

SABER

AL > DETALLE



GOBIERNO DE COLOMBIA

EDICIÓN

02

Bogotá D.C.

Octubre de 2018

ISSN: En trámite

Publicación trimestral

Instituto Colombiano para la Evaluación de la Educación, ICFES

Oficinas: Calle 26 No. 69-76 Torre 2, piso 15

Edificio Elemento, Bogotá • Colombia

Directora General: María Figueroa Cahnspeyer

Directora de Evaluación: Natalia González Gómez

Subdirectora de Análisis y Divulgación: Ana María Restrepo Sáenz

Subdirección de Estadísticas: Natalia González Gómez (E)

Subdirección de Diseño de Instrumentos: Javier Toro

Coordinación General: Dirección de Evaluación

¿Qué diseño del armado se emplea en el Icfes para medir las pruebas Saber?

¿Qué diseño del armado se emplea en el Icfes para medir las pruebas Saber?

Al realizar evaluaciones estandarizadas de gran escala se busca recolectar información sobre todas las preguntas (a lo que llamamos ítems) que componen una prueba, con el fin de estimar la habilidad de los evaluados. Sin embargo, que cada evaluado resuelva la totalidad de los ítems que componen una prueba Saber resulta complejo, al existir una gran cantidad de ítems construidos para cada prueba y una restricción de tiempo para la presentación de la misma. Con esto en mente, es necesario (a) emplear una metodología que permita evaluar todo el rango de dificultad de una prueba contemplando las dos restricciones mencionadas, (b) garantizar la comparabilidad entre los resultados y (c) optimizar la forma como se presentan los ítems en cada caso. Una aproximación para lograr lo anterior consiste en construir diferentes formas de medición de cada prueba en las cuales se configuren combinaciones particulares de bloques, lo cual se expondrá más adelante.

Para la aplicación de las pruebas Saber 359, Saber 11, Saber TyT y Saber PRO se emplea un diseño para la configuración de las formas de medición conocido como Bloques Incompletos Balanceados (BIBs). Este consiste en construir grupos de ítems con características particulares (que llamamos bloques) y agrupar conjuntos de bloques para usarlos como formas de medición, siempre que respeten las especificaciones¹ de la prueba. Por lo cual el diseño resulta un proceso complejo. El objetivo del diseño en BIBs es recolectar datos de manera eficiente, de tal manera que cada prueba Saber sea una herramienta válida y confiable para evaluar las competencias por áreas del conocimiento. La ventaja que tiene esta metodología es que, al construir muchas formas de medición, disminuye la posibilidad de cometer fraude. Así mismo, es posible pilotear ítems en las formas para emplearlos en próximas aplicaciones de las pruebas Saber.

1. ¿Por qué es importante pensar en un diseño en BIBs para configurar las pruebas Saber?

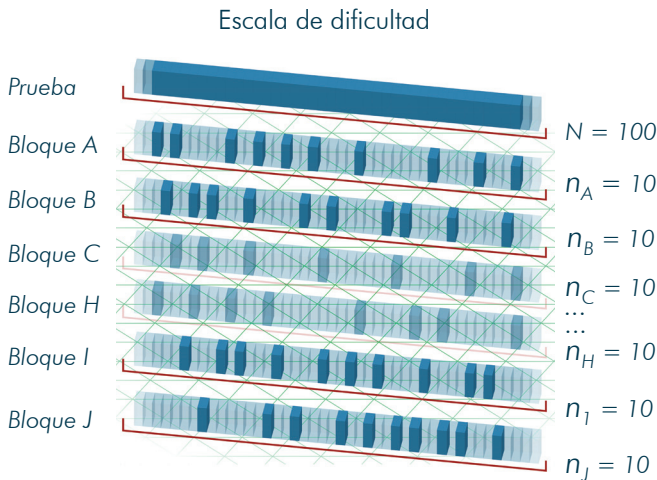
Cuando se emplea un diseño completo en una prueba se presentan todos los ítems a los evaluados en un solo bloque para compararlos entre sí. Sin embargo, un diseño de este tipo resulta ineficiente cuando una prueba está compuesta por un número grande de ítems, ya que cada bloque tiene un tamaño limitado. Resulta más eficiente dividir los ítems en varios bloques donde se comparen un mismo número de veces para que no sea necesario que los evaluados respondan todos los ítems de una prueba: esta es la dinámica del diseño en BIBs. Tal diseño permite evaluar un gran número de ítems y maximizar la información recolectada en una prueba, mientras se reduce tanto el número de ítems que presenta cada evaluado como el número de veces que se repiten los ítems.

