

IGUALACIÓN LINEAL DE TRES VERSIONES DEL EXAMEN DE HABILIDADES Y CONOCIMIENTOS BÁSICOS (EXHCOBA)

Luz Elena Antillón, Norma Larrazolo y Eduardo Backhoff

Revista Iberoamericana de Evaluación Educativa 2008 - Volumen 1, Número 2

<http://www.rinace.net/riee/numeros/vol1-num2/art12.pdf>

Fecha de recepción: 12 de septiembre de 2008

Fecha de aceptación: 8 de octubre de 2008



La evaluación educativa es un campo cuyo principal objetivo es proporcionar evidencias para ayudar en la toma de decisiones educativas con respecto a un individuo, grupo, institución o programa. Una forma de proporcionar esta información son las pruebas estandarizadas que se aplican a grupos numerosos en una institución, estado o país y que tienen un gran impacto; es decir, que las decisiones que se toman con respecto a sus resultados, afectan a los individuos, instituciones y programas. Estas pruebas a gran escala se clasifican en pruebas normativas y criterioles.

Las pruebas normativas relacionan el desempeño de un individuo con respecto al grupo que realiza el mismo examen. Por su parte, las pruebas criterioles evalúan qué tanto el individuo consigue el criterio previamente establecido. Para regular la elaboración, aplicación, calificación e interpretación de las pruebas educativas y psicológicas, se han establecido indicadores internacionales que se utilizan para juzgar la calidad técnica de un instrumento (AERA, APA, NCME, 1999).

Uno de estos indicadores de calidad de una prueba, es la igualación (*equating*). Cuando las pruebas se administran en múltiples ocasiones es necesario generar versiones paralelas, que coincidan en contenido, formato, tipo de ítems, estructura y especificaciones estadísticas (Kolen y Brennan, 1995). Sin embargo, dos o más versiones de una prueba, no llegan a tener una equivalencia exacta en su dificultad, lo que hace necesario utilizar un proceso estadístico de igualación, que permita el ajuste de las puntuaciones para hacer equivaler la dificultad de versión a versión (Holland y Rubin, 1982). Sin embargo, es deseable que desde el momento en que se diseñan y construyen las distintas versiones de una prueba, se considere que éstas sean lo más parecidas y equiparables posibles, tanto en su estructura como en su contenido y extensión. Con lo anterior se logrará una mejor equiparación y se evitará el riesgo de tener que hacer ajustes estadísticos al final del proceso.

La Teoría Clásica de los Tests (TCT) dispone de distintos métodos de igualación y diseños de recolección de datos. En general, los métodos clásicos de igualación caen dentro de dos categorías principales: igualación lineal e igualación equipercenil (Hambleton, Swaminathan y Roger, 1991). El método de igualación lineal implica cambiar la escala de manera que se modifique la media y la desviación estándar, mientras se conserva exactamente la forma de la distribución y, en consecuencia las posiciones relativas de los individuos en dicha distribución (Magnusson, 1993). Este método es práctico, porque comprende solamente una transformación simple de la puntuación cruda a una puntuación escala. De esta manera las puntuaciones de dos versiones de una prueba se consideran igualadas si las dos puntuaciones corresponden al mismo número de desviaciones estándar arriba o abajo de la media en el grupo de examinados (Kolen y Brennan, 1987). El segundo método de igualación equipercenil se aborda en Antillón, L. E., Larrazolo, N., y Backhoff, E. (2006).

En este trabajo, se realizó la estimación del proceso estadístico de igualación de versiones de una prueba con el método lineal en un examen de admisión de nivel superior "Examen de Habilidades y Conocimientos Básicos" (EXHCOBA). Este examen, desarrollado por Backhoff y Tirado (1992), como instrumento computarizado para seleccionar estudiantes de nivel superior y medio superior, se utiliza en varias instituciones públicas y privadas del país como parte fundamental de sus procesos de ingreso (por ejemplo, la Universidad Autónoma de Baja California, la Universidad de Sonora, la Universidad de Guanajuato, entre otras instituciones educativas)

El EXHCOBA cuenta actualmente con varias versiones paralelas y desde su inicio ha venido acumulando evidencias de su calidad técnica, mismas que se han publicado en revistas y foros nacionales e

internacionales (Backhoff y Tirado, 1993; 1994, Larrazolo, 1997a; 1997b, Backhoff, Tirado, Larrazolo y Antillón, 1996; Backhoff, Larrazolo y Rosas, 2000, Backhoff, Tirado y Larrazolo, 2001). Sin embargo, aún no se reportan los indicadores de la igualación de las versiones paralelas generadas en dicho examen. La finalidad de estimar la igualación lineal es que las puntuaciones de los estudiantes sean más justas y equivalentes: de no implementarse esta estrategia de ajustes algunos estudiantes se verán desfavorecidos, es decir, los individuos que utilicen una versión más difícil tendrán desventajas sobre aquellos individuos que utilicen una versión más fácil, debido a las variaciones en las dificultades de los ítems de cada versión.

