

---

## PROCESOS DE CONTROL DE CALIDAD DEL EXAMEN DE HABILIDADES Y CONOCIMIENTOS BÁSICOS (EXHCOBA)

---

FELIPE TIRADO SEGURA / MANUEL GONZÁLEZ MONTESINOS /  
EDUARDO BACKHOFF ESCUDERO

### RESUMEN:

Se presentan resultados de un proceso de control de calidad que se aplica de manera continua al instrumento EXHCOBA que se utiliza para selección de aspirantes al ingreso en varias instituciones de educación superior del país. Se describen los procedimientos clásicos de análisis de propiedades métricas para reactivos y escalas en exámenes de aptitudes y desempeño. Los resultados del análisis clásico se comparan con los que se obtienen a través de análisis Rasch. Se destacan coincidencias y diferencias de estos enfoques así como su utilidad conjunta para la toma de decisiones sobre el monitoreo del comportamiento del instrumento.

**PALABRAS CLAVE:** Psicometría, Teoría Clásica, Teoría Respuesta al Ítem, Modelo Rasch

### INTRODUCCIÓN

El Examen de Habilidades y Conocimientos Básicos (EXHCOBA), que se aplica rutinariamente a los aspirantes a ingresar en diversas universidades del país, ha sido sometido a diversos procesos de control de calidad, de tal forma que cumpla con cabalidad los propósitos para los que fue diseñado y se puedan optimizar todas sus potencialidades.

Para ello se configuró una estructura y marco normativo bajo tres instancias colegiadas: un *Consejo Consultivo*, que tiene por función proponer y sugerir políticas generales de desarrollo del examen, que permitan atender a los intereses específicos de las instituciones que utilizan los servicios del EXHCOBA;

---

un *Comité Técnico*, cuya función es analizar y proponer las medidas técnicas requeridas para evaluar la calidad del examen en tanto su adecuada operación, así como el diseño de posibles desarrollos que permitan aprovechar las potencialidades de la evaluación; finalmente los *Comités de Área* responden a la propia estructura del *Examen*, *constituida* por cuatro áreas de conocimiento las cuales son: 1) *Matemáticas*, 2) *Lenguaje*, 3) *Ciencias Naturales* y 4) *Ciencias Sociales*. La función de los *Comités de Área* es dar cumplimiento a los acuerdos del *Comité Técnico* en cuanto los ajustes que son requeridos en los contenidos y reactivos particulares de las áreas o subáreas temáticas respectivas, para su adecuada operación y desarrollo.

En este reporte se presenta un ejercicio que se ha desarrollado por parte del *Comité Técnico* del Examen, con el propósito de revisar la calidad técnica de los reactivos del instrumento, a partir de datos empíricos y analizados bajo dos acercamientos teóricos que se basan en la estadística de la Teoría Clásica de la Medida y en la Teoría de Respuesta al Ítem, utilizando el modelo de Rasch; de tal manera que se generen comparaciones de estos indicadores en torno a reactivos específicos, para despertar la reflexión sobre los contrastes y coincidencias que se observan.

## **INSTRUMENTO DE EVALUACIÓN**

El Examen de Habilidades y Conocimientos Básicos está diseñado bajo una estructura compleja, los cuales corresponden a tres niveles: *básico*, *medio básico* y *medio superior*. La estructura también responde a cuatro áreas temáticas: *matemáticas*, *lenguaje*, *ciencias naturales* y *ciencias sociales*.

El EXHCOBA se aplica bajo un procedimiento computarizado *sui generis*, que permite generar un número indefinido de versiones, al presentar los reactivos y sus opciones de respuesta bajo un procedimiento aleatorio, de acuerdo con el nivel y las áreas temáticas correspondientes, extraídos de bancos de reactivos que constituyen sus diversas versiones. Esto le otorga al instrumento una de sus más grandes fortalezas, ya que se generan exámenes prácticamente únicos para

---

cada sustentante, dado que cambian tanto los reactivos como el orden de su presentación y de las opciones de respuesta, por lo cual es muy difícil que pueda darse el factor copia.

