



Serie de Cuadernillos Técnicos

Análisis de ítems con jMetrik



Licenciada Cynthia del Aguila Mendizábal
Ministra de Educación

Licenciada Evelyn Amado de Segura
Viceministra Técnica de Educación

Licenciado Alfredo Gustavo García Archila
Viceministro Administrativo de Educación

Doctor Gutberto Nicolás Leiva Alvarez
Viceministro de Educación Bilingüe e Intercultural

Licenciado Eligio Sic Ixpancoc
Viceministro de Diseño y Verificación de la Calidad Educativa



Directora Lcda. Luisa Fernanda Müller Durán

Subdirección de Análisis de Datos

Autoría

M.A. Mario Quim Can

M.A. José Adolfo Santos Solares

Revisión de texto y diagramación

Lcda. María Teresa Marroquín Yurrita

Diseño de portada

Lic. Roberto Franco

Dirección General de Evaluación e Investigación Educativa

© Digeduca 2015 todos los derechos reservados.

Se permite la reproducción de este documento total o parcial, siempre que no se alteren los contenidos ni los créditos de autoría y edición.

Para efectos de auditoría, este material está sujeto a caducidad.

Para citarlo: Quim, M. y Santos, J. (2015). *Análisis de ítems con jMetrik*. Guatemala: Dirección General de Evaluación e Investigación Educativa, Ministerio de Educación.

Disponible en red: <http://www.mineduc.gob.gt/Digeduca>

Impreso en Guatemala

divulgacion_digeduca@mineduc.gob.gt

Guatemala, 2015

Contenido

Introducción	4
Procesamiento para el análisis de ítems	5
Preparar la base de datos para cargar a jMetrik	5
Cargar los datos en jMetrik	6
Codificación de ítems en jMetrik.....	11
Gestión del análisis de ítems	14
Retroalimentación sobre el análisis de ítems.....	17
a. Dificultad	18
b. Discriminación.....	20
Análisis de distractores.....	22
Análisis de omisiones.....	24
Aspectos que son analizados del instrumento en su conjunto	25
a. Confiabilidad del instrumento	25
Referencias	29

Lista de figuras

Figura 1. Guardado de base de datos en archivo de Excel separado por comas.....	6
Figura 2. Creación de <i>Workspace</i>	7
Figura 3. Ubicación de <i>Workspace</i>	7
Figura 4. Área de trabajo	8
Figura 5. Crear base de datos.....	8
Figura 6. Importar datos.....	9
Figura 7. Ubicación de la importación de datos.....	10
Figura 8. Tabla de datos importada	10
Figura 9. Vista de variables.....	11
Figura 10. Indicación de clave de calificación	12
Figura 11. Ítem con clave	12
Figura 12. Replicar codificación	13
Figura 13. Variables con claves replicadas	14
Figura 14. Cuadro para elección de los ítems a analizar	15
Figura 15. Ítems preparados para correr el análisis de distractores.....	15
Figura 16. Cuadro de resultados de los ítems, distractores	16
Figura 17. Resultados de confiabilidad de la prueba	17

Lista de tablas

Tabla 1. Análisis de la dificultad del ítem	19
Tabla 2. Rango de dificultad de los ítems	19
Tabla 3. Rango de discriminación de los ítems	21
Tabla 4. Análisis de la discriminación del ítem.....	21
Tabla 5. Análisis de distractores.....	23
Tabla 6. Análisis de omisión	24
Tabla 7. Resumen del modelo.....	27
Tabla 8. Índices de fiabilidad	28

Introducción

El presente cuadernillo de análisis de ítems forma parte de una colección que tiene el propósito de describir los procesos que se desarrollan en la Subdirección de Análisis de Datos de la Dirección General de Evaluación e Investigación Educativa –Digeduca– del Ministerio de Educación de Guatemala.

En dicha subdirección, existe una fase que se relaciona con la retroalimentación del funcionamiento de las pruebas y los ítems que la componen, además de generar resultados de la evaluación del sistema educativo en las áreas de Lectura y Matemáticas; esto con la misión principal de proveer información objetiva, transparente y actualizada con rigor científico. Retroalimentar el diseño de instrumentos de evaluación asegura a la Digeduca mantener la mejora continua de sus procesos, los cuales buscan el perfeccionamiento de los instrumentos de medición que aplica y la generación de resultados basados en el análisis de datos con métodos estadísticos que muestren una mayor confiabilidad.

Un análisis que se realiza previo a la generación de resultados es el estudio psicométrico de cada uno de los instrumentos que se utilizan en los diversos niveles y áreas de aprendizaje. Este análisis comprende la fiabilidad del instrumento de evaluación, la dificultad de los ítems, la discriminación de los ítems y el funcionamiento de los distractores de estos ítems.

El proceso de análisis de ítems tiene su génesis en el ingreso de los datos al *software* libre y abierto jMetrik, que genera datos considerando la psicometría clásica y moderna. Contiene funciones de análisis, compatibles a la teoría de respuesta al ítem y equiparación de pruebas.

Procesamiento para el análisis de ítems

Para realizar este tipo de análisis se pueden utilizar distintos *software*. En la Digeduca se usa tanto el análisis de ítems en Teoría Clásica de los Test para el cual se emplea el *software* SPSS y la Teoría de Respuesta al Ítem, cuyos programas son Winsteps y jMetrik. Con estos se realiza el proceso de calificar las pruebas tanto de Matemáticas como de Lectura y generar los resultados.

En este apartado se explica el uso del *software* jMetrik para el análisis de los ítems de las pruebas estandarizadas de Lectura y Matemáticas.



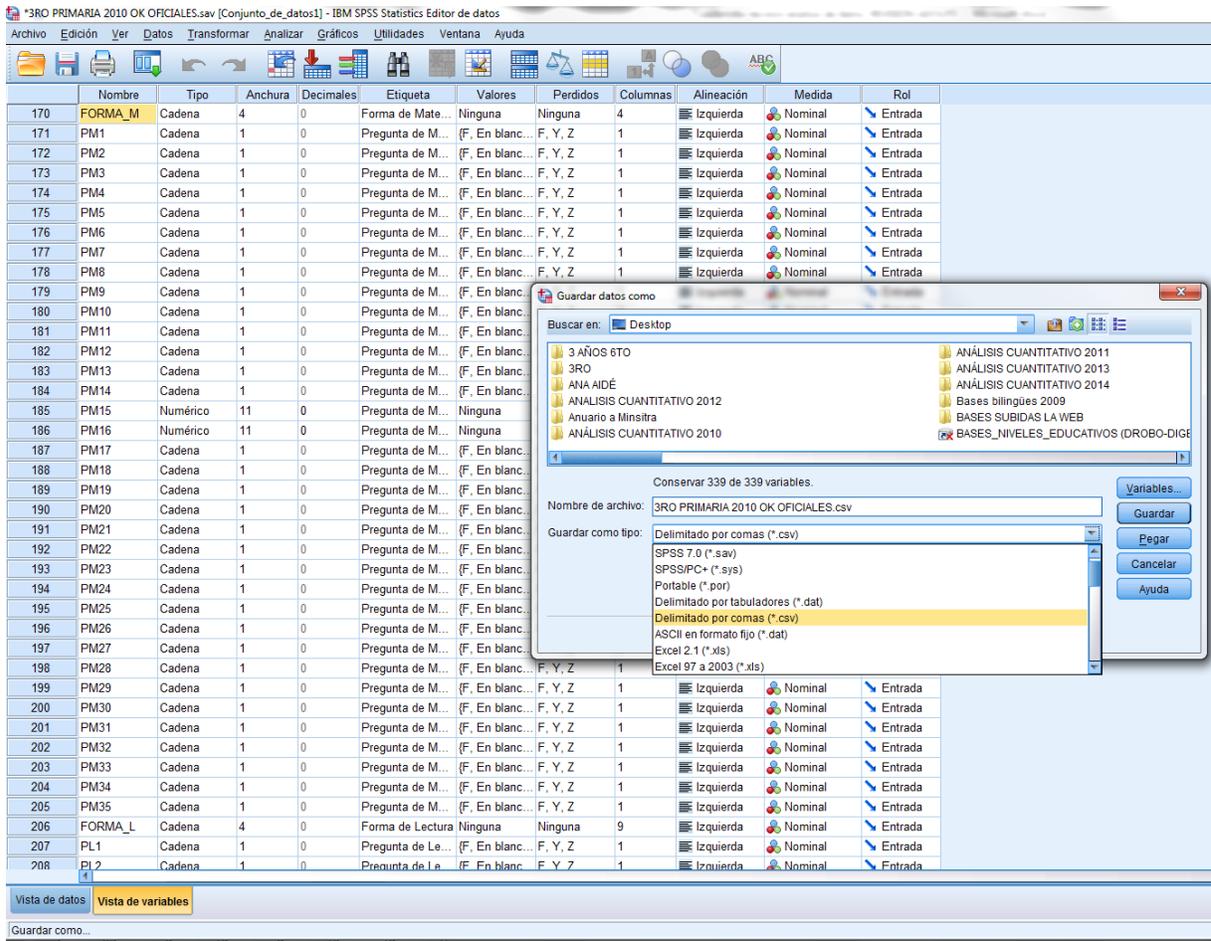
Preparar la base de datos para cargar a jMetrik