Por lo anterior, el objetivo principal es la estimación de la igualación lineal de tres versiones del EXHCOBA (V2, V3 y V 4) por área temática, así como la estimación de la dificultad en cada una de ellas.

1. METODOLOGÍA

1.1. Sujetos

La Universidad Autónoma de Baja California ha venido administrando el EXHCOBA desde 1992 a la fecha como examen de admisión. En 1998 realizó un estudio con 8,925 aspirantes con la modalidad computarizada de ese examen. Del total de estudiantes, aproximadamente, 10 % respondió a una de las tres versiones del examen de manera aleatoria. Así, a 883 estudiantes se les administró la versión 2, a 887 la versión 3 y finalmente, a 888 aspirantes se les aplicó la versión 4. El resto de los estudiantes (70%) respondió a versiones mezcladas de las tres versiones, que el sistema hace en forma automática al momento que el alumno inicia su examen, razón por la que no participaron en este estudio.

1.2. Instrumento

El EXHCOBA está conformado en tres secciones: Habilidades básicas (habilidades verbales y habilidades cuantitativas), Conocimientos básicos (español, matemáticas, ciencias naturales y ciencias sociales) y Conocimientos básicos para especialidad (matemáticas estadística, económico-administrativa, biología, física, química, humanidades, lenguaje, matemáticas cálculo y ciencias sociales). La primera sección corresponde a aprendizajes a nivel primaria, la segunda a nivel secundaria y la tercera a nivel bachillerato (Tabla I).

TABLA I. ESTRUCTURA DEL EXHCOBA

Niveles	Secciones	Áreas temáticas	Ítems
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	Ítems
Bachillerato	Conocimientos Básicos para Especialidad	Estadística	20
		Cs. Sociales	20
		Cs. Económico-administrativas	20
		Cálculo	20
		Biología	20
		Química	20
		Física	20
		Lenguaje	20
	Humanidades	20	

Todas las versiones del EXHCOBA, o formas paralelas, se ajustan a los criterios de estructura, contenido, tipo de ítems, y formato; sin embargo, por razones naturales éstos difieren en sus dificultades relativas (por ejemplo, forma de redactar el ítem, distintas opciones de respuesta). La Tabla II muestra un ejemplo de su estructura temática.

TABLA II. ESTRUCTURA TEMÁTICA DEL EXHCOBA

<i>Habilidades verbales (versión 2)</i>	<i>Habilidades verbales (versión 3)</i>	<i>Habilidades verbales (versión 4)</i>
<p>Área del nodo: Comprensión de palabras.</p> <p>Nodo: Vocabulario (antónimos)</p> <p>Reactivo: Escoge la palabra que más se acerque al significado opuesto del concepto: <u>Aceptable</u>.</p> <p>a) Inadmisible b) Tolerante c) Aprobable d) Imposible e) No sé</p>	<p>Área del nodo: Comprensión de palabras.</p> <p>Nodo: Vocabulario (antónimos)</p> <p>Reactivo: Escoge la palabra que más se acerque al significado opuesto del concepto: <u>Precavido</u>.</p> <p>a) Cautó b) Prudente c) Inepto d) Negligente e) No sé</p>	<p>Área del nodo: Comprensión de palabras.</p> <p>Nodo: Vocabulario (antónimos)</p> <p>Reactivo: Escoge la palabra que más se acerque al significado opuesto del concepto: <u>Inminente</u>.</p> <p>a) Remoto b) Cercano c) Apremiante d) Inaplazable e) No sé</p>

Este examen es de modalidad 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. En este trabajo, se utilizaron las versiones 2, 3 y 4 (la primera versión se utilizó como guía de estudio).