Para los fines del estudio que se presenta en este reporte, se seleccionó la base de datos correspondiente al banco de reactivos de una de las versiones del Examen.

Los reactivos del nivel básico comprenden 30 preguntas de *habilidades verbales* y 30 de *habilidades cuantitativas*; los del nivel medio básico se integran por 15 reactivos de *lenguaje (español)*, 15 de *matemáticas*, 20 de *ciencias naturales* y 20 *ciencias sociales*, haciendo un total de 130 reactivos analizados. El tercer segmento del Examen, el de conocimientos de especialidad, explora 60 habilidades y conocimientos correspondientes a tres áreas específicas de la formación profesional del sustentante, de manera que un aspirante que desea estudiar la carrera de ingeniería no responde los mismos reactivos temáticos que otro que aspira estudiar filosofía

## **PROCEDIMIENTOS**

### **Indicadores psicométricos clásicos**

En los indicadores psicométricos clásicos se utilizaron cuatro criterios para valorar el comportamiento de los reactivos. El primer indicador es el valor “p”, que nos indica la proporción de estudiantes que responden correctamente el reactivo, con el cual se valora si el ítem se comporta dentro de los rangos de dificultad esperados. Como ya se mencionó el *EXHCOBA* está diseñado bajo una matriz de rangos de dificultad múltiples. En el nivel *básico*, los 60 reactivos más fáciles se aceptan si son contestados correctamente entre el 30% y el 90% de los sustentantes; ya que fuera de esos rangos se estima que el reactivo es demasiado difícil o demasiado fácil, por lo que debe ser revisado o sustituido.

El segundo parámetro tiene que ver con la correlación ítem-total, es decir con la relación entre lo que mide el reactivo y lo que mide la escala. Éste es un

---

indicador de consistencia, donde se espera que los reactivos tengan una correlación con la escala de al menos 0.2. En este ejercicio las 30 preguntas de habilidades verbales y las 30 de habilidades cuantitativas fueron analizadas de manera independiente.

El tercer indicador se refiere a la capacidad que tiene el reactivo para discriminar a los estudiantes de acuerdo con su habilidad. Este parámetro se calcula comparando la proporción de aciertos, en cada reactivo, que obtienen los sustentantes con las puntuaciones más altas en la escala (Grupo superior), con relación a la proporción de reactivos que obtienen el grupo de alumnos con las calificaciones más bajas en la misma escala (Grupo inferior). El valor mínimo aceptable en este indicador es de 0.20. Debido a la distribución de las respuestas, para el caso de habilidades verbales cada grupo se conformó con 40% de los estudiantes. Así, el Grupo inferior quedó conformado por 40% de los estudiantes con las calificaciones más bajas en esta subescala, mientras que el Grupo superior se conformó con 40% de los alumnos con las calificaciones más altas. En el caso de las habilidades cuantitativas, los dos grupos quedaron conformados por, aproximadamente, 30% de estudiantes en los dos extremos de de la escala (inferior y superior).

El cuarto criterio utilizado fue el análisis de confiabilidad de cada subescala, calculado con el procedimiento de Alpha de Cronbach (Cronbach, 1990). En este procedimiento se analiza la consistencia interna de la escala eliminando uno a uno sus reactivos, con lo cual se puede conocer la aportación que hace cada uno de ellos a la confiabilidad total de la escala (Alpha sin el reactivo). Al igual que en los dos casos anteriores, las dos subescalas se analizaron de manera independiente.

