



Bogotá, Marzo 29 de 2016

Ingeniero

JESUS EVELIO ORTEGA AREVALO

Coordinador Contrato Consultoría N° 231

Universidad de Pamplona

En atención a su solicitud del día de ayer, nos permitimos reiterar lo expuesto en varias oportunidades tanto en los manuales técnicos de las pruebas como por este medio en relación con el proceso de validación de las pruebas escritas de la convocatoria 022 de 2013 del Consejo Superior de la Judicatura:

Con el objeto de regular el proceso de construcción de pruebas y garantizar un uso adecuado de sus resultados, se han desarrollado estándares que prescriben lo que se considera una buena práctica en la elaboración y el uso de instrumentos de evaluación, los cuales han sido adoptados y desarrollados por instituciones que en Colombia apoyan los procesos masivos de selección para la provisión de empleos públicos.

La construcción de instrumentos de evaluación requiere de procedimientos sistemáticos y rigurosos. En las pruebas con altas consecuencias, como es el caso de los exámenes de selección para el acceso a empleos públicos, estas exigencias cobran aun mayor relevancia. El proceso se inicia con la identificación del propósito de la evaluación y finaliza una vez que se demuestra que la prueba sirve para los objetivos para los cuales fue diseñada.

Diferentes estándares internacionales sobre la construcción de pruebas definen una serie de pasos que deben respetarse al elaborar instrumentos de evaluación que implican consecuencias serias para los individuos. Respetarlos da garantías a los usuarios de que se someterán a un proceso justo y confiable.

Uno de estos pasos busca determinar las propiedades estadísticas de los ítems que permitan y demostrar su poder predictivo, es decir, que se observa correlación entre el puntaje obtenido y el rendimiento del aspirante en el empleo al que aspira y, cuando corresponda, eliminar los ítems que no cumplen con los criterios preestablecidos.

Adicionalmente, en las pruebas de altas consecuencias como estas, se garantiza que los resultados sean confiables y no producto del acceso fraudulento a las preguntas. En este sentido, cobra importancia el cuidado con las filtraciones de preguntas; por ello, el proceso de análisis y validación de ítems se debe realizar en la misma aplicación de las pruebas y no antes.



Este análisis de ítems es un proceso cuantitativo y cualitativo mediante el cual se establece la calidad de los ítems de un instrumento, en relación con los propósitos para los cuales ellos fueron elaborados. Su realización implica un saber profundo sobre el objeto de evaluación, la población evaluada, los propósitos de la evaluación y se requiere, además, conocer debidamente las técnicas de procesamiento de datos para hacer una adecuada interpretación de los indicadores estadísticos disponibles. El proceso de análisis de ítems debe conducir a la toma de decisiones en relación con la inclusión, exclusión, o modificación de ítems, a partir de la identificación clara de las posibles problemáticas de los mismos.”¹

Cabe aclarar que esta eliminación es un procedimiento técnico usado a partir de los resultados del análisis de ítems, con el cual se busca mejorar la calidad de la prueba en cuanto a confiabilidad sin disminuir la cantidad de reactivos necesaria para evaluar un constructo o contenido. De esta manera, al identificar y eliminar los elementos que aumentan el error estadístico de medida de todo el instrumento, mejora la confiabilidad del mismo y permite evaluar a los examinados con los ítems que permiten detectar a los mejores aspirantes.

Así se informó en el Manual Técnico de la Prueba de Conocimientos, Fase II: Análisis Estadístico y Estandarización de las Pruebas, pagina 30:

“2. EXCLUSIÓN DE ÍTEMS

Las pruebas fueron previamente analizadas por el equipo encargado del diseño y construcción de las pruebas, teniendo en cuenta los conocimientos, habilidades y características requeridas por los aspirantes en este proceso de selección. Cada uno de los ítems superó al menos dos validaciones de jueces expertos y fueron analizados psicométricamente, de tal manera que los ítems que presentaron indicadores inferiores al standard adoptado, fueron eliminados de la calificación.”

Los standards adoptados fueron:

- Dificultad del ítem.
- Índices de discriminación negativos o cercanos a cero, es decir, que no eran respondidos por prácticamente ninguna persona.