2. PROCEDIMIENTO

Para este estudio se dispuso de tres bases de datos. Cada una comprendió lo siguiente: Ficha de cada estudiante, tipo de examen (de siete que existen), versión asignada, y los resultados de 310 reactivos del examen en código binario (respuestas correctas e incorrectas). En el EXHCOBA los reactivos son de opción múltiple, con cuatro respuestas posibles y una quinta opción denominada "no sé".

El trabajo se realizó por área temática del examen: habilidades verbales, habilidades cuantitativas, español, matemáticas, ciencias naturales, ciencias sociales, matemáticas estadística, ciencias sociales (especialidad), económico-administrativo, matemáticas cálculo, biología, química, física, lenguaje y humanidades.

El método de igualación utilizado en este trabajo fue el conocido como *equiparación lineal*, para lo cual se utilizó el procedimiento descrito por Kolen y Brennan (1995). Se eligió este procedimiento de igualación lineal debido a la sencillez de sus cálculos, a lo robusto de sus resultados y a la facilidad con

que se interpretan sus resultados (Yang y Houang, 1996). Primeramente, se determinó el nivel de dificultad, el número de aciertos, media, desviación estándar, sesgo y *kurtosis* de las puntuaciones crudas de cada una de las 15 áreas temáticas de las 3 versiones (2, 3 y 4); posteriormente, se realizó la estimación de las equivalencias con el método lineal, para la transformación de las puntuaciones crudas de la versión 3 y 4 a las puntuaciones crudas de la versión 2 por área temática y; por último se obtuvo la media, desviación estándar, sesgo y *kurtosis* de las puntuaciones resultantes de la estimación de la igualación lineal de las 15 áreas temáticas de las tres versiones (2, 3 y 4). Este proceso se realizó con la ayuda de los paquetes de cómputo estadísticos para las ciencias sociales *SPSS* (versión 10) y la hoja de cálculo *EXCEL* (versión 5).

La ecuación para la transformación lineal de las puntuaciones observadas de la versión X a la escala de la versión Y (Kolen y Brennan, 1995), es la siguiente (1.0):

$$I_Y(x) = y = \frac{\sigma(Y)}{\sigma(X)}x + \left[\mu(Y) - \frac{\sigma(Y)}{\sigma(X)}\mu(X) \right] \quad (1.0)$$

Donde:

$I_Y(x)$ = Es la ecuación de la conversión lineal de las puntuaciones observadas en la versión X a la escala de

la versión Y

$\sigma(Y)$ \square Es la desviación estándar de la versión Y

$\sigma(X)$ \square Es la desviación estándar de la versión X

y = Representa un puntaje particular en la versión Y

x = Representa un puntaje particular en la versión X

$\mu(Y)$ \square Es la media de la versión Y de una población examinada

$\mu(X)$ \square Es la media de la versión X de una población examinada

De acuerdo con Yang y Houang (1996), la selección de un modelo de igualación dependerá del propósito que se persigue, de los fundamentos teóricos, de la factibilidad y precisión del modelo; así como de las características de los examinados y los datos de la prueba. Sin embargo, seleccionar un método de igualación dependerá del diseño de recolección de datos que se emplee y no a la inversa. En este estudio, el EXHCOBA utiliza el diseño de grupos aleatorios para el método estadístico tradicional de igualación lineal que describen Kolen y Brennan (1995). La característica principal de éste diseño consiste en asignar al azar una versión distinta del examen a cada grupo de examinados.

El diseño de grupos aleatorios junto con el método de igualación lineal es apropiado utilizarlo cuando se tengan las condiciones adecuadas de estandarización y control de calidad, y sus versiones se construyan con las mismas especificaciones estadísticas, las muestras sean pequeñas y la dificultad de las versiones de una prueba sean similares (Kolen y Brennan, 1995); condiciones que se satisfacen para el caso de las

tres versiones del EXHCOBA analizadas como se verá a continuación.

3. RESULTADOS

Los autores (Backhoff, *et al.*, 2000) señalan en su publicación "Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA)", que el nivel medio de dificultad del examen, medido por la proporción de aciertos, debe oscilar entre 0.5 y 0.6. Entre más alta sea la proporción que se obtenga, la dificultad será menor, lo que significa que el ítem será más fácil; y viceversa, cuando la proporción sea menor, la dificultad será mayor, lo que indicará que el ítem será más difícil. En la tabla III se muestra el nivel de dificultad por área temática de las tres versiones del EXHCOBA (2, 3 y 4), donde se observa de manera general que existen áreas como por ejemplo química con un mayor nivel de dificultad en las tres versiones, es decir, los ítems son más difíciles. Existen otras áreas como por ejemplo ciencias naturales que tienen un nivel medio de dificultad en las tres versiones. En general, podemos observar que entre las tres versiones de cada área temática, hay similitud en el nivel de dificultad.

