento de las tres versiones igualadas a lo largo de toda la escala de puntuaciones. Con propósitos ilustrativos se presentan las figuras 1 y 2, que muestran las puntuaciones igualadas acumuladas de dos áreas temáticas del examen. La primera de ellas, muestra el caso donde la equiparación resultó homogénea a lo largo de la escala; la segunda muestra un caso donde la igualación no fue tan homogénea.

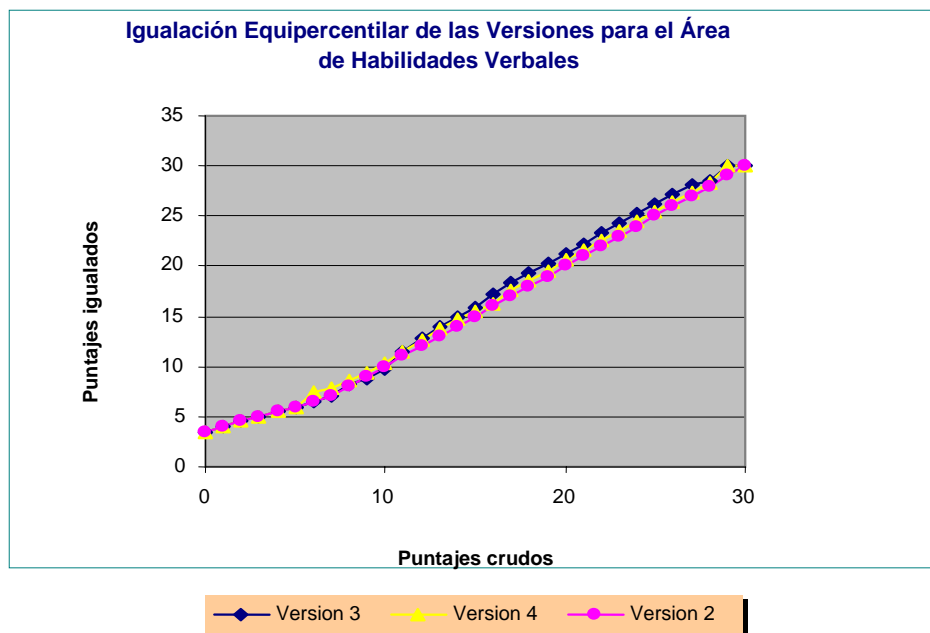


Figura 1. Ejemplo de igualación homogénea a lo largo de la escala

Como se puede observar, en la figura 1, la igualación de las tres versiones del área de Habilidades Verbales muestra patrones gráficos muy similares, tanto en los extremos como en su parte media de la escala.

En contraste, la figura 2 muestra los resultados de la equiparación para el caso del área de Español, donde se pueden observar dife-

rencias en los puntajes igualados en ciertas regiones de la escala; es decir zonas donde las distintas versiones pueden resultar más fáciles o más difíciles para los estudiantes. Por ejemplo, Las versiones 2 y 4 resultaron ligeramente más fáciles que la versión 3 en la parte inferior y ligeramente más difíciles en la parte superior.