Dificultad del Ítem

Este Indicador permite apreciaciones directas a partir de la lectura del número que la expresa, el cual va de cero (0) a uno (1). El índice de dificultad hace referencia al grado

¹ Pardo, Carlos; Rocha, Martha; Avendaño, Bertha; Barrera, Libardo. MANUAL DE PROCESAMIENTO DE DATOS Y ANÁLISIS DE ÍTEMS. Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (Llece). 2005.



en el cual una población la responde correctamente. Por esto se define a partir de una población de examinados y corresponde a la proporción de individuos que responde correctamente el ítem. Las inferencias se hacen sobre los estimados de dificultad de la pregunta a los parámetros de dificultad para una población bien definida. La inferencia es a un grupo de personas.

Se calcula por medio de:

$$P_i = \frac{\sum U_i}{N}$$

Donde

P_i = índice de dificultad de la pregunta

U_i = respuestas a la pregunta

Si la respuesta es correcta, **U_i** = 1

Si la respuesta es incorrecta, **U_i** = 0

N = total de personas que abordan la pregunta

El índice de dificultad es mayor si la muestra tiene habilidad mayor que el promedio de habilidad en la población.

Interpretación: los valores de dificultad oscilan entre menos infinito y más infinito en la escala logit, aunque en términos prácticos los ítems asumen valores entre -3.0 y +3.0, cuando el promedio de dificultades del grupo de ítems se centra en cero.

Valores positivos y altos indican alta dificultad y los valores negativos indican baja dificultad.

Criterio de aceptación: regularmente se analiza la distribución de valores de dificultad del instrumento en relación con los valores de habilidad de la población evaluada para conceptuar sobre lo apropiado de la medición de dicho instrumento, de acuerdo con los propósitos que lo inspiraron. Un aspecto importante de análisis está dado por la densidad de ítems en un punto de la escala de habilidad en particular; así, se espera que no haya más de dos ítems de un mismo componente o contenido que midan con la misma dificultad. El criterio de aceptación se fijó en +0,10 o superior.



Discriminación del Ítem

Es, junto con la dificultad, parámetro fundamental de los Ítems dentro del modelo de análisis. Indica en qué grado el ítem es respondido correctamente por las personas de alta habilidad e incorrectamente por las personas de baja habilidad.

Es la varianza, que en una variable dicótoma se puede expresar en términos de la proporción de personas que responde incorrectamente una pregunta. Se calcula a partir de:

$$S_i^2 = p_i q_i$$

Donde:

p_i = índice de dificultad de la pregunta

q_i = 1 - p_i

Es claro que la magnitud de la varianza es determinada por cualquiera de los dos valores p ó q ; también lo es que el valor máximo es de .25 que ocurre cuando p y q valen 0.5 cada una, y este valor disminuye a medida que p ó q se desvían de este punto. Al producirse una varianza grande, la incertidumbre respecto a la puntuación verdadera de cualquier persona es mayor.

Como depende tan estrechamente de la proporción de respuestas correctas, si ésta se afecta por la población, la discriminación lo hace en el mismo sentido.

Interpretación: los valores de discriminación oscilan, teóricamente, entre menos infinito y más infinito, aunque, por lo general, los ítems presentan valores de discriminación entre -2 y $+2$. Valores que se aproximan a más infinito se corresponden con un patrón de Guttman (discriminación perfecta).

Criterio de aceptación: Se estableció como aceptables los ítems con valores de discriminación superiores o iguales a 0.10.

Así las cosas, cuando una pregunta presenta índices de dificultad o de discriminación bajos o negativos, la validez y la confiabilidad de la prueba se afectan. Esto implica que es necesario dejar dentro de la evaluación aquellas preguntas que más aportan a la medición adecuada del conocimiento, atributo o cualidad que se pretende medir, y para esto deben dar cuenta de unos niveles de dificultad y discriminación aceptables (mínimo ,10, respectivamente). En este caso se retiran las mismas preguntas para todos los aspirantes poder afinar el instrumento de evaluación y tener unos parámetros de

medición ajustados al grupo, sin afectar a ninguno de ellos ya que son tratados de la misma manera.