TABLA III. NIVEL DE DIFICULTAD POR ÁREA TEMÁTICA

Áreas temáticas	Nivel de dificultad por área temática de tres versiones		
	Versión 2	Versión 3	Versión 4
Habilidades verbales	0.66	0.62	0.64
Habilidades cuantitativas	0.65	0.61	0.58
Español	0.62	0.57	0.59
Matemáticas	0.45	0.47	0.51
Ciencias naturales	0.62	0.67	0.64
Ciencias sociales	0.59	0.62	0.52
Matemáticas estadística	0.48	0.49	0.46
Ciencias sociales especialidad.	0.62	0.50	0.50
Económico-Administrativa	0.56	0.53	0.60
Matemáticas cálculo	0.51	0.51	0.47
Biología	0.49	0.58	0.49
Química	0.43	0.40	0.38
Física	0.49	0.53	0.40
Lenguaje	0.59	0.58	0.54
Humanidades	0.64	0.60	0.57

Respecto a los resultados de equivalencia estadística (sin igualar) de cada área temática en las versiones 2, 3 y 4 del EXHCOBA, se utilizó el análisis de varianza (ANOVA) para determinar si existen diferencias significativas entre las medias de cada versión. Esta es una prueba estadística que analiza si más de dos grupos difieren significativamente entre sí en sus medias. Las hipótesis a probar son las siguientes:

Ho: Propone que los grupos no difieren significativamente

Ha: Los grupos difieren significativamente entre sí.

Si la significancia de "F" (F prob.) es menor a un nivel de significancia de $\alpha = .05$, entonces "F" resulta significativo, lo que quiere decir que los grupos difieren significativamente entre sí, por lo que la hipótesis alterna está fundamentada y la hipótesis nula se rechaza. Cuando la significancia de "F" (F prob.) es mayor a un nivel de significancia $\alpha = .05$, entonces "F" resulta no significativo, lo que quiere decir que

no hay diferencias significativas en los grupos, por lo tanto la hipótesis nula se acepta. Sin embargo, este análisis de varianza unidireccional solamente nos indica si la diferencia entre las medias y las distribuciones de los grupos son o no significativos, pero no nos indica a favor de qué grupo lo es, por ello, el análisis de Tukey DSH (Diferencia Significativa Honesta) realiza comparaciones múltiples de las medias y las distribuciones de los grupos (Kohler, 1999).

La tabla IV muestra los indicadores estadísticos del EXHCOBA de las tres versiones *sin igualar*, por área de conocimiento, como la media, la desviación estándar, el sesgo y la kurtosis. En esta tabla se pueden observar las diferencias entre estos indicadores estadísticos cuando se comparan las versiones del examen 2, 3 y 4, en cada área temática.