Se debe tener la base de datos en Excel con valores separados por coma; esto implica estar guardada bajo la extensión “*.csv.” en una carpeta nueva para que el proceso sea ordenado y separado de otros. Como las bases de datos limpias se tienen guardadas en SPSS, se puede hacer la conversión con el menú Archivo, en Guardar datos como, de manera que se elija la opción “*.csv” tal como se muestra en la Figura 1. La conversión en este tipo de documento permite representar los datos en forma de tabla, separando las columnas por comas.

Luego se debe hacer una revisión del archivo de Excel separado por comas para asegurar que el contenido esté completo, tanto los casos como las variables y que no se tenga basura que pueda afectar los datos. Además, hay que verificar que no se tengan datos vacíos, de manera que previamente se debieron haber tomado las medidas para convertir las omisiones en una categoría adicional de respuesta.

Las variables que se incluyen esencialmente en el archivo son las de identificación del estudiante, la variable de forma de la prueba y las respuestas dadas por los alumnos a cada ítem. Se utilizará una tabla de datos para Matemáticas y otra para Lectura, que se analizan de forma separada.

Figura 1. Guardado de base de datos en archivo de Excel separado por comas



Fuente: Dgeduca 2014.

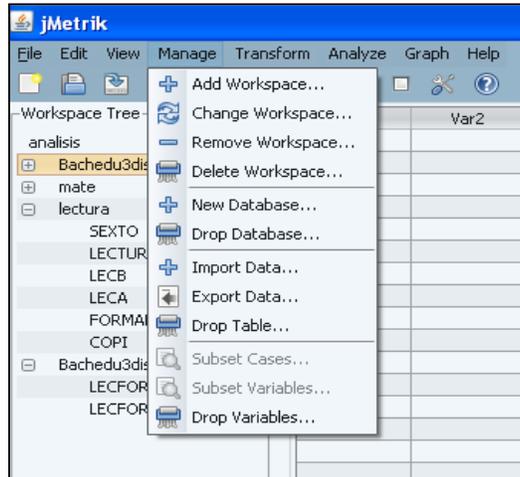
Cargar los datos en jMetrik

Al tener abierto el programa jMetrik, se abre el archivo de Excel desde este para que se cargue la base de datos, la cual estará lista para calificar la prueba. Luego se procede a indicar la clave de los ítems que permite identificar en el programa cuál es la respuesta correcta de cada uno de ellos. Los siguientes son pasos que detallan el procedimiento para dejar lista la base de datos para calificar en JMetrik.

1. Crear una nueva área de trabajo, ir al menú **Manage**, luego clic en **+Add Workspace**. Aparecerá una ventana donde se deberá escribir el nombre que se le dará a la nueva área de trabajo (*Workspace Name*) [ver la Figura 2]. También se deberá establecer su

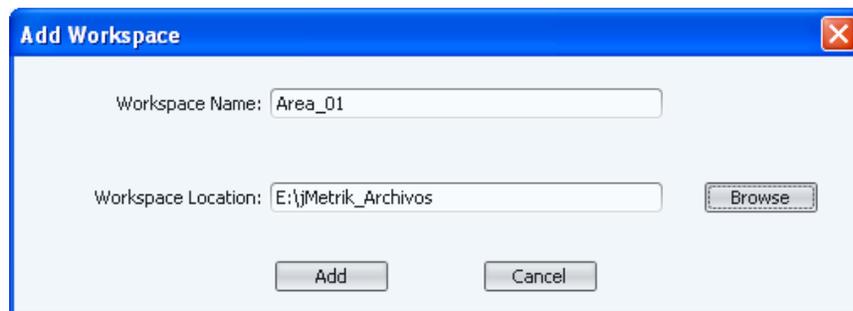
ubicación (*Workspace Location*) en alguna unidad de almacenamiento de datos, fija o extraíble (ver la Figura 3).

Figura 2. Creación de *Workspace*



Fuente: Digeduca 2014.

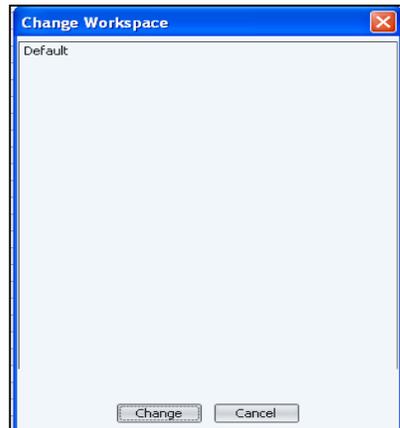
Figura 3. Ubicación de *Workspace*



Fuente: Digeduca 2014.

2. Para crear una nueva área de trabajo, hay que buscar en el menú **Manage** la opción **Change Workspace** (ver Figura 4). Se deberá dar un clic sobre el nombre del área de trabajo para seleccionarla y por último presionar el botón **Change**. Deberá aparecer el área de trabajo seleccionada en el diagrama de árbol.

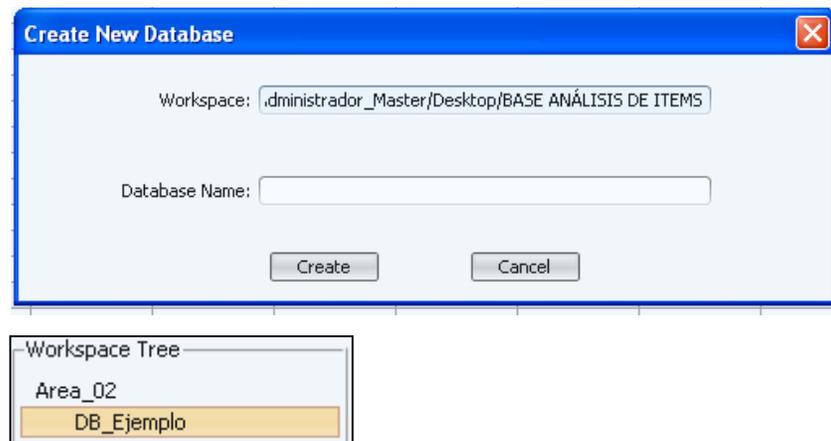
Figura 4. Área de trabajo



Fuente: Digeduca 2014.