### **Indicadores psicométricos Rasch**

El modelo de Rasch postula *que un reactivo eficiente sólo debe ser contestado correctamente por aquellos sustentantes que posean la habilidad requerida por las demandas cognitivas del reactivo. A la inversa, un reactivo eficiente no debe ser*

---

*contestado correctamente por aquellos sustentantes que no posean la habilidad requerida para responderlo. Bajo estas dos premisas el análisis Rasch procede primero a determinar las dificultades de los reactivos y las habilidades de los sustentantes con una escala común que se basa en una unidad de medida especial, el lógito, que se aplica a una muestra de respuestas emitidas por un grupo de  $n$  sustentantes a un conjunto de  $k$  reactivos. Bajo esta óptica, el procedimiento crea una expectativa probabilística para cada reactivo y cada sustentante.*

Posteriormente, se identifican con toda precisión aquellos reactivos y sustentantes que se comportaron de manera distinta a las expectativas probabilísticas del modelo Rasch. El proceso identifica aquellos reactivos que no fueron contestados correctamente por sustentantes que poseen la habilidad para hacerlo. Por extensión se identifican también aquellos reactivos que fueron contestados correctamente por sustentantes que no poseen la habilidad requerida para hacerlo. De igual manera el procedimiento identifica aquellos sustentantes que no contestaron correctamente a reactivos dentro de su nivel de habilidad, o que contestaron correctamente a reactivos fuera de su nivel de habilidad.

En resumen, el procedimiento identifica como anomalías todos los casos que se apartan de la expectativa del modelo Rasch. Los conjuntos de anomalías se suman y al rebasarse un límite definido de tolerancia en los índices de bondad de ajuste *INFIT* y *OUTFIT*.

La interpretación de estos índices es:

- 1) El valor de **Ajuste Interno** (*INFIT*) debe mantenerse en el rango de **.80 a 1.30**, lo que significa que el ítem **SI** fue respondido correctamente por los sustentantes que están **CERCA** del nivel de habilidad requerido por el ítem.

- 
- 2) El valor de **Ajuste Externo (OUTFIT)** debe mantenerse en el rango de **.80 a 1.30**, lo que significa que el ítem **NO** fue respondido correctamente por los sustentantes que están **LEJOS** del nivel de habilidad requerido por el ítem.
  - 3) El valor de la Correlación Punto Biserial *RPTBIS* debe ser de **.20 o mayor**, lo que indica que el **ÍTEM PERTENECE** a la escala correspondiente. Es decir el ítem mide el mismo rasgo que la escala en conjunto.
  - 4) El valor del Índice de Discriminación (**DISCR**) debe ser de **1.00**, lo que indica que el ítem **DISTINGUE CON PRECISION** a los sustentantes de alta y baja habilidad con respecto al ítem. Si el valor de *DISCR* rebasa 1.00 el ítem tiene mayor poder de distinguir con precisión entre los extremos de los niveles de habilidad de los sustentantes.

## **RESULTADOS**

Las tablas 1 y 2 presentan los indicadores psicométricos de los reactivos del EXHCOBA de las subescalas de habilidades verbales y de habilidades cuantitativas, respectivamente. En ambos casos, se presentan los resultados utilizando procedimientos de la Teoría Clásica de la Medida y del modelo Rasch.<sup>1</sup> Para cada reactivo, en negritas se señalan los parámetros que están fuera de rango, de acuerdo con la metodología correspondiente.

### **Habilidades verbales**

La tabla 1, muestra los indicadores de los 30 ítems de la subescala de habilidades verbales. Con el análisis clásico, se obtuvieron los siguientes resultados:

- Seis reactivos presentan valores “p” fuera del rango establecido: los reactivos 5, 8, 16, 19 y 25 tiene valores p por encima de 0.90, lo que

---

<sup>1</sup> Realizado con el programa Winsteps 3.60 (Linacre, 2008).

---

significa que son demasiado fáciles; mientras que el reactivo 2 resultó demasiado difícil ya que solo lo respondió el 16% de los estudiantes.

- Las correlaciones *ítem-total* total muestran que más de la mitad de los reactivos no logran cumplir con este indicador, el que casi empata con el índice de discriminación.
- Finalmente, la quinta columna muestra el cambio que sufre la confiabilidad de la subescala (Alfa sin ítem) cuando se remueve el reactivo. Se podrá apreciar que sólo en un caso este indicador, cuya medida para la escala es de 0.62, permanece igual (reactivo 26); lo que quiere decir que este reactivo no aporta elementos de consistencia a la escala, y que si se eliminara, su confiabilidad permanecería igual.