TABLA IV. INDICADORES ESTADÍSTICOS DE LAS VERSIONES 2, 3 Y 4 SIN IGUALAR POR ÁREAS DE CONOCIMIENTO DEL EXHCOBA

Áreas de conocimiento del EXHCOBA	k	Versión 2 Sin igualar					Versión 3 Sin igualar					Versión 4 Sin igualar				
		N	μ	σ	Sesgo	Kurtosis	N	μ	σ	Sesgo	Kurtosis	N	μ	σ	Sesgo	Kurtosis
Habilidades verbales	30	883	19.72	4.31	-.34	-.29	887	18.59	4.11	-.20	-.45	888	19.19	4.34	-.34	-.28
Habilidades cuantitativas	30	883	19.39	5.72	-.34	-.65	887	18.42	5.98	-.24	-.69	888	17.49	6.03	-.10	-.87
Español	15	883	9.30	2.43	-.19	-.37	887	8.57	2.68	.01	-.34	888	8.79	2.40	-.07	-.38
Matemáticas	15	883	6.70	3.42	.47	-.43	887	7.02	3.16	.32	-.55	888	7.69	3.04	.28	-.53
Ciencias naturales	20	883	12.33	3.15	-.10	-.28	887	13.49	2.97	-.33	-.22	888	12.73	2.91	-.47	.39
Ciencias sociales	20	883	11.79	3.47	-.02	-.47	887	12.40	3.43	-.17	-.47	888	10.48	3.57	.04	-.48
Matemáticas estadística	20	404	9.51	4.05	.19	-.63	422	9.84	4.05	.32	-.52	401	9.22	4.04	.38	-.40
Ciencias sociales especialidad	20	529	12.33	2.79	-.92	2.5	540	9.93	3.28	-.07	-.26	515	9.94	2.79	-.37	1.1
Económico administrativas	20	280	11.25	3.86	-.30	-.12	297	10.62	3.51	-.33	.32	273	12.07	3.82	-.88	1.5
Matemáticas cálculo	20	213	10.15	4.25	.06	-.80	219	10.11	4.01	.23	-.77	244	9.39	3.94	.20	-.50
Biología	20	155	9.86	3.51	.10	-.13	143	11.57	3.77	-.13	-.96	146	9.78	3.35	-.46	.63
Química	20	326	8.56	3.30	.12	.57	316	8.03	3.36	-.30	.36	344	7.65	3.86	-.12	-.20
Física	20	340	9.75	3.53	-.17	-.14	332	10.55	2.80	-.48	.55	356	8.00	3.11	-.16	.38
Lenguaje	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
Humanidades	20	125	12.79	2.98	-.15	-.23	118	11.99	3.68	-.16	-.51	114	11.40	2.58	.00	-.43

Donde: k = número de ítems; N = número de estudiantes; μ = media; σ = desviación estándar

Sin embargo, es importante tomar en cuenta el análisis de varianza efectuado, que se presentan en la tabla V, que nos indica con mayor precisión si existen diferencias significativas entre ellas. Aquí podemos observar que efectivamente en la mayoría de las áreas temáticas de las tres versiones del examen existen diferencias significativas, no así en la sección de matemáticas estadísticas y matemáticas cálculo.

En la misma tabla V se presentan los resultados con el método de Tukey HSD que muestra la significancia de las diferencias, de acuerdo a su versión. Se observa que en las áreas temáticas de las versiones 2 y 3 se presentan diferencias significativas en las áreas de habilidades verbales, habilidades cuantitativas, español, ciencias naturales, ciencias sociales, ciencias sociales (especialidad), biología y física. Por el contrario, no se observan diferencias significativas en las áreas de matemáticas, matemáticas estadística, económico-administrativo, matemáticas cálculo, química, lenguaje, y humanidades no presentan diferencias significativas.

Por su parte, en las versiones 2 y 4, las diferencias significativas se observan en las áreas de habilidades verbales, habilidades cuantitativas, español, matemáticas, ciencias naturales, ciencias sociales, ciencias sociales (especialidad), económico-administrativo, química, física, lenguaje y humanidades. En estas

mismas versiones no se observan diferencias significativas en las áreas de matemáticas estadística, matemáticas cálculo, biología.

TABLA V. ANÁLISIS DE VARIANZA (ANOVA) DE LAS MEDIAS ENTRE LAS VERSIONES DEL EXHCOBA POR ÁREA TEMÁTICA

Áreas temáticas	ANOVA Significancia de "F" $\alpha = .05$	Tukey HSD Significancia $\alpha = .05$	
		Versión 2 con versión 3	Versión 2 con versión 4
Habilidades verbales	.000	.000	.024
Habilidades cuantitativas	.000	.002	.000
Español	.000	.000	.000
Matemáticas	.000	.086	.000
Ciencias naturales	.000	.000	.015
Ciencias sociales	.000	.001	.000
Matemáticas estadística	.087	.473	.555
Ciencias sociales especialidad	.000	.000	.000
Económico Administrativas	.000	.102	.026
Matemáticas cálculo	.073	.994	.110
Biología	.000	.000	.977
Química	.004	.144	.003
Física	.000	.003	.000
Lenguaje	.001	.727	.001
Humanidades	.003	.112	.002