3. Para crear una nueva base de datos, seleccione del menú **Manage**, la opción **New Database**. Se deberá dar un nombre a la base de datos para luego presionar el botón **Create**.

Figura 5. Crear base de datos



Fuente: Digeduca 2014.

Deberá aparecer en el diagrama de árbol del área de trabajo, la nueva base de datos creada dentro de su respectiva área de trabajo.

4. Para la realizar la importación de datos, se deberá seleccionar el tipo de archivo (ver Figura 6). El programa jMetrik reconoce dos tipos de archivos:

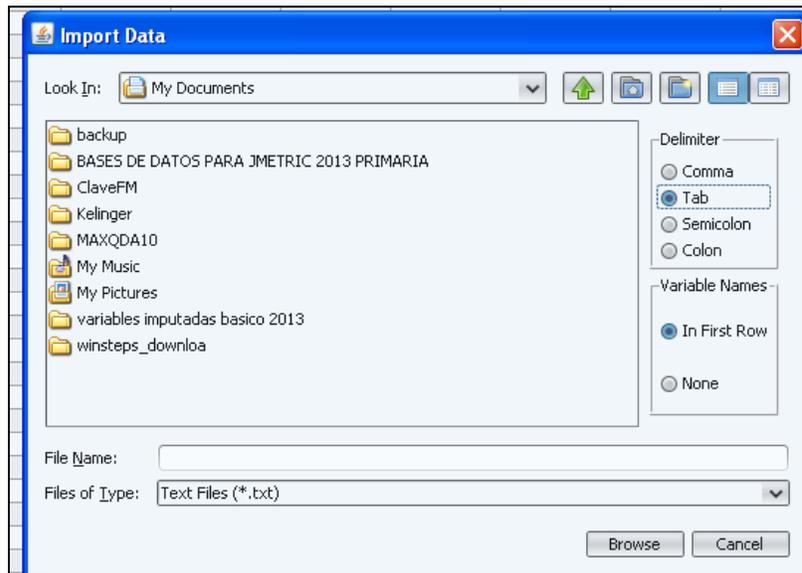
- a) Archivos de texto – Text Files(*.txt)
- b) Archivos delimitados por coma – Comma Delimited Files (*.csv)

Los archivos de texto y los delimitados por comas puede ser visualizados y editados con el *Notepad* o bloc de notas en Windows, otro editor de textos, Excel o similar. Al seleccionar un archivo para importar, también deberá indicar qué tipo de delimitador posee. Los principales son:

- **Comma:** Delimitador “,” (para datos separados por comas).
- **Tab:** Delimitador “ ” (es un espacio entre cada variable, datos separados por espacios).
- **Semicolon:** Delimitador, punto y coma “;” (para datos separados por punto y coma).
- **Colon:** Delimitador, dos puntos “:” (para datos separados por dos puntos).

Por último, por defecto, en *Variable Names* se indicará *In First Row*, es decir los nombres de las variables se encontrarán en la primera fila de la tabla. Si el archivo no contiene el nombre de las variables, se deberá seleccionar *None*.

Figura 6. Importar datos

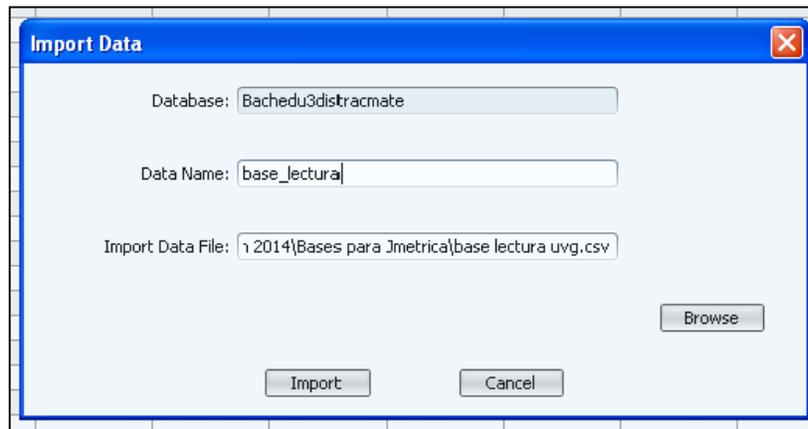


Fuente: Digeduca 2014.

Para importar datos, en el menú **Manage** habrá que buscar la opción **Import Data**. Se deberá dar un nombre a la tabla (**Table Name**) e indicar la ubicación del archivo en la unidad de almacenamiento, fija o extraíble. Al momento de presionar el botón **Browse**, aparecerá la ventana que se muestra en la Figura 7.

Se debe tener cuidado al momento de seleccionar el delimitador correcto; de lo contrario los datos se importarán con errores. Se debe seleccionar el botón **Browse** e indicar la ubicación.

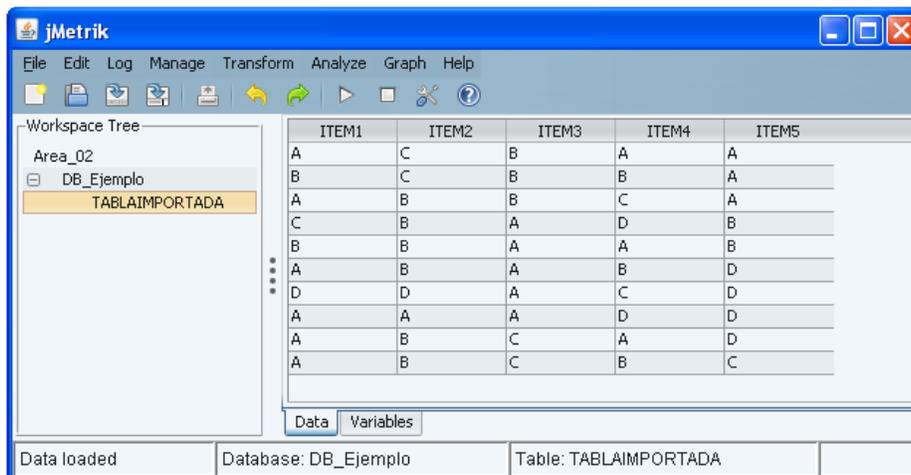
Figura 7. Ubicación de la importación de datos



Fuente: Dgeduca 2014.

En el cuadro de texto, **Rows to Scan** (número de filas que deben analizarse para comprobar los tipos de datos) puede dejarse vacío. Para finalizar la importación de datos, se presiona el botón **Import**. La tabla de datos importada se verá como lo muestra la Figura 8.

Figura 8. Tabla de datos importada



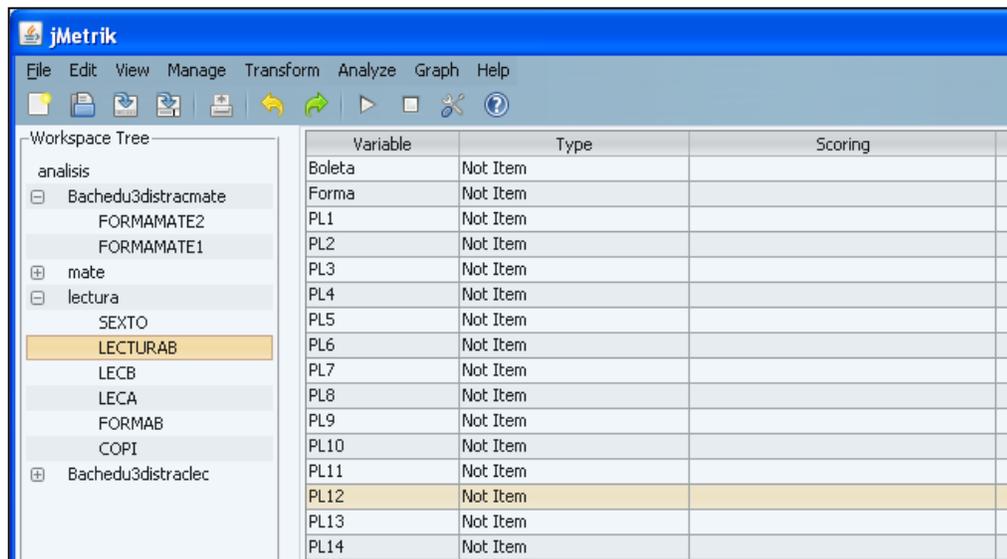
Fuente: Dgeduca 2014.