Con estos indicadores podemos tomar ciertas decisiones en términos de modificar o eliminar los reactivos que componen la escala. Por ejemplo, los reactivos 2, 5, 8, 16, 19, 25 y 26 tienen tres indicadores por abajo de los criterios establecidos; mientras que el 11, 21, 26 y 29 no cumplen con dos parámetros.

Tabla 1. Indicadores psicométricos de los ítems de habilidades verbales

Ítem	Teoría Clásica de la Medida				Modelo Rasch				
	$p$	$r$ ítem-total	Discriminación	Alpha sin ítem	MLOG	INFIT	OUFIT	RPTBIS	DISC
<b>Escala</b>	<b>0.71</b>	<b>0.17</b>	<b>0.21</b>	<b>0.62</b>	<b>0.00</b>	<b>1.00</b>	<b>.98</b>	<b>.18</b>	<b>.99</b>
1	0.40	0.20	0.27	0.60	1.62	0.99	0.99	0.20	1.03
2	<b>0.16</b>	<b>0.09</b>	<b>0.10</b>	0.61	<b>3.02</b>	1.07	1.13	<b>0.09</b>	0.93
3	0.81	0.21	0.21	0.60	-0.48	0.97	0.92	0.21	1.04
4	0.52	0.20	0.26	0.60	1.09	1.01	0.98	0.20	1.00
5	<b>0.92</b>	<b>0.18</b>	<b>0.13</b>	0.60	-1.44	0.97	0.83	<b>0.18</b>	1.03
6	0.87	0.23	0.20	0.60	-0.92	0.95	0.87	0.23	1.05
7	0.38	<b>0.19</b>	0.25	0.60	1.74	1.00	1.00	<b>0.19</b>	0.99
8	<b>0.91</b>	<b>0.18</b>	<b>0.12</b>	0.60	-1.40	0.97	0.89	<b>0.18</b>	1.02
9	0.45	0.25	0.35	0.59	1.39	0.96	0.99	0.25	1.14
10	0.60	0.24	0.35	0.59	0.70	0.97	0.97	0.24	1.10
11	0.89	<b>0.09</b>	<b>0.10</b>	0.61	-1.08	1.03	1.10	<b>0.09</b>	0.97
12	0.73	0.27	0.31	0.59	0.04	0.95	0.97	0.27	1.10
13	0.57	0.21	0.32	0.60	0.83	0.99	0.99	0.21	1.03
14	0.73	0.24	0.30	0.59	0.06	0.96	0.99	0.24	1.06
15	0.57	<b>0.12</b>	0.20	0.61	0.83	1.06	1.10	<b>0.12</b>	<b>0.71</b>
16	<b>0.93</b>	<b>0.17</b>	<b>0.11</b>	0.60	-1.61	0.97	0.84	<b>0.17</b>	1.03
17	0.72	<b>0.13</b>	0.21	0.61	0.13	1.04	1.01	<b>0.13</b>	0.93
18	0.62	0.22	0.33	0.60	0.62	0.98	1.01	0.22	1.05
19	<b>0.90</b>	<b>0.11</b>	<b>0.10</b>	0.61	-1.28	1.01	0.97	<b>0.11</b>	0.99
20	0.85	0.23	0.23	0.60	-0.72	0.96	0.85	0.23	1.05
21	0.88	<b>0.15</b>	<b>0.14</b>	0.60	-1.07	1.00	0.94	<b>0.15</b>	1.01
22	0.80	<b>0.12</b>	0.20	0.61	-0.35	1.03	1.10	<b>0.12</b>	0.95
23	0.78	<b>0.19</b>	0.25	0.60	-0.23	0.99	0.98	<b>0.19</b>	1.02
24	0.80	0.27	0.28	0.59	-0.35	0.94	0.89	0.27	1.09
25	<b>0.95</b>	<b>0.16</b>	<b>0.07</b>	0.60	-2.12	0.97	0.83	<b>0.16</b>	1.02
26	0.67	<b>0.03</b>	<b>0.13</b>	<b>0.62</b>	0.38	1.11	1.11	<b>0.03</b>	<b>0.70</b>
27	0.66	<b>0.14</b>	0.24	0.60	0.42	1.03	1.03	<b>0.14</b>	0.91
28	0.65	<b>0.18</b>	0.27	0.60	0.47	1.01	1.00	<b>0.18</b>	0.97
29	0.84	<b>0.14</b>	<b>0.13</b>	0.60	-0.67	1.01	1.04	<b>0.14</b>	0.98
30	0.67	<b>0.12</b>	0.19	0.61	0.37	1.05	1.08	<b>0.12</b>	0.84