Por otra parte, se llevó a cabo la estimación de la igualación con el método lineal para la transformación de las puntuaciones crudas de la versión 3 y 4 a las puntuaciones crudas de la versión 2 por área temática. Magnusson (1993) señala que una transformación lineal implica cambiar la escala de manera que se modifique la media y la desviación estándar, mientras se conserva exactamente la forma de la distribución y, en consecuencia las posiciones relativas de los individuos en dicha distribución. Es decir, la igualación lineal resulta efectiva si se ajustan los momentos de la media y desviación estándar en las tres versiones por área temática y se mantiene la misma distribución. En la tabla VI se pueden observar los resultados de los indicadores estadísticos como la media, desviación estándar, sesgo y kurtosis de las tres versiones (2, 3 y 4) *igualadas* por área de conocimientos del EXHCOBA. Los resultados muestran que se ajustaron la media y desviación estándar de las puntuaciones de la versión 3 y 4 con las puntuaciones de la versión 2, además, la distribución se mantuvo igual, es decir, no se ajustaron el sesgo y la kurtosis de las versiones 3 y 4 con la versión 2.

Por cuestión de espacio, sólo se muestra la figura 1, del área de química, de las 16 que se elaboraron para este estudio, con los resultados de la igualación lineal de las versiones 2, 3 y 4. Se observó que en todas las áreas temáticas, las diferencias entre las puntuaciones transformadas con el método lineal de las tres versiones resultaron mayores en los extremos de la distribución, especialmente en el extremo inferior, no así en la mitad y en los niveles superiores de la distribución. Estas diferencias muestran la

dificultad de las versiones, donde la versión 2 del área de química es más fácil que la versión 3, y ésta a su vez, más fácil que la versión 4. Puede compararse lo antes mencionado, observando los resultados obtenidos en la tabla III de los niveles de dificultad para el área de química.

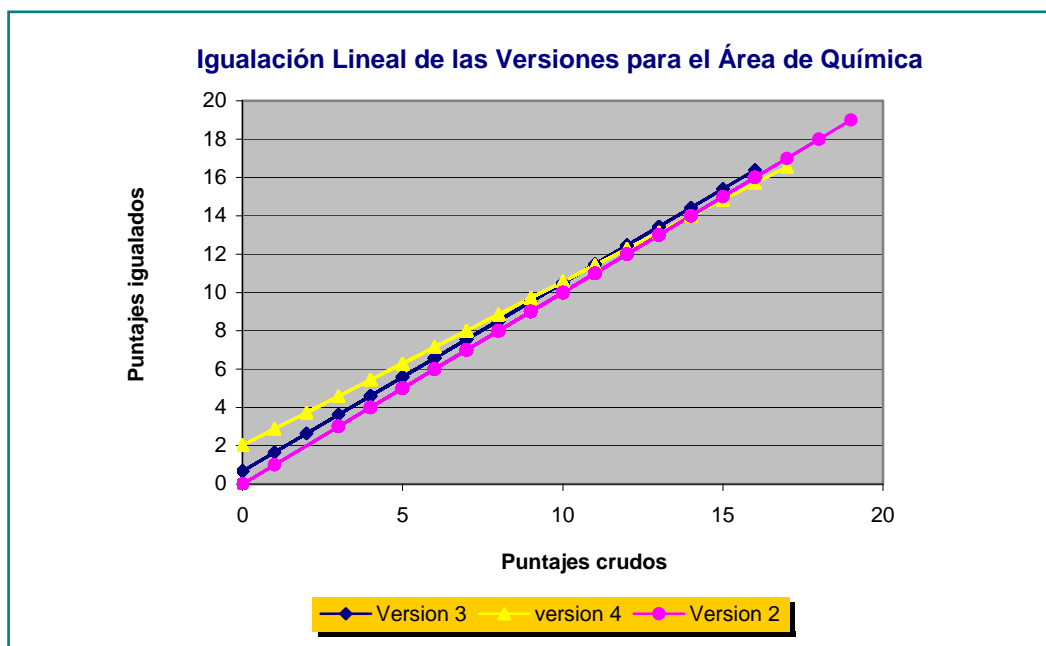
TABLA VI. INDICADORES ESTADÍSTICOS DE LA VERSIÓN 3 Y 4 IGUALADAS A LA VERSIÓN 2 POR ÁREAS DE CONOCIMIENTO CON EL MÉTODO LINEAL