Una de las grandes ventajas que tiene el diseño en BIBs es que permite evaluar todo el rango de dificultad de una prueba en una población objetivo, como por ejemplo establecimientos educativos, con el fin de obtener información sobre varios niveles de habilidad. Para ilustrar este enfoque, se presenta en la figura 1 una prueba

hipotética compuesta por 100 ítems agrupada en 10 bloques {A, B, C, D, E, F, G, H, I, J}, en el que cada bloque contiene 10 ítems. Cada prueba Saber tiene una cantidad definida de bloques que se mantiene entre aplicaciones. Sin embargo, la cantidad de bloques es diferente entre pruebas. Así, en la prueba Saber 359 son 15 bloques, en la prueba Saber 11 son 8 bloques, y en las pruebas genéricas de Saber TyT y Saber PRO el máximo 4 bloques cada una. Se puede apreciar que cada bloque comprende ítems con diferentes niveles de dificultad, y al interior de cada bloque los niveles de dificultad son dispersos. Agrupar distintos bloques (a los que llamamos formas) permite cubrir diferentes niveles de habilidad. De esta manera, tener distintas formas de medición para una prueba Saber permite contemplar el espectro completo de la escala de dificultad sin necesidad de responder todos los ítems, y como la escala de dificultad y habilidad es la misma, se estima todo el rango de habilidades de los evaluados. Como se expondrá más adelante, el diseño en BIBs de las pruebas Saber hace que el armado de las formas sea un proceso de optimización que tiene en cuenta muchas variables simultáneamente.

¹ En el caso de las pruebas Saber, las competencias se configuran a partir del Modelo Basado en Evidencia (MBE) que brinda los lineamientos sobre las especificaciones de la prueba. El modelo MBE indica que cada competencia que se pretenda evaluar requiere de ciertas afirmaciones que detallan alguna capacidad, habilidad o conocimiento, y cada prueba evalúa niveles particulares del ciclo de aprendizajes. Se encuentra un análisis detallado en <http://www.icfes.gov.co/instituciones-educativas-y-secretarias/acerca-de-las-evaluaciones/como-se-elaboran-las-pruebas>

Figura 1. Escala de dificultad o habilidad por bloques de una prueba



Fuente: ICFES, 2017

2. ¿En qué consiste el diseño en BIBs?

El nombre de la metodología proviene de un diseño experimental con el mismo nombre en el que se busca recolectar datos sobre unidades experimentales o tratamientos y realizar comparaciones entre sí. A raíz de este diseño experimental, hay un hallazgo importante alrededor de la rotación de bloques que evidencia la comparabilidad entre formas. La innovación del diseño experimental aplicado a evaluaciones de desempeño es la organización de tratamientos en bloques de ítems para generar dichas comparaciones, de tal manera que no es necesario tener un mismo número de tratamientos que de bloques, sino que el número de tratamientos pueda ser mayor o igual que el número de bloques. Que el diseño sea balanceado implica que cada bloque y que cada pareja de bloques aparecen el mismo número de veces a lo largo de las formas de medición/tratamientos. Así, la metodología de diseño en BIBs sigue estos lineamientos y emplea ítems como objeto de análisis, en el que se administran conjuntos de bloques a los evaluados. Bajo el enfoque de diseño experimental, se busca optimizar el número de agrupaciones de tratamientos, mientras que bajo la metodología en BIBs se maximizan las agrupaciones de ítems.

Siguiendo el ejemplo de la prueba compuesta por 100 ítems descrito anteriormente, se presenta en la figura 2 un diseño en BIBs, donde los ítems son de respuesta múltiple con única respuesta. Se generan diez subgrupos de 10 ítems por bloque, de tal manera que todos los ítems de la aplicación se

agrupan en 10 bloques $\{A, B, C, D, E, F, G, H, I, J\}$. Teniendo en cuenta que cada evaluado presenta una forma compuesta por 3 bloques, se estructuran 30 formas posibles para este diseño. Aplicando la metodología en BIBs, cada uno de los 10 bloques aparece 9 veces al interior de las formas de la prueba y cada par de bloques aparece solo dos veces en las 30 formas presentadas. Nótese que cada Bloque A, por ejemplo, aparece en las formas 1, 5, 7, 10, 13, 15, 16, 25 y 26 y el par (A, B) se repite en las formas 1 y 5.

De esta manera, al diseñar diferentes formas se obtienen respuestas individuales de todos los grupos de bloques y por tanto de cada uno de los ítems, abarcando todos los niveles de dificultad de la prueba. Al estimar las características de los ítems (a lo que llamamos parámetros), es posible calcular la habilidad de cada evaluado a través de la estimación un modelo de calificación de teoría respuesta al ítem (TRI). En particular, un modelo logístico de tres parámetros (3PL).