Codificación de ítems en jMetrik

Los pasos para preparar la base de datos en jMetrik de manera que se identifiquen los códigos ingresados, se detallan a continuación:

1. Presionar el botón de la pestaña de variables para activarla. Se mostrará un listado de las variables con su respectiva información, como se observa en la Figura 9.

Figura 9. Vista de variables



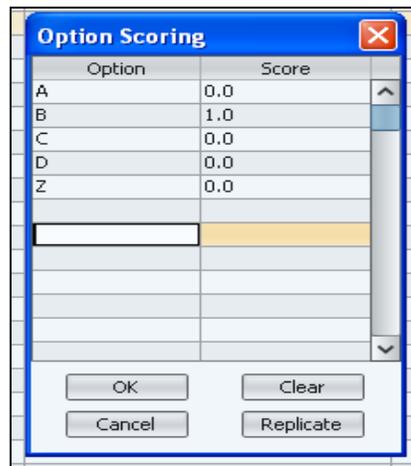
Variable	Type	Scoring
Boleta	Not Item	
Forma	Not Item	
PL1	Not Item	
PL2	Not Item	
PL3	Not Item	
PL4	Not Item	
PL5	Not Item	
PL6	Not Item	
PL7	Not Item	
PL8	Not Item	
PL9	Not Item	
PL10	Not Item	
PL11	Not Item	
PL12	Not Item	
PL13	Not Item	
PL14	Not Item	

Fuente: Dgeduca 2014.

2. Para cada ítem hay situarse en la fila que coincide con la columna **Scoring** y dar un clic sobre la celda seleccionada. Aparecerá una pequeña ventana. Se registrará cada alternativa de respuesta con su respectivo valor.

Por ejemplo, si para el ítem 1, con alternativas de respuesta A,B, C, D y Z, si la alternativa B es la correcta, entonces en **Score** se escribirá el valor "1", para las demás alternativas incorrectas se escribirá el valor "0", tal como se muestra en la Figura 10.

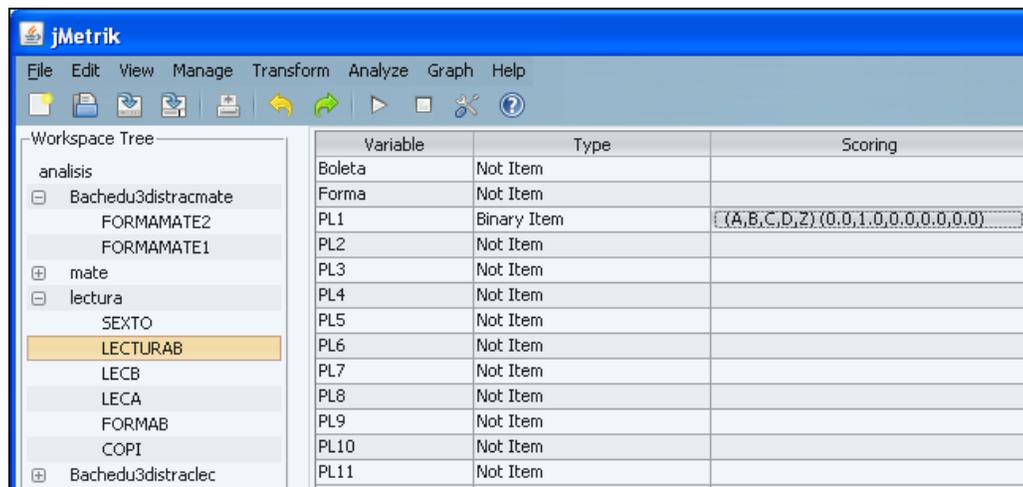
Figura 10. Indicación de clave de calificación



Fuente: Dgeduca 2014.

Al presionar el botón **OK**, aparece la vista de variables como se muestra en la Figura 11. El primer ítem identificado como PL1 ahora aparece como tipo **Binary Item** y en **Scoring** se asignan los valores entre “0” y “1” para las alternativas de respuesta, indicando cuál es la opción correcta.

Figura 11. Ítem con clave



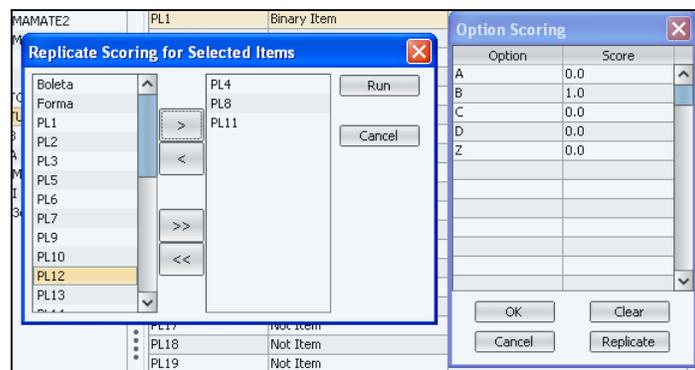
Fuente: Dgeduca 2014.

Una forma adecuada que ayuda a optimizar el tiempo de codificación de los ítems es hacer un listado en los que la respuesta correcta es A, B, C y D. Se usará la letra Z para indicar que el ítem no fue respondido. Los listados de ítems por opción de respuesta se observan a continuación:

- A: Ítem 2, 6, 7, 9, 12, 13, 18.
- B: Ítem 1, 4, 8, 11.
- C: Ítem 3, 10, 14, 15.
- D: Ítem 5, 16, 17, 19, 20.
- Z: En la base se designa con esta clave a las omisiones de los estudiantes a cada ítem.

Como en el ejemplo mostrado, la clave de respuesta B está en el primer ítem y es la misma para los ítems 4, 8 y 11, entonces se puede usar la función **Replicate** del **Scoring**. Para esto se ubica sobre la variable PL1 y se selecciona **Scoring**. En la ventana que aparecerá (como en la Figura 12) se debe presionar el botón **Replicate**, luego en la ventana siguiente, se elige la variable a replicar y se presiona el botón ">" (en este caso se replicaron las variables PL4, PL8 y PL11). Por último se presionan los botones **Run** y **Ok**.

Figura 12. Replicar codificación

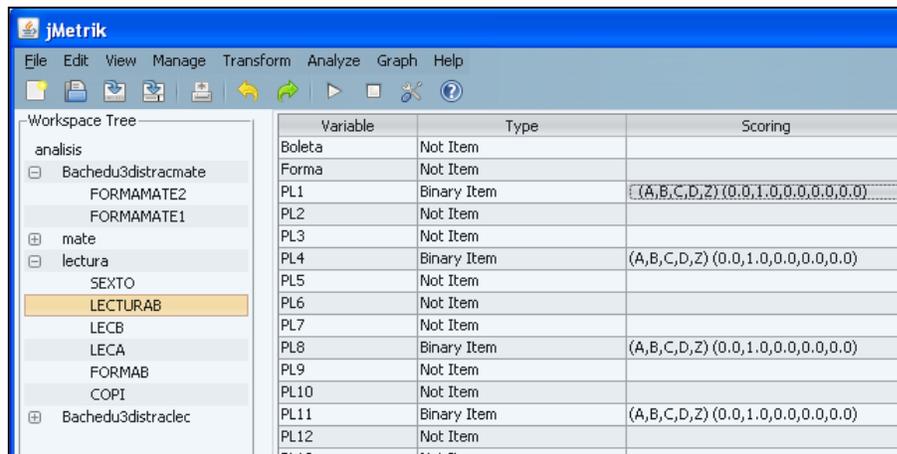


Fuente: Dgeduca 2014.