Áreas de conocimiento del EXHCOBA	k	Versión 2					MÉTODO LINEAL Versión 3					MÉTODO LINEAL Versión 4				
		N	μ	σ	Sesgo	Kur-tosis	N	μ	σ	Sesgo	Kur-tosis	N	μ	σ	Sesgo	Kur-tosis
		Habilidades verbales	30	883	19.72	4.31	-.34	-.29	887	19.72	4.31	-.20	-.45	888	19.72	4.31
Habilidades cuantitativas	30	883	19.39	5.72	-.34	-.65	887	19.39	5.72	-.24	-.69	888	19.39	5.72	-.10	-.87
Español	15	883	9.30	2.43	-.19	-.37	887	9.30	2.43	.01	-.34	888	9.30	2.43	-.07	-.38
Matemáticas	15	883	6.70	3.42	.47	-.43	887	6.70	3.42	.32	-.55	888	6.70	3.43	.28	-.53
Ciencias naturales	20	883	12.33	3.15	-.10	-.28	887	12.33	3.12	-.33	-.22	888	12.33	3.15	-.47	.39
Ciencias sociales	20	883	11.79	3.47	-.02	-.47	887	11.80	3.47	-.17	-.47	888	11.79	3.47	.04	-.48
Matemáticas estadística	20	404	9.51	4.05	.19	-.63	422	9.51	4.05	.32	-.52	401	9.51	4.05	.38	-.40
Ciencias sociales especialidad	20	529	12.33	2.79	-.92	2.5	540	12.33	2.79	-.07	.26	515	12.33	2.79	-.37	1.1
Económico administrativas	20	280	11.25	3.86	-.30	-.12	297	11.25	3.86	-.33	-.32	273	11.25	3.86	-.88	1.5
Matemáticas cálculo	20	213	10.15	4.25	.06	-.80	219	10.15	4.25	.23	-.77	244	10.15	4.25	.20	-.50
Biología	20	155	9.86	3.51	.10	-.13	143	9.86	3.51	-.13	-.96	146	9.86	3.51	-.46	.63
Química	20	326	8.56	3.30	.12	.57	316	8.56	3.30	-.30	.36	344	8.56	3.30	-.12	-.20
Física	20	340	9.75	3.53	-.17	-.14	332	9.76	3.53	-.48	.55	356	9.75	3.53	-.16	.38
Lenguaje	20	277	11.84	3.88	-.67	.54	274	11.84	3.88	-1.3	1.8	271	11.84	3.88	-.76	1.0
Humanidades	20	125	12.79	2.98	-.15	-.23	118	12.79	2.98	-.16	-.51	114	12.79	2.98	.00	-.43

Donde: k = número de ítems; N = número de estudiantes; μ = media; σ = desviación estándar

En esta gráfica queda claro que la evaluación lineal fue adecuada para las tres versiones del EXHCOBA, con lo cual se equiparan las puntuaciones de los estudiantes. Sin embargo, también es cierto que la mejor equiparación se logró en la parte media superior de las escalas, después en la parte superior y, finalmente, la equiparación no se logró del todo en la parte inferior.

FIGURA 1. IGUALACIÓN LINEAL DE LAS VERSIONES PARA EL ÁREA DE QUÍMICA



4. CONCLUSIÓN

Como lo señala Magnusson (1993), la igualación lineal resulta efectiva si se ajustan los momentos de la media y desviación estándar en las versiones y se mantiene la misma distribución. En este estudio, la estimación de la igualación lineal de las versiones 3 y 4 con la versión 2, originó como resultado un cambio en los dos momentos (media y desviación estándar) de las versiones 3 y 4, manteniendo sin cambios la forma de la distribución (sesgo y kurtosis).

No obstante, este método ignora la variabilidad de la dificultad de la prueba en distintas regiones de la distribución, debido a que sólo ajusta los momentos de la media y desviación estándar de las distribuciones.

Lord (1982, citado en Linn, 1993) señala que es posible que dos pruebas nunca sean igualadas perfectamente. Sin embargo para que sea satisfactoria la igualación debe tomarse en consideración la construcción en las especificaciones de contenido (la prueba a ser igualada debe medir la misma habilidad, característica o función psicológica), las especificaciones estadísticas, así como la administración de la prueba, de esta forma, las versiones serán equivalentes cuando provengan del mismo dominio y produzcan distribuciones iguales en las calificaciones.

Admitir una equivalencia igual de las versiones de un examen cuando en realidad no lo sean o emplear métodos de igualación no adecuados, tendrá efectos desfavorables en la calidad de la información.