Figura 2. Diseño en BIBs

	Bloque 1	Bloque 2	Bloque 3
BLOQUES			
A			
B			
C			
D			
E			
F			
G			
H			
I			
J			
Forma 1	A	D	B
Forma 2	E	H	D
Forma 3	I	C	B
Forma 4	B	H	E
Forma 5	F	B	A
Forma 6	B	G	F
Forma 7	C	D	A
Forma 8	H	B	I
Forma 9	C	E	D
Forma 10	A	E	I
Forma 11	D	I	F
Forma 12	J	C	B
Forma 13	F	A	E
Forma 14	D	I	J
Forma 15	A	C	G
Forma 16	J	A	H
Forma 17	F	J	C
Forma 18	E	J	G
Forma 19	G	H	D
Forma 20	B	D	G
Forma 21	G	E	F
Forma 22	E	I	C
Forma 23	H	F	C
Forma 24	C	G	H
Forma 25	H	A	J
Forma 26	I	G	A
Forma 27	D	F	J
Forma 28	J	B	E
Forma 29	I	F	H
Forma 30	G	J	I

Fuente: ICFES, 2017

En el diseño en BIBs se consideran una serie de parámetros que permiten dar solución al problema de diseño, teniendo en cuenta que es un sistema de variables sujeta a restricciones. El primer parámetro consiste en el número total de ítems (N). El segundo, está asociado a los bloques (j) y corresponde al número de subgrupos de ítems. El tercer y cuarto parámetro es el número de ítems por bloque (n_j) y el número de bloques que compone una forma, (k), respectivamente. El quinto y sexto se denotan por ($\lambda=2$) y (r) y corresponden al número de veces que aparece un par de bloques dentro de cada forma del diseño, y al número de veces que se repite un bloque en el diseño. Cabe señalar que a veces no existen soluciones para algunos diseños en BIBs, por lo cual es necesario verificar cada uno de los parámetros en el diseño.

Con el diseño en BIBs es posible hablar de formas de medición equivalentes, ya que se asignan ítems a los evaluados optimizando una característica del ítem relevante

para la evaluación, como lo es la habilidad de los estudiantes. Esto permite que se logre calcular la distribución de habilidades de los evaluados, que es el parámetro de interés. En este sentido se busca que los bloques tengan dificultades promedio similares entre sí y que al interior de cada bloque los ítems tengan dificultades muy diferentes, con el fin de poder evaluar niveles de la habilidad muy bajas y muy altas.

Teniendo en cuenta la definición de cada parámetro, para el ejemplo de la figura 2 se considera un número total de ítems ($N=100$), los cuales se reparten en ($j=10$) bloques $\{A,B,C,D,E,F,G,H,I,J\}$ donde cada bloque contiene ($n_j=10$) número de ítems. En este diseño cada bloque se repite ($r=9$) veces y cada par de bloques aparece ($\lambda=2$) veces. Teniendo en cuenta que hay treinta formas, es decir 30 agrupaciones de bloques, cada una de ellas está compuesta por ($k=3$) bloques (lo que se conoce como una triada).

3. ¿Cómo se construyen los bloques?

Para que un diseño sea considerado en BIBs se deben tener en cuenta restricciones estructurales y restricciones prácticas. Por un lado, las restricciones estructurales hacen referencia al número de bloques que se asigna por forma, así como un límite en el número de formas diseñadas para la aplicación de la prueba y las combinaciones de bloques asignadas por formas. Por otra parte, las restricciones funcionales se centran

en las particularidades de los ítems, tales como contenido, formato y nivel cognitivo, así como el tiempo esperado de respuesta por bloque y las características psicométricas de los ítems. En cada una de las formas, se procura mantener una ponderación equiparable por contenido, teniendo en cuenta que los ítems que conformen una prueba sean independientes entre sí y no incluyan la clave de respuestas entre sí.

Bibliografía

- Craig, B. A.** (2018). Purdue university statistics. <http://www.stat.purdue.edu/~bacraig/notes1/topic13/>.
- Fariah, Y., A., R., and Munir, A.** (2015). Construction of balanced incomplete block designs using cyclic shifts. *Designs Using Cyclic Shifts and Computation*, 44(2):525-532.
- ICFES** (2017a). Implementación de un diseño de bloques incompletos balanceados en las pruebas genéricas de saber 11°. *Dirección de Evaluación. Instituto Colombiano para la Evaluación de la Educación.*
- ICFES** (2017b). Implementación de un diseño de bloques incompletos balanceados en las pruebas genéricas de saber 3°, 5° y 9°. *Dirección de Evaluación. Instituto Colombiano para la Evaluación de la Educación.*
- ICFES** (2017c). Implementación de un diseño de bloques incompletos balanceados en las pruebas genéricas de saber pro y saber tyt. *Dirección de Evaluación. Instituto Colombiano para la Evaluación de la Educación.*
- Van der Linden, W., Veldkamp, B., and Carlson, J.** (2004). Optimizing balanced incomplete block designs for educational assessments. *Applied Psychological Measurement*, 28(5):317-331.
- Weller, S. C. and Romney, A. K.** (1998). *Systematic Data Collection*. Sage Publications.

SABER

AL > DETALLE

¿Qué diseño del armado se emplea en el Icfes para medir las pruebas Saber?



GOBIERNO DE COLOMBIA



@icfescol



ICFES



icfescol



YouTube: ICFES

Instituto Colombiano para la Evaluación de la Educación, ICFES

Oficinas: Calle 26 NO. 69-76 Torre 2, piso 15

Edificio Elemento, Bogotá • Colombia