En la ventana de variables se puede observar que ahora existen con puntuación las variables ítem 1, 4, 8 y 11 (ver Figura 13).

Se hará lo mismo con todas las variables hasta que se tenga completas las claves en toda la tabla de datos. Con esto se podrá proceder a calificar las pruebas tanto de Matemáticas como la de Lectura.

Figura 13. Variables con claves replicadas



Variable	Type	Scoring
Boleta	Not Item	
Forma	Not Item	
PL1	Binary Item	(A,B,C,D,Z) (0.0,1.0,0.0,0.0,0.0)
PL2	Not Item	
PL3	Not Item	
PL4	Binary Item	(A,B,C,D,Z) (0.0,1.0,0.0,0.0,0.0)
PL5	Not Item	
PL6	Not Item	
PL7	Not Item	
PL8	Binary Item	(A,B,C,D,Z) (0.0,1.0,0.0,0.0,0.0)
PL9	Not Item	
PL10	Not Item	
PL11	Binary Item	(A,B,C,D,Z) (0.0,1.0,0.0,0.0,0.0)
PL12	Not Item	
PL13	Not Item	

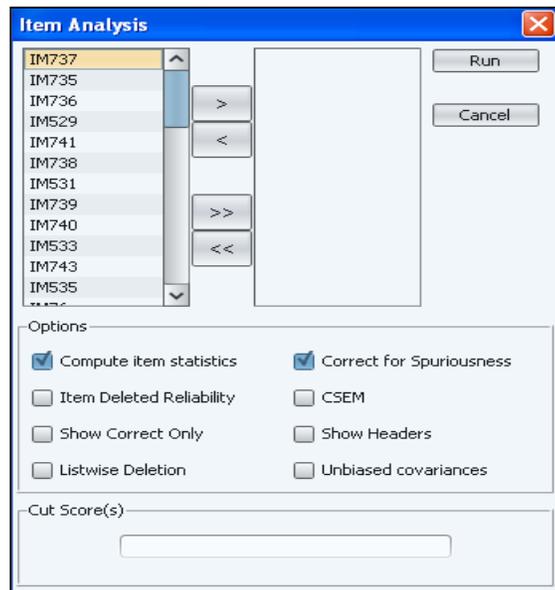
Fuente: Dgeduca 2014.

► Gestión del análisis de ítems

Si todas las tablas de datos correspondientes a cada forma han sido cargadas al programa jMetrik, se pueden comenzar a hacer análisis de ítems y calificación. Los siguientes pasos muestran cómo obtener estadísticas de los ítems y confiabilidad de las pruebas.

1. Ingresar por el menú **Analyze** para luego presionar el submenú **Item Analysis**. La Figura 14 muestra la ventana que se abrirá, en la cual se deben colocar de lado derecho los ítems que se necesitan analizar, tal y como aparece en la Figura 15.

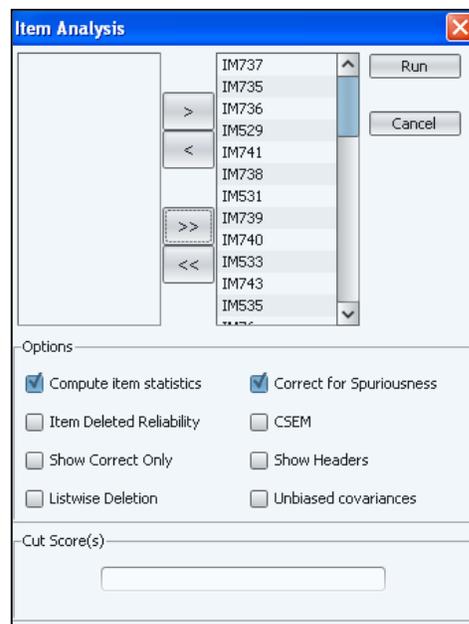
Figura 14. Cuadro para elección de los ítems a analizar



Fuente: Digeduca 2014.

2. Para seleccionar los ítems a analizar, habrá que añadirlos mediante el botón ">". Si se necesita analizar todos los ítems, entonces se presiona el botón ">>". En la parte de abajo se eligen las opciones de las cuales se desea obtener información en relación a los ítems, para finalmente presionar el botón **Run**.

Figura 15. Ítems preparados para correr el análisis de distractores



Fuente: Digeduca 2014.

3. Como consecuencia aparece un reporte como el de la Figura 16. Se verán los resultados por cada ítem como el índice de dificultad, la desviación estándar, la correlación de persona, ítem-total y la correlación serial puntual, así como los distractores.
4. Estos resultados se trasladan a formato Excel para facilitar el análisis de ítems, colocándolos en formatos que ayuden a la mejor interpretación y análisis de los mismos al momento de retroalimentar al área de Desarrollo de instrumentos de evaluación, ya que son quienes tomarán esta información al alimentar datos en el banco de ítems. La Figura 17 muestra los resultados de confiabilidad de la prueba, en donde el Alpha de Cronbach es el más utilizado.

Figura 16. Cuadro de resultados de los ítems, distractores

ITEM ANALYSIS					
C:/Documents and Settings/Administrador_Master/Desktop/BASE ANÁLISIS DE ÍTEMS/					
May 26, 2014 12:13:50					
Item	Option (Score)	Difficulty	S.D.	Item-Total Pearson	Correlation Polyserial
IM737	Item	0.2500	0.4339	0.2530	0.3447
	A(1.0)	0.2500	0.4339	0.2530	0.3447
	B(0.0)	0.4449	0.4980	-0.2443	-0.3073
	C(0.0)	0.3008	0.4596	-0.2456	-0.3236
	Z(0.0)	0.0042	0.0651	-0.0613	-0.3190
IM735	Item	0.3814	0.4868	0.1275	0.1625
	A(0.0)	0.2246	0.4182	-0.2302	-0.3207
	B(0.0)	0.3644	0.4823	-0.1724	-0.2209
	C(1.0)	0.3814	0.4868	0.1275	0.1625
	Z(0.0)	0.0297	0.1700	-0.1159	-0.2916
IM736	Item	0.3771	0.4857	0.1084	0.1383
	A(1.0)	0.3771	0.4857	0.1084	0.1383
	B(0.0)	0.2161	0.4125	-0.2228	-0.3129
	C(0.0)	0.3432	0.4758	-0.1549	-0.2000
	Z(0.0)	0.0636	0.2445	-0.1173	-0.2296
IM529	Item	0.1907	0.3937	0.0106	0.0153
	A(0.0)	0.3008	0.4596	-0.0489	-0.0644
	B(0.0)	0.1441	0.3519	-0.0849	-0.1313
	C(0.0)	0.2924	0.4558	-0.1115	-0.1476
	D(1.0)	0.1907	0.3937	0.0106	0.0153
	Z(0.0)	0.0720	0.2591	-0.2313	-0.4357

Fuente: Dgeduca 2014.

Figura 17. Resultados de confiabilidad de la prueba

RELIABILITY ANALYSIS			
Method	Estimate	95% Conf. Int.	SEM
Guttman's Lambda_2	0.6458	(0.5780, 0.7075)	2.8385
Cronbach's Alpha	0.6272	(0.5558, 0.6921)	2.9123
Feldt-Gilmer	0.6333	(0.5631, 0.6972)	2.8883
Feldt-Brennan	0.6301	(0.5593, 0.6945)	2.9071
Raju's Beta	0.6272	(0.5558, 0.6921)	2.9123

Fuente: Dgeduca 2014.