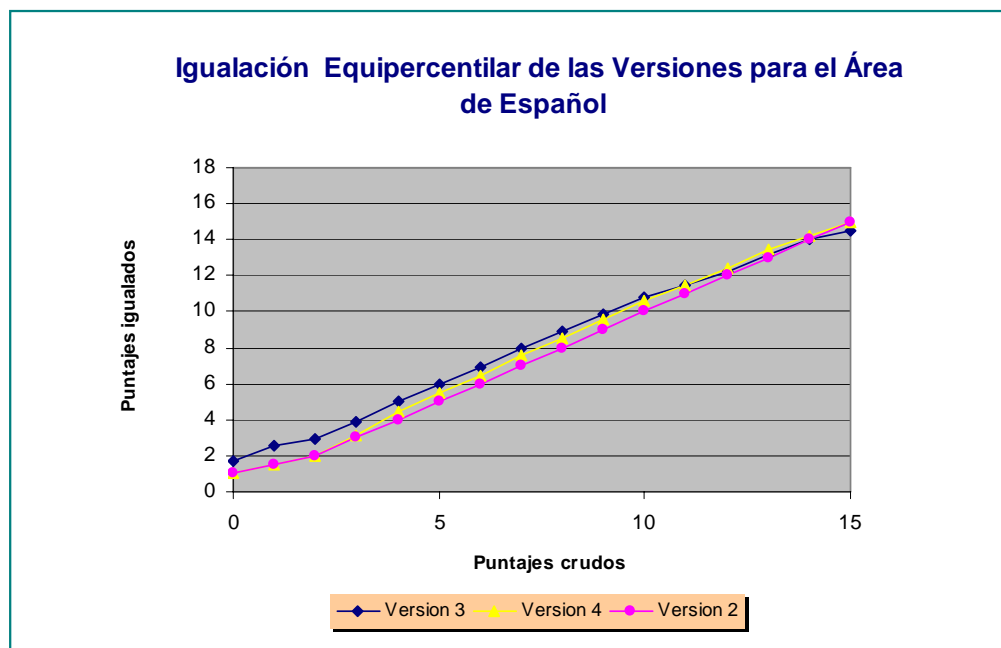


Figura 2. Ejemplo de igualación heterogénea a lo largo de la escala

Con respecto a las trece áreas temáticas restantes, cuyas gráficas no se presentan debido a su extensión, se pudo observar de manera general que las diferencias entre las puntuaciones transformadas resultaron mayores en los extremos de la distribución de las puntuaciones, por lo que queda claro que el método equipercentil produce resultados más similares en las partes medias de la misma.

DISCUSIÓN

Por el frecuente uso que se les da, los exámenes de gran escala y alto impacto utilizan más de una forma o versión de examen, procurando que éstos sean lo más similares posible. En teoría, dos versiones son equivalentes cuando comparten la misma estructura, provienen del mismo dominio y producen distribuciones equivalentes. Sin embargo, como en la práctica esto es imposible de lograr, es necesario utilizar procedimientos que igualen las distintas versiones de un examen. En esencia, la igualación comprime y/o extiende la escala de las puntuaciones de una versión de tal manera que corresponda a la escala de las puntuaciones de la versión de examen con la que se está equiparando.

Siendo el EXHCOBA un examen de ingreso a la educación superior, el cual se utiliza en varias universidades públicas del país, fue necesario que sus versiones se equipararan con un método robusto, como es el caso de la igualación equipercentil, que ha mostrado superioridad sobre el método lineal (Arce-Ferrer y Backhoff, 2002). Por lo anterior, el presente trabajo tuvo el propósito de equiparar y estudiar el comportamiento de tres versiones del EXHCOBA (versiones 2, 3 y 4) con el procedimiento propuesto por Kolen y Brennan (1995), en cada una de las 15 áreas temáticas del examen, cuya equiparación se realizó por separado.

Los resultados del estudio muestran que, antes de la equiparación, las distribuciones de las tres versiones del EXHCOBA diferían

en dificultad, así como en el resto de sus momentos estadísticos, y que el método de igualación equipercentil utilizado en este estudio resultó muy efectivo en ajustar los cuatro momentos (media, desviación estándar, sesgo y kurtosis) de las versiones 3 y 4 con la versión 2, para cada una de sus áreas temáticas.

Sin embargo, como ya lo habían señalado Arce-Ferre y Backhoff (2002), el modelo parece funcionar mejor en la mitad de la escala y es menos preciso en sus extremos. Por esta razón, varios autores han recomendado utilizar estrategias que mejoren los resultados de la igualación equipercentil; particularmente, Kolen y Brennan (1995) sugieren utilizar la estrategia conocida como “suavizado”.

El suavizado se utiliza para disminuir las irregularidades que típicamente se observan en la función estimada de igualación y que, por lo general, son el resultado de errores de muestreo (Hanson, 1991). Kolen (1984) demostró que una adecuada selección de la cantidad de suavizado mejora los resultados de la igualación. Sin embargo, también se ha reconocido que el uso inadecuado del suavizado introduce errores sistemáticos al proceso de igualación (Kolen, 1991), que podría causar un sesgo en los resultados de la igualación, mucho más serios que los errores aleatorios de la igualación equipercentil misma (Glowacki, 1991).

Por lo anterior, queda pendiente realizar un estudio que valore el grado en que esta estrategia mejora los resultados obtenidos hasta ahora con el EXHCOBA; aunque sabemos, en principio, que es poco probable que dos versiones de una prueba se puedan igualar perfectamente (Lord, 1982, citado por Lynn, 1993).

Queremos terminar apoyando la propuesta de Petersen, Kolen & Hoover (1989), quienes recomiendan que la decisión de seleccionar un método de igualación para una prueba en particular debe hacerse después de haber

explorado su comportamiento psicométrico. Esta aproximación empírica nos acercará, sin duda, a la meta de disponer de un sistema de evaluación mucho más sólido cuyas calificaciones sean lo más equivalentes posible y, por consiguiente, más justo cuando se tomen decisiones de seleccionar a los estudiantes que desean ingresar a una institución de educación.