Nota: en negritas se señalan los parámetros fuera de rango

Ahora bien, el análisis de la modelización Rasch de los reactivos de habilidades verbales se presenta en la tabla 1. Como se podrá apreciar sólo el reactivo 2 tiene una dificultad excesiva, mayor a 3 lógitos. En las columnas de ajuste interno (INFIT) y externo (OUFIT), los 30 reactivos de la escala de habilidades verbales están dentro de las expectativas del modelo Rasch ya que todos sus valores están dentro de los rangos de bondad de ajuste requeridos. Las implicaciones inmediatas de estos resultados son:

---

Los reactivos fueron respondidos correctamente por los sustentantes que fueron ubicados cerca del nivel de habilidad cognitiva requerido por cada reactivo.

- Los reactivos no fueron respondidos correctamente por los sustentantes que fueron ubicados lejos del nivel de habilidad cognitiva requerido por cada reactivo.
- La escala es unidimensional y como tal se centra en el constructo de interés para el cual fue diseñada la escala (*habilidades verbales*).
- El comportamiento empírico de la escala modelada con el Análisis Rasch proporciona una pieza importante de evidencia apoyar la conclusión de que la escala mide la variable latente de interés.

Por lo que respecta al indicador de RPTBIS los valores obtenidos en este ejemplo son los mismos que los que se obtienen bajo el análisis de la Teoría Clásica. Se señalan en negritas los reactivos que se apartan del límite mínimo de .20.

En cuanto a indicador de DISCR, se han señalado los reactivos 15 y 26 con valores de .71 y .70, ya que estos dos casos se apartan del valor esperado de 1.0.

### **Habilidades cuantitativas**

Los mismos análisis se realizaron para la subescala de habilidades cuantitativas, cuyos resultados se muestra en la tabla 2. Desde la óptica clásica, los reactivos de esta sección tienen mejores indicadores que los de la sección anterior, ya que sólo tres de ellos están por arriba del nivel de dificultad aceptado (reactivos 8, 11 y 29); sólo uno presenta una correlación ítem-total por debajo de 0.2 (reactivo 29); dos de ellos presentan una discriminación baja (reactivos 8 y 29) y cuatro de ellos no contribuyen a su confiabilidad (1, 8, 28 y 29). Los reactivos 8 y 29, que presentan problemas en cuando menos tres indicadores, deberán ser modificados o sustituidos para mejorar la escala.