Retroalimentación sobre el análisis de ítems

La subdirección de Desarrollo de la Dgeduca tiene en su agenda posterior a la aplicación de los instrumentos de evaluación, utilizar el análisis de los resultados de la aplicación de cada uno de los instrumentos que han desarrollado.

Para garantizar el cumplimiento y condiciones psicométricas que requiere un instrumento, en este caso el ítem (unidad básica que constituye una prueba) pasa por una serie de análisis antes de colocarse en una prueba, con la finalidad de responder a la población objetivo para el cual fue creado. Un análisis de ítems generado puede ser tanto de un pilotaje de los mismos o posterior a la aplicación definitiva, lo cual se conoce como retroalimentación del funcionamiento de las pruebas y sus ítems.

El procedimiento por el cual se capturan los datos psicométricos de cada ítem es denominado pilotaje y se realiza en una muestra de la población que será evaluada. Para esto se requiere un mínimo de siete estudiantes por cada ítem en la prueba, con el propósito de garantizar suficiente variabilidad de los resultados, es decir, que pueda capturar una suficiente varianza que existe en la población. Los aspectos principales que se analizan de un ítem son:

- a) Dificultad
- b) Discriminación
- c) Distractores
- d) Omisiones

Los aspectos principales que analizan del instrumento en su conjunto son:

- a) Confiabilidad de la prueba o forma
- b) Ajuste al modelo de calificación o escala

Tal como lo señala autores como Morales, Muñiz y Ponce de León, estos análisis aportan información complementaria al instrumento, pero no necesariamente garantizan la validez del mismo.

a. Dificultad

El indicador de dificultad del ítem proviene de la proporción de sujetos que respondieron correctamente el ítem (A) entre el total de sujetos que respondieron el ítem (N). El indicador tiene valores entre 0 y 1 en donde 0 indica la máxima expresión de dificultad y 1 la mínima.

$$ID = A / N$$

El índice de dificultad de los ítems a su vez puede construir la media de dificultad de la prueba completa, ya que es la media de la dificultad de los ítems que la conforman.

$$ID = \frac{A - E / (K - 1)}{N}$$

Donde A es el número de sujetos que aciertan el ítem, E es el número de sujetos que fallan, K es el número de alternativas del ítem y N es el número de sujetos que intentan resolverlo.

Tabla 1. Análisis de la dificultad del ítem

No. ÍTEMS	DIFICULTAD	DIFICULTAD DE OPCIONES DE RESPUESTA
Ítem 1	0.6183	Porcentaje de respuesta correcta
A	0.1505	El distractor está bien
B	0.6183	Respuesta correcta
C	0.1237	El distractor está bien
D	0.086	El distractor está bien
Z (No respuesta)	0.0215	Respuesta menor al 5 %
Ítem 2	0.2473	Porcentaje de respuesta correcta
A	0.2204	El distractor está bien
B	0.2043	El distractor está bien
C	0.2473	Respuesta correcta
D	0.2043	El distractor está bien
Z (No respuesta)	0.1237	Respuesta mayor al 5 %

Fuente: Digeduca/Mineduc, 2014.

En la Tabla 1 puede observarse un ejemplo del reporte de dificultad de los ítems para una prueba de opción múltiple con cuatro alternativas de respuesta. En ella se analizan los distractores en función a su dificultad, esta es la proporción de estudiantes que los eligen. Por ejemplo, la opción A del ítem 1 muestra que 15 % de los estudiantes la eligieron, en este caso equivocadamente. También puede notarse que la respuesta correcta es la que tiene mayor proporción de elección, de lo contrario el ítem no funciona bien. Además, la Z representa a quienes dejan sin respuesta el ítem. Si la cantidad de estudiantes que no lo respondió es mayor al 5 % (0.05) se alerta para que el ítem sea revisado, como se observa en el ítem 2 del ejemplo. El ítem 2 es más difícil que el primero, pues los distractores tienen proporciones de dificultad similar a la respuesta correcta y muchos estudiantes dejan el ítem sin responder.

Tabla 2. Rango de dificultad de los ítems

RANGO DE DIFICULTAD	CATEGORÍA	ÍTEM DE LA PRUEBA	CANTIDAD DE ÍTEM	PORCENTAJE
I.D. < 0.25	Muy difíciles	3,16,19,22,44,46,50	7	14 %
0.25 ≤ I.D. < 0.45	Difíciles	4,6,7,12,14,15,18,23,24,29,33,41,45,48,49	15	30 %
0.45 ≤ I.D. < 0.55	Normales	1,5,8,9,10,17,26,34,35,37,47	11	22 %
0.55 ≤ I.D. < 0.75	Fáciles	13,21,30,32,36,38,39,40,42,43	10	20 %
I.D. ≥ 0.75	Muy fáciles	2,11,20,25,27,28,31	7	14 %
TOTAL DE ÍTEMS			50	

Fuente: Adaptado de García, J., González, M.A., y Ballesteros, Belén.

Para la construcción de una prueba definitiva, García, González y Ballesteros (2001) recomiendan utilizar algunos criterios considerando la dificultad de los ítems, los cuales pueden ser revisados de la siguiente manera:

- ❖ 10 % de elementos muy difíciles.
- ❖ 20 % de elementos difíciles.
- ❖ 40 % de elementos normales.
- ❖ 20 % de elementos fáciles.
- ❖ 10 % de elementos muy fáciles.

En el ejemplo de la Tabla 2, los resultados muestran índices de dificultad difíciles, siendo muy reducido el número de ítems pertenecientes a la categoría fácil; cabe mencionar que ítems muy fáciles o muy difíciles no contribuirían a la confiabilidad del instrumento.

b. Discriminación

Por discriminación se entiende la capacidad que tiene el ítem en la prueba para separar a las personas que saben un atributo del constructo evaluado y las personas que no lo saben. Según Morales (2009) “los índices de discriminación expresan en qué medida cada pregunta o ítem diferencia a los que más y menos saben”.

Muñiz (2005), menciona que el objetivo de la medición de discriminación y homogeneidad para los ítems, responde a la necesidad que cada elemento de una prueba debe poder diferenciar a las personas que puntúan alto y aquellas que puntúan bajo; por lo tanto, un buen ítem debe ser acertado en mayor proporción por aquellos que puntúan alto en la prueba.

Es importante comprender que el índice de discriminación proviene entonces de la correlación del resultado en el ítem con la calificación total de la prueba y se aduce que el resultado del ítem fue eliminado del total de la prueba para hacer la correlación.

Para Muñiz (2005), los ítems cuyo índice se ubican entre 0.00 y 0.10, no son adecuados para discriminar; entre 0.10 y 0.19, se encuentran en un límite que requiere que el ítem sea mejorado y arriba de 0.20 son adecuados para discriminar; a medida que crece el índice la calidad de discriminar del ítem va mejorando. En cuanto a índices negativos que a veces se obtienen, el autor indica que el ítem está midiendo otro constructo, por lo tanto debe cambiarse.

Tabla 3. Rango de discriminación de los ítems

VALORES	INTERPRETACIÓN	ITEMS DE LA PRUEBA	TOTAL	PORCENTAJE
Igual o mayor que 0.40	El ítem discrimina muy bien	1,5,10,13,15,20,21,25,28,31,36,37,38,40,41	15	30 %
Entre 0.30 y 0.39	El ítem discrimina bien	2,3,9,12,17,23,24,27,35,39,42,43,48	13	26 %
Entre 0.20 y 0.29	El ítem discrimina poco	8,11,16,18,30,32,45,46,47	9	18 %
Entre 0.10 y 0.19	Ítem límite. Se debe mejorar	4,6,7,14,19,22,26,29,33,34,44,49,50	13	26 %
Menor de 0.10	El ítem carece de utilidad para discriminar.		0	0 %
TOTAL DE ÍTEMS			50	100 %

Fuente: Adaptado de Muñiz, J., Fidalgo A.M., García-Cueto, E. y Moreno R.