Los índices obtenidos por los ítems que fueron eliminados en cada uno de los 14 cuadernillos de las pruebas aplicadas el 07 de diciembre de 2014 en la Convocatoria N° 22, Acuerdo PSAA13-9939 de 2013 fueron:

1. Índices de dificultad y de discriminación para las preguntas eliminadas de los cuadernillos en el Componente Común:

Pregunta	Índice de dificultad	Índice de discriminación
4 *	0.11	-0.09
11	0.10	0.10
14	0.16	0.02
16	0.04	-0.29
22	0.10	0.08
42	0.10	-0.07

* Esta pregunta solo se excluyó de la calificación en el grupo 4 (cuadernillos 7 a 9).

2. Índices de dificultad y de discriminación para las preguntas eliminadas en la prueba específica:

Prueba - Cuadernillo N° 1. No se eliminaron preguntas del Componente Específico.

Prueba - Cuadernillo N° 2. Se eliminaron las preguntas 55 y 92, con los siguientes índices:

Pregunta	Índice de dificultad	Índice de discriminación
55	0.14	-0.05
96	0.11	-0.23

Prueba - Cuadernillo N° 3. Se eliminaron las preguntas 83 y 87, con los siguientes índices:

Pregunta	Índice de dificultad	Índice de discriminación
83	0.10	-0.23
87	0.07	-0.08

Prueba - Cuadernillo N° 4. Se eliminaron las preguntas 62, 65 y 86, con los siguientes índices:

Pregunta	Índice de dificultad	Índice de discriminación
62	0.08	-0.06
65	0.07	-0.08
86	0.05	0.02

Prueba - Cuadernillo N° 5. Se eliminaron las preguntas 65 y 94, con los siguientes índices:

Pregunta	Índice de dificultad	Índice de discriminación
65	0.009	0.3
94	0.03	0.09

Prueba - Cuadernillo N° 6. Se eliminaron las preguntas 57 y 80, con los siguientes índices:

Pregunta	Índice de dificultad	Índice de discriminación
57	0.03	-0.13
80	0.10	0.02

Prueba - Cuadernillo N° 7. Se eliminaron las preguntas 52 y 58, con los siguientes índices:

Pregunta	Índice de dificultad	Índice de discriminación
52	0.02	- 0.18
58	0.10	0.3

Prueba - Cuadernillo N° 8. Se eliminaron las preguntas 82 y 95, con los siguientes índices:

Pregunta	Índice de dificultad	Índice de discriminación
82	0.08	- 0.03
95	0.11	- 0.2

Prueba - Cuadernillo N° 9. Se eliminaron las preguntas 62 y 63, con los siguientes índices:

Pregunta	Índice de dificultad	Índice de discriminación
62	0.06	- 0.07
63	0.04	- 0.20

Prueba - Cuadernillo N° 10. Se eliminaron las preguntas 70 y 77, con los siguientes índices:

Pregunta	Índice de dificultad	Índice de discriminación
70	0.06	- 0.3
77	0.03	- 0.17

Prueba - Cuadernillo N° 11. Se eliminaron las preguntas 52, 74, 82, 86 y 95, con los siguientes índices:



Pregunta	Índice de dificultad	Índice de discriminación
52	0.05	-0.17
74	0.09	0.03
82	0.02	-0.58
86	0.08	-0.24
95	0.03	-0.39

Prueba - Cuadernillo N° 12. No se eliminaron preguntas del Componente Específico.

Prueba - Cuadernillo N° 13. Se eliminaron las preguntas 61 y 82, con los siguientes índices:

Pregunta	Índice de dificultad	Índice de discriminación
61	0.06	-0.05
82	0.06	-0.05

Prueba - Cuadernillo N° 14. Se eliminaron las preguntas 68 y 70, con los siguientes índices:

Pregunta	Índice de dificultad	Índice de discriminación
68	0.09	0.3
70	0.02	-0.73

Atentamente,

DIEGO LEÓN REYES
División Técnica
Alpha Gestión SAS