Tabla 2. Indicadores psicométricos de los ítems de habilidades cuantitativas

Ítem	Teoría Clásica de la Medida				Modelo Rasch				
	<i>p</i>	<i>r</i> ítem-total	Discri- minación	Alpha sin ítem	MLOG	INFIT	OUTFIT	RPTBIS	DISC
<b>Escala</b>	<b>0.71</b>	<b>0.32</b>	<b>0.41</b>	<b>0.82</b>	<b>0.00</b>	<b>1.00</b>	<b>.98</b>	<b>.32</b>	<b>.99</b>
31	0.86	0.22	0.24	<b>0.82</b>	-0.92	1.05	1.03	0.22	0.95
32	0.87	0.31	0.30	0.81	-1.10	0.96	0.80	0.31	1.05
33	0.47	0.33	0.48	0.81	1.42	1.04	1.07	0.33	<b>0.88</b>
34	0.83	0.24	0.27	0.81	-0.66	1.04	1.11	0.24	0.94
35	0.65	0.32	0.48	0.81	0.43	1.03	1.03	0.32	0.93
36	0.85	0.30	0.31	0.81	-0.82	0.97	1.02	0.30	1.02
37	0.60	0.32	0.47	0.81	0.74	1.06	1.10	0.32	<b>0.84</b>
38	<b>0.97</b>	0.20	<b>0.08</b>	<b>0.82</b>	-2.71	0.97	<b>0.72</b>	0.20	1.03
39	0.47	0.48	0.66	0.80	1.42	0.88	0.87	0.48	1.30
40	0.76	0.35	0.46	0.81	-0.19	0.97	0.93	0.35	1.05
41	<b>0.91</b>	0.28	0.20	0.81	-1.54	0.96	<b>0.76</b>	0.28	1.04
42	0.72	0.38	0.50	0.81	0.05	0.96	0.94	0.38	1.07
43	0.72	0.46	0.61	0.81	0.06	0.88	<b>0.75</b>	0.46	1.25
44	0.74	0.43	0.55	0.81	-0.09	0.90	0.80	0.43	1.18
45	0.86	0.33	0.33	0.81	-0.93	0.96	<b>0.79</b>	0.33	1.06
46	0.73	0.32	0.41	0.81	0.01	1.01	1.09	0.32	0.97
47	0.61	0.45	0.62	0.81	0.69	0.91	0.86	0.45	1.24
48	0.51	0.31	0.46	0.81	1.20	1.08	1.11	0.31	<b>0.78</b>
49	0.35	0.39	0.51	0.81	2.05	0.97	0.94	0.39	1.06
50	0.76	0.25	0.30	0.81	-0.17	1.08	1.09	0.25	<b>0.88</b>
51	0.64	0.38	0.56	0.81	0.51	0.98	0.99	0.38	1.05
52	0.55	0.25	0.39	0.81	0.96	1.13	1.18	0.25	<b>0.63</b>
53	0.77	0.32	0.40	0.81	-0.25	1.00	1.00	0.32	1.00
54	0.83	0.25	0.29	0.81	-0.66	1.03	1.12	0.25	0.95
55	0.70	0.30	0.42	0.81	0.18	1.05	1.08	0.30	0.91
56	0.72	0.29	0.39	0.81	0.05	1.05	1.13	0.29	0.90
57	0.80	0.43	0.48	0.81	-0.44	0.88	<b>0.74</b>	0.43	1.17
58	0.44	0.23	0.35	<b>0.82</b>	1.55	1.15	1.25	0.23	<b>0.61</b>
59	<b>0.92</b>	<b>0.19</b>	<b>0.14</b>	<b>0.82</b>	-1.64	1.00	1.25	<b>0.19</b>	0.98
60	0.59	0.38	0.56	0.81	0.78	0.99	0.95	0.38	1.05

Nota: en negritas se señalan los parámetros fuera de rango.

Desde la perspectiva del modelo Rasch, puede observarse que la mayoría de los reactivos se mantienen dentro de los límites de bondad de ajuste. Pero se puede apreciar que bajo el criterio de OUTFIT los reactivos: 38, 41, 43, 45 y 57 presentan valores inferiores a .80. Los valores por debajo de este límite indican que los reactivos están “atenuados”. Esto significa que estos reactivos no son lo suficientemente sensibles para captar la información de los respondentes que

---

están lejos del nivel de habilidad requerido para responderlos correctamente. Estos reactivos deben ser separados para un análisis puntual y posterior modificación o sustitución. Por lo demás, bajo el criterio de INFIT, se confirma que los 30 reactivos de esta subescala son lo suficientemente sensibles para captar la información de los respondentes que están cerca del nivel de habilidad requerido para responderlos correctamente.

Por lo que respecta al indicador de RPTBIS se observa que los valores obtenidos coinciden con los respectivos del análisis clásico. Bajo ambos análisis en esta subescala no se detectaron valores inferiores al límite de .20.

En cuanto al indicador de DISCR se señalan los reactivos: 33, 37, 48, 50 y 58 que se apartan del valor esperado de 1.0.