Con el objetivo de garantizar buenas pruebas en el Ministerio de Educación, la Dgeduca utiliza el punto de corte de 0.20 como mínimo en el índice de discriminación. Aquellos ítems que no lo alcanzan, son reportados por el departamento de Análisis cualitativo o a la subdirección de Desarrollo de instrumentos, quienes tomarán las decisiones sobre su remoción.

Tabla 4. Análisis de la discriminación del ítem

ÍTEMS	DISCRIMINACIÓN	DISCRIMINACIÓN DE LOS DISTRACTORES POR SIGNO	DISCRIMINACIÓN DE CLAVE	RECOMENDACIÓN DEL ÍTEM
Ítem 1	0.4573		El ítem discrimina bien	El ítem está bien
A	-0.4816	El distractor discrimina negativo		
B	0.4573	Respuesta correcta		
C	-0.4177	El distractor discrimina negativo		
D	-0.3816	El distractor discrimina negativo		
No respuesta	-0.4032			
Ítem 3	-0.0273		Carece de utilidad para discriminar	Problema con la clave, posible doble respuesta
A	-0.4396	El distractor discrimina negativo		
B	0.0376	El distractor discrimina positivo		
C	-0.0648	El distractor discrimina negativo		
D	-0.0273	Respuesta correcta		
No respuesta	0.0482			
Ítem 4	0.3338		El ítem discrimina bien	El ítem está bien
A	0.3338	Respuesta correcta		
B	-0.4608	El distractor discrimina negativo		
C	-0.5846	El distractor discrimina negativo		
D	-0.445	El distractor discrimina negativo		
No respuesta	-0.0566			

Fuente: Dgeduca/Mineduc, 2014.

En el ejemplo presentado en la Tabla 3, el 26 % de los ítems necesitan ser mejorados. Aunque se toma como valor frontera mayor a 0.15 y menor a 0.20, se recomienda revisarlos y pueden no ser removidos de la prueba. En la Tabla 4 se encuentra un ejemplo del reporte de la discriminación y la recomendación técnica que produce.

Algo muy importante a considerar en el análisis de ítems es lo que Morales (2009) señala:

Las preguntas muy fáciles o muy difíciles no discriminan, no establecen diferencias; nos dicen que todos saben o no saben una pregunta, pero no quién sabe más y quién sabe menos. Estas preguntas no contribuyen a la fiabilidad, pero eso no quiere decir necesariamente que sean malas preguntas (son malas para discriminar...).

Análisis de distractores

Cada ítem de opción múltiple tiene alternativas que no son la respuesta correcta. A ellos se les denomina distractores. Un buen distractor es aquél que desvía la atención de la respuesta al presentarse como una posible alternativa si no se conoce cuál es la respuesta correcta; para esto, el estudiante necesita saber bien del tema que mide el ítem y así descartarlo.

En la Tabla 5 se encuentra un ejemplo del reporte de la dificultad de opciones de respuesta y la recomendación técnica que produce.

La mayoría de programas para análisis de funcionamiento de ítems coinciden en analizar la frecuencia en que cada opción incorrecta es seleccionada y cuando uno de ellos es seleccionado en muy baja o nula proporción, se puede concluir que el distractor no está cumpliendo su función adecuadamente, dividiendo la probabilidad de adivinar la respuesta correcta en las opciones restantes; por ende el ítem resulta ser más fácil. Asimismo, si la proporción de selección de una opción incorrecta es mayor que la respuesta, se entiende que esta opción está confundiendo a los evaluados y es recomendable reestructurar el ítem.

Tabla 5. Análisis de distractores

ÍTEMS	DIFICULTAD	DIFICULTAD DE OPCIONES DE RESPUESTA	RECOMENDACIÓN DEL ÍTEM
Ítem 7	0.5962		Revisar distractor
A	0.2692	El distractor está bien	
B	0.1154	El distractor está bien	
C	0.5962	El distractor está bien	
D	0.0192	El distractor tiene baja frecuencia	
No respuesta	0	Respuesta menor al 5 %	
Ítem 8	0.6474		Revisar distractor
A	0.6474	El distractor está bien	
B	0.0577	El distractor está bien	
C	0.0321	El distractor tiene baja frecuencia	
D	0.2564	El distractor está bien	
No respuesta	0.0064	Respuesta menor al 5 %	

Fuente: Dgeduca/Mineduc, 2014.

Tomando bajo consideración las situaciones anteriores, el reporte que realiza el departamento de Análisis cuantitativo luego de calificar las pruebas toma las siguientes consideraciones:

- a) Cada distractor debe ser seleccionado al menos por el 5 % de los evaluados.
- b) La respuesta correcta debe ser la opción con mayor porcentaje de selección.
- c) La elección de los tres distractores debe ser aproximada en proporción.
- d) La “no respuesta” al ítem no debe superar el 5 % de los evaluados.

► Análisis de omisiones

Respecto a la omisión de respuesta en un ítem, se considera que si una proporción alta de evaluados no lo responde, es porque este es confuso, es muy difícil o la longitud de la prueba impidió que los estudiantes llegaran a la posición en la que el ítem se encontraba.

En la Tabla 6 se observa un ejemplo del reporte que es entregado para retroalimentación. Para cada ítem se marca con “R/C” la respuesta correcta en la columna “Dificultad de opciones de respuesta”, se agrega una descripción de las opciones según el porcentaje de elección incluyendo la opción Z (No respuesta). En el ejemplo, los ítems 1 y 9 cumplen la función de selección de la respuesta correcta ya que fueron escogidos por una proporción más grande de estudiantes, pero dejaron de ser contestados por más del 5 % de los evaluados, por lo que se debe “revisar el distractor”.

Tabla 6. Análisis de omisión

ÍTEMS	DIFICULTAD	DIFICULTAD DE OPCIONES DE RESPUESTA	RECOMENDACIÓN DEL ÍTEM
Ítem 1	0.7628		Revisar distractor
A	0.0585	El distractor está bien	
B	0.0492	El distractor tiene baja frecuencia	
C	0.7628	Respuesta correcta	
D	0.0128	El distractor tiene baja frecuencia	
Z (No respuesta)	0.1166	Respuesta mayor al 5 %	
Ítem 9	0.4542		Revisar distractor
A	0.1974	El distractor está bien	
B	0.4542	Respuesta correcta	
C	0.1624	El distractor está bien	
D	0.0434	El distractor tiene baja frecuencia	
Z (No respuesta)	0.1426	Respuesta mayor al 5 %	

Fuente: Digeduca/Mineduc, 2014.



Aspectos que son analizados del instrumento en su conjunto

a. Confiabilidad del instrumento

El coeficiente de fiabilidad es una estimación de la correlación esperada con una prueba semejante y por lo tanto “este indica en qué medida en exámenes semejantes los alumnos hubieran quedado ordenados de manera parecida” (Morales, 2009).

Se debe tener en cuenta tres factores que inciden en la magnitud de este coeficiente: la homogeneidad de ítems, la diferencia entre los examinados y el número de ítems.

Si los ítems miden lo mismo, la fiabilidad será mayor; si los ítems contienen mucha variación entre las preguntas o si no existe relación, la fiabilidad será más baja. De igual forma, si los examinados obtienen resultados similares, la fiabilidad bajará. Mientras mayor sea la cantidad de ítems, será más fácil diferenciar a los estudiantes.