REFERENCIAS

- AERA, APA, NCME (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association (APA).
- Arce-Ferrer, A. & Backhoff, E. (2002). Efecto del método de igualación en la equivalencia de los exámenes nacionales. *Revista Mexicana de Psicología* 19(2), 101-114.
- Backhoff, E. (2001). *Desarrollo, validación e implementación de un sistema para la selección de estudiantes a la Universidad Autónoma de Baja California*. Aguascalientes: Universidad Autónoma de Aguascalientes.
- Backhoff, E., Ibarra, M. & Rosas, M. (1995). Sistema Computarizado de Exámenes (SI-CODEX). *Revista Mexicana de Psicología*, 10(1), 55-62.
- 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. & Tirado, F. (1992). Desarrollo del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista de la Educación Superior*, 83, 23-44.
- Glowacki, M. (1991). An Analysis of Test Equating Models for the Alabama High School Graduation Examination. Design. 27p. Lexington, KY: Paper Presented at the *Annual Meeting of the Mid-South Educational Research Association* (Eric Document Reproduction Service No. ED 340 720).
- Hanson, B. (1991). A comparison of bivariate smoothing methods in common-item equipercentile equating. *Applied Psychological Measurement*, 15, 391-408.
- Holland, P. & Rubin, D. (1982). *Test equating*. Nueva York: Academic Press.
- Kolen, M. (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational Statistics*, 9, 25-44.
- Kolen, M. (1991). Smoothing methods for estimating test score distributions. *Journal of Educational Measurement*, 28, 257-282.
- Kolen, M. & Brennan, R. (1995). *Test equating*. Nueva York: Springer-Verlag.
- Linn, R. L. (1993). *Educational measurement*. (3rd. ed.). Nueva York: American Council on Education. Series on Higher Education. Oryx Press.
- Martínez, F. Backhoff, E., Castañeda, S., Del la Orden, A., Schmelkes, S., Solano-Flores, G., Tristán, A. y Vidal, R. (2000). *Estándares de calidad para instrumentos de evaluación educativa*. México: CENEVAL.
- Petersen, N., Kolen, M. & Hoover, H. (1989). Scaling, norming, and equating. En R. Linn (Ed.), *Educational measurement*. New York, NY: Macmillan.
- Tristán, A. (1996). Igualación o equivalencia de calificaciones utilizando la computadora. 2do. *Memorias del Foro Nacional de Evaluación Educativa*. México: CENEVAL.
- Walberg, H. & Haertel, G. (1992). *The international encyclopedia of educational evaluation*. New York: Pergamon Press.

ABOUT THE AUTHORS / SOBRE LOS AUTORES

Luz Elena Antillón. Maestra en Ciencias Educativas por el Instituto de Investigación y Desarrollo Educativo (IIDE) de la Universidad Autónoma de Baja California (UABC), en Ensenada, México. Su campo de interés es la Evaluación Educativa, la Estadística informática y la Psicometría para la Investigación Educativa. Ha colaborado en varios proyectos de investigación como son el Examen de Habilidades y Conocimientos Básicos (EXHCOBA), el Examen de Egreso del Idioma Ingles (EXEDII), la conformación de redes para la producción de textos científicos y de divulgación en educación y humanidades de la UABC, Innovación en la enseñanza de estadística mediante el uso de objetos de aprendizaje y constructivismo. Ha participado en la obtención y aseguramiento de acreditación y certificación de programas de licenciatura y posgrado. Actualmente es técnica académica de Investigación del Instituto de Investigación y Desarrollo Educativo, y docente de la carrera de Psicología en la Facultad de Ciencias Administrativas y Sociales de la UABC

Norma Larrazolo. Licenciada en psicología por la Universidad Nacional Autónoma de México, Maestra en Educación por la Universidad de Washington y Doctora en Ciencias Educativas por la Universidad Iberoamericana del Noroeste. Su campo de interés es la evaluación del aprendizaje en particular las pruebas a gran escala. Es responsable del Examen de Habilidades y Conocimientos Básicos (EXHCOBA) que se aplica en varias universidades públicas del país como examen de admisión desde 1994 y colabora en el Examen de Egreso del Idioma Inglés. Ha participado en congresos y publicado textos en el área de la evaluación educativa y ha sido docente de los programas de maestría y doctorado en educación.