## **CONCLUSIONES**

Al hacer la comparaciones entre los indicadores clásicos de psicometría y los del análisis Rasch, se encontró que estos convergen en un criterio: la correlación punto-biserial. Éste es el criterio ancla que identifica los mismos reactivos en ambas escalas. Pero la información derivada bajo el método Rasch sobre los reactivos y escalas es distinta en contenido e impacto a la que se obtiene mediante los métodos clásicos. Esto puede apreciarse a partir de las medidas de dificultad, el error estándar asociado a cada medida, el índice de correlación del reactivo con el puntaje total (PTBIS) y el índice de discriminación (DISCR).

La información de las tablas 1 y 2 revela que ambos procedimientos generan información distinta pero complementaria:

- Los indicadores de análisis clásico aportan información útil sobre la composición y consistencia interna de las escalas del EXHCOBA.
- Desde el análisis clásico se obtienen determinaciones puntuales sobre la consistencia interna de las escalas y sobre la contribución específica cada reactivo a esta propiedad. Esta aproximación confirma cuando los

---

reactivos trabajan en la misma dirección para medir el constructo común, lo que constituye la propiedad más básica o fundamental de toda escala o instrumento de medición.

- Desde el análisis Rasch se obtienen determinaciones sobre las escalas en una unidad métrica aplicable a las dificultades de los reactivos y las habilidades de los sustentantes. A partir de las comparaciones entre estos parámetros empíricamente derivados, se establecen criterios probabilísticos sólidamente fundados, y que funcionan como estándares de calidad que cada reactivo debe satisfacer como elemento constituyente de una escala unidimensional.

Aunque el instrumento *EXHCOBA* es ya un examen consolidado, es indispensable que todas sus versiones continúen siendo examinadas bajo los procedimientos y criterios de control de calidad de la *Teoría Clásica y del Modelo Rasch*, ya que las características de los sustentantes van variando con el tiempo. Estas medidas constituyen una evaluación continua del instrumento. Las principales ventajas de esta meta-evaluación son:

- 1) Permite identificar los reactivos que poseen las propiedades métricas deseables y los reactivos que no concuerdan con el diseño original de las escalas.
- 2) Permite acumular evidencia para establecer la unidimensionalidad de las escalas y la validez de constructo para cada escala y para el instrumento en conjunto a partir de todas sus versiones en uso.

---

## REFERENCIAS

- Backhoff, E. y Tirado, F. (1992). "Desarrollo del Examen de Habilidades y Conocimientos Básicos", en: *Revista de la Educación Superior* (México: Asociación Nacional de Universidades e Instituciones de Educación Superior), vol. XXI, núm. 3 (83), pp. 95-118.
- Backhoff, E. y Tirado, F. (1993). "Habilidades y conocimientos básicos del estudiante universitario: hacia los estándares nacionales", *Revista de Educación Superior* (México: Asociación Nacional de Universidades e Instituciones de Educación Superior), vol. XXI, núm. 3 (88), pp. 45-65.
- Backhoff, E. y Tirado, F. (1994). "Estructura y lógica del Examen de Habilidades y Conocimientos Básicos", *Revista Sonorense de Psicología*, vol. 8, núm. 1, pp. 21-33.
- Cronbach, L.J. (1990). *Essentials of Psychological Testing*, Nueva York: Harper Collins Publishers, Inc.
- González, M. M. (2008). *El análisis de reactivos a través del modelo de Rasch*. Publicación electrónica: <http://www.winsteps.com/recursos>
- Linacre, J. M. (1991- 2008). *Rasch-Model Computer Programs* ©, Chicago, IL, publicación electrónica: [www.winsteps.com](http://www.winsteps.com)
- Linacre, J. M. (2006). *A User's Guide to Winsteps Ministeps: Rasch-Model Computer Programs*, Chicago, IL, publicación electrónica: [www.winsteps.com](http://www.winsteps.com)
- Tirado, F.; Backhoff, E.; Larrazolo, N. y Rosas, M. (1997). "Validez predictiva del Examen de Habilidades y Conocimientos Básicos (EXHCOBA)", en: *Revista Mexicana de Investigación Educativa*, vol. II, núm. 3.