La fiabilidad depende de las diferencias entre los sujetos, por lo que se puede cuestionar la fiabilidad de un *test* o de una prueba objetiva como indicador necesario de su calidad: si todos saben todo o casi todo (o casi nada), la fiabilidad tiende a bajar y esto no quiere decir que el *test* sea malo o que se trate de un mal resultado ... Una fiabilidad alta nos dice que el examen deja a cada uno en su sitio; en exámenes parecidos (con otras preguntas del mismo estilo) los alumnos quedarían ordenados de manera semejante ... No hay que olvidar que una fiabilidad alta no es sinónimo sin más de calidad porque puede faltar lo que es más importante, la validez: preguntas que se pueden responder correctamente de memoria cuando lo que queremos es comprobar comprensión o interpretación, etc. (Morales, 2009).

Para poder calcular la fiabilidad de una prueba, esta debe estar formada por un conjunto de ítems cuyas puntuaciones se suman y el resultado es el que se interpreta. Aunque existen fórmulas más sofisticadas, el uso frecuente de este coeficiente es debido a la practicabilidad de su uso, ya que requiere una sola administración de la prueba.

La forma más sencilla de calcular el Alfa de Cronbach es multiplicar el promedio de todas las correlaciones observadas en los ítems por el número de ítems que componen una escala y luego dividir el producto entre el resultado de la suma de 1 más el producto de la multiplicación del promedio de todas las correlaciones observadas, por el resultado de la resta de 1 al número de ítems: $[a=n*p / 1 + p (n-1)]$, donde n es el número de ítems y p es el promedio de todas las correlaciones (Oviedo y Campo-Arias, 2005).

Varios autores como Oviedo y Campo-Arias (2005) advierten que:

Aunque no existen reglas al respecto, lo más común para interpretar los resultados es tomar 0.70 como el valor mínimo aceptable. Valores más bajos indican una consistencia interna baja y obtener un valor mayor a 0.90 puede indicar que la prueba es redundante debido a que varios ítems miden lo mismo. Lo más recomendable es que los valores oscilen entre 0.80 y 0.90.

De acuerdo a Bogliaccini, Cardozo y Rodríguez (2005), el análisis de confiabilidad a través de la consistencia interna permite estudiar la propiedad en la prueba, pero el valor del análisis es el que permite observar la relación entre cada ítem y el funcionamiento total de la prueba; por lo tanto, se puede saber qué tanto mejoraría o empeoraría la consistencia de la prueba si se decide eliminar el ítem. Muñiz (2005) añade que el valor radica en la optimización de la prueba como instrumento de medición. Igualmente, explica que el índice de confiabilidad (IF) es una función del índice de discriminación (ID) y la desviación estándar (DS) del ítem:

$$IF = DS_i ID_i.$$

El reporte que se genera en el departamento de Análisis cuantitativo, responde a la premisa que indica que cada ítem debe aportar a la consistencia interna de la prueba; por lo tanto, si su índice muestra que al eliminarlo de la prueba, la consistencia interna aumentaría, la recomendación técnica es no usar dicho ítem.

En el departamento se utiliza el estadístico *Reliability*, el cual se obtiene al generar el Análisis Factorial Confirmatorio en el *software* jMetik. Los datos estadísticos que producen, se muestra en la Tabla 7, la cual sirve de ejemplo para mostrar que se obtuvo la fiabilidad de 0.8791.

Tabla 7. Resumen del modelo

Statistic	Congeneric	Tau-Equivalent	Parallel
Fmin	1.9832	2.3286	6.8031
Chi ²	1525.0876	1790.68	5231.6205
df	1175	1224	1273
p-value	0	0	0
GFI	0.9207	0.9069	0.7279
AGFI	0.9139	0.903	0.7274
RMR	0.0084	0.0238	0.0283
RMSEA	0.0197	0.0245	0.0636
Reliability	0.8791	0.5122	0.3645

Fuente: Digeuca/Mineduc, 2014.

Según Tomás y Oliver (2004):

Tanto el índice GFI (Goodness-of-Fit Index) (Jöreskog & Sörbom, 1986) como su versión ajustada el AGFI (Adjusted Goodness-of-Fit Index), oscilan entre 0 y 1, y son los índices absolutos de mejor funcionamiento (Hoyle & Panter, 1995; Marsh et al., 1988). Se consideran como modelos adecuados a aquellos que superan el 0.9. Por otra parte, el RMSEA es una medida de error por grado de libertad del modelo, y permite tener una idea de la parsimonia del modelo (Loehlin, 1998), es decir su simplicidad. Reduciendo el número de parámetros, se considera que a igualdad de otras consideraciones es preferible un modelo simple a otro más complejo. Los valores del índice se interpretan de forma que un modelo con RMSEA menor de 0.08 indicaría que es un modelo razonable, mientras 0.05 indicaría un modelo con buen ajuste (Browne & Cudeck, 1993).

Un valor de *Reliability* (confiabilidad) que está en 0.80 se considera aceptable. Adicionalmente dentro del cuadro de resultados que se reporta a la subdirección de Desarrollo, se incluyen los índices de confiabilidad tradicionales que se generan del análisis de ítems, según la Tabla 8.

Tabla 8. Índices de fiabilidad

Method	Estimate	95% Conf. Int.	SEM
Guttman's Lambda_2	0.8586	(0.8440, 0.8725)	2.8467
Cronbach's Alpha	0.8538	(0.8387, 0.8682)	2.8947
Feldt-Gilmer	0.8577	(0.8429, 0.8717)	2.8562
Feldt-Brennan	0.8571	(0.8423, 0.8711)	2.8643
Raju's Beta	0.8538	(0.8387, 0.8682)	2.8947

Fuente: Digeduca/Mineduc, 2014.

De acuerdo con los resultados del índice de confiabilidad, cuando estos son menores a 0.80, se identifica la necesidad de realizar un cuidadoso análisis de la discriminación de cada uno de los ítems, para identificar cuáles no sería conveniente usar y así alcanzar un mayor porcentaje de fiabilidad en la prueba.

Referencias

- Bogliaccini, J.; Cardoso, M y Rodríguez, F. (2005) *Construcción de índices. Confiabilidad: Alpha de Cronbach*. Uruguay: Laboratorio Metodológico, Facultad de Ciencias Humanas Universidad Católica del Uruguay.
- García, J.; González, M.A. y Ballesteros, B. (2001) *Introducción a la investigación en educación*. Primera Edición. Universidad Nacional de Educación a Distancia. Madrid: Editorial IMPRESA.
- Mineduc, Perú. (2010) Manual básico de Usuario versión para Windows. Recuperado de: <http://www.perueduca.pe/documents/60563/9a3784fd-1b51-4290-8b4f86bd0a66972b>
- Morales, P. (2009). *Análisis de Ítems en las Pruebas Objetivas*. Universidad Pontificia Comillas. Facultad de Ciencias Humanas y Sociales. Madrid. España. Recuperado de: <http://web.upcomillas.es/personal/peter/otrosdocumentos/analisisitemspruebasobjetivas.pdf>
- Muñiz, J.; Fidalgo, A.; García-Cueto, E. y Moreno, R. (2005) *Análisis de los Ítems*. Primera Edición. Madrid: Editorial La Muralla S.A.
- Oviedo, H. y Campo-Arias, A. (2005). "Aproximación al uso del coeficiente alfa de Cronbach". *Revista Colombiana de Psiquiatría*, vol. XXXIV, núm. 4, 2005, pp. 572-580. Recuperado de: <http://www.redalyc.org/pdf/806/80634409.pdf>
- Tomás, J. y Oliver, A. (2004). "Análisis Psicométrico Confirmatorio de una Medida Multidimensional del Autoconcepto en Español". *Revista Interamericana de Psicología/Interamerican Journal of Psychology - 2004, Vol. 38, N