Eduardo Backhoff (backhoff@inee.edu.mx). Licenciado en psicología por la Universidad Nacional Autónoma de México, Maestro en Educación por la Universidad de Washington y Doctor en Educación por la Universidad Autónoma de Aguascalientes. Su campo de interés es el desarrollo y validación de pruebas de aprendizaje de gran escala y la evaluación asistida por computadora. Es coautor del Examen de Habilidades y Conocimientos Básicos (EXHCOBA), que se utiliza en varias universidades públicas del país para seleccionar a sus estudiantes. Es responsable del desarrollo de los Exámenes de la Calidad y el Logro Educativos (Excale), que se utilizan para evaluar la calidad del Sistema Educativo de México. Desde febrero de 2004 es Director de Pruebas y Medición del Instituto Nacional para la Evaluación de la Educación (INEE).

ARTICLE RECORD / FICHA DEL ARTÍCULO

Reference / Referencia	Antillón, Luz Elena, Larrazolo, Norma y Backhoff, Eduardo(2006). Igualación equipercenil del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). <i>RELIEVE</i> , v. 12, n. 2. http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2_2.htm . Consultado en (poner fecha).
Title / Título	Igualación equipercenil del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). [<i>Equipercenile equating of the Basic Ability and Knowledge Test (EXHCOBA)</i>]
Authors / Autores	Luz Elena Antillón, Norma Larrazolo y Eduardo Backhoff
Review / Revista	Revista ELectrónica de Investigación y EValuación Educativa (RELIEVE), v. 12, n. 2
ISSN	1134-4032
Publication date / Fecha de publicación	2006 (Reception Date: 2005 September 26; Approval Date: 2006 July 10; Publication Date: 2006 July 27)
Abstract / Resumen	<p><i>Equipercenile equating method is a statistical procedure where student raw scores of two different versions of the same test are considered equated if they correspond to the same percentile range. A graphic curve is presented to describe the difficult differences from version to version of a test. This work was aimed to estimate the equipercenil equating values, by academic content area, of the Basic Ability and Knowledge Test (EXHCOBA, by its Spanish acronym) that is used by the University of Baja California (UABC) as a selection student test. This norm-referenced test has excellent quality standards, a high technological development, several reliability and validity support studies, and others good psychometric parameters. Estimation of equating parameters was done applying the analytic method described by Kolen and Brennan (1995), with the random group procedure utilized by UABC to collect data. Results shows that equating was effective to adjust four statistical moments (mean, standard deviation, bias, and kurtosis) of the frequency distributions of EXHCOBA's version 3 and 4 compared with version 2, by content area, producing equal score distributions. Nevertheless, irregularities appeared at the ends of the curves that suggests the need of a smoothing procedure.</i></p> <p>La igualación equipercenil es un método estadístico en el cual los puntajes crudos de dos versiones de una prueba se consideran igualados si ellos corresponden al mismo rango percentilar en un grupo de examinados. En la igualación equipercenil se presenta una curva para describir las diferencias de dificultad de versión a versión. Este trabajo tuvo como objetivo estimar la igualación equipercenil sin suavizado de las versiones 3 y 4, con la versión 2, por área temática del Examen de Habilidades y Conocimientos Básicos (EXHCOBA) que utiliza la Universidad Autónoma de Baja California (UABC) para la selección de aspirantes, examen que posee un nivel de calidad excelente, un desarrollo tecnológico elevado y cuenta con estudios constantes de fiabilidad, validez y otros indicadores psicométricos. La estimación de la igualación equipercenil se llevó a cabo con el método analítico que describen Kolen y Brennan (1995) y con el diseño de grupos aleatorios que utiliza la UABC para la recolección de datos. Los resultados de este estudio indicaron que la igualación resultó efectiva al ajustar cuatro momentos (media, desviación estándar, sesgo y curtosis) de las distribuciones de las versiones 3 y 4 con la versión 2 por área temática, produciendo distribuciones iguales en las puntuaciones. Sin embargo, algunas irregularidades se encontraron en los extremos de las curvas que hacen suponer la falta de un proceso de suavizado.</p>
Keywords / Descriptores	<i>Test equating, equipercenil equating method, Basic Abilities and Knowledge Test</i> Igualación de pruebas, método equipercenil, Examen de Habilidades y Conocimientos Básicos
Institution / Institución	Universidad Autónoma de Baja California e Instituto Nacional para la Evaluación de la Educación (México)
Publication site / Dirección	http://www.uv.es/RELIEVE
Language / Idioma	Spanish (Title, abstract and keywords in English)

**Revista ELectrónica de Investigación y EValuación Educativa
(RELIEVE)**

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).