Utilizar un proceso estadístico de igualación adecuado, cambiará las calificaciones de los estudiantes para que sean más justas y equivalentes, cuando se apliquen versiones diferentes de un mismo examen, estableciendo así un sistema de evaluación mucho más sólido, donde la información que proporcione ayude en la toma de decisiones a nivel individual, institucional y político, lo cual contribuye a mejorar la calidad de los servicios evaluativos.

Los resultados de este trabajo apoyan esta afirmación, ya que ponen de manifiesto que la igualación

lineal ayuda a resolver el problema de calificar lo más justo posible a los estudiantes. Aunque, en términos globales, la equiparación lineal utilizada en este trabajo resultó adecuada, es importante señalar que actualmente se investigan funciones no lineales, que pueden resultar más apropiados y equitativos (Arce-Ferrer y Backhoff, 2002).

REFERENCIAS BIBLIOGRÁFICAS

- Aebli, H. (1989). Evaluación educativa: base de las decisiones pedagógicas. *Perfiles Educativos*, 45-46.
- AERA, APA, NCME (1999). *Standards for Educational and Psychological Testing*. Washington: American Psychological Association (APA).
- Antillón, L.E., Larrazolo, N., y Backhoff, E. (2006). Igualación equipercenil del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Electrónica de Investigación y Evaluación Educativa (RELIEVE)*. Vol 12, n.2. http://www.uv.es/RELIEVE/v12n2/RELIEVE_v12n2_2.htm.
- Arce-Ferrer, A. y Backhoff, E. (2002). Efecto del método de igualación en la equivalencia de los exámenes nacionales. *Revista Mexicana de Psicología* 19(2), pp. 101-114.
- Backhoff, E., Ibarra, M. y Rosas, M. (1995). Sistema Computarizado de Exámenes (SICODEX). *Revista Mexicana de Psicología*, 10(1), pp. 55-62.
- Backhoff, E., Larrazolo, N. y Rosas, M. (2000). Nivel de dificultad y discriminación del Examen de Habilidades y Conocimientos Básicos EXHCOBA. *Revista Electrónica de Investigación Educativa*, 2(1), pp. 1-16.
- Backhoff, E. y Tirado, F. (1992). Desarrollo del Examen de Habilidades y Conocimientos Básicos. *Revista de la Educación Superior*, 83.
- Backhoff, E. y Tirado, F. (1993). Habilidades y Conocimientos Básicos del Estudiante Universitario: hacia los estándares nacionales. *Revista de la Educación Superior*, 88, pp. 45-65.
- Backhoff, E. y Tirado, F. (1994). Estructura y lógica del Examen de Habilidades y Conocimientos Básicos. *Revista Sonorense de Psicología*, 8(1), pp. 21-33.
- Backhoff, E., Tirado, F. y Larrazolo, N. (2001). Ponderación diferencial de reactivos para mejorar la validez de una prueba de ingreso a la universidad. *Revista Electrónica de Investigación Educativa*, 3(1), pp. 1-16.
- Backhoff, E., Tirado, F., Larrazolo N. y Antillón L.E. (1996). Desigualdad de la educación básica en México. *Revista Latinoamericana de Estudios Educativos*, XXVI(3), pp. 73-88
- Hambleton, R., Swaminathan H. y Rogers H. (1991). *Fundamentals of item response theory*. Newbury Park/ London/ New Delhi: Sage publications.
- Holland, P. y Rubin, D. (1982). *Test equating*. New York: Academic Press.
- Kohler Heinz. (1999). *Estadística para Negocios y Economía*. México: CECSA.
- Kolen, M. y Brennan, R. (1995). *Test equating*. New York: Springer-Verlag.
- Kolen, M. y Brennan, R. (1987). Linear equating models for the common-item nonequivalent-populations design. *Applied Psychological Measurement*. 11(3), pp. 263-277.
- Larrazolo, R.N. (1997). Nuevas Tendencias en los Exámenes de Admisión en Universidades. *Revista de Psicología Contemporánea*, 4(2).

- Magnusson, D.. (1993). *Teoría de los tests*. México: Editorial Trillas.
- 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*, 2(3).
- Yang, Wen-Ling y Houang, R. T. (1996). *The effect of anchor length and equating method on the accuracy of test equating: Comparisons of linear and IRT-based equating using an anchor-Item*. Design. 84p. New York: Paper Presented at the Annual Meeting of the American Educational Research Association. (Eric Document Reproduction Service No. ED 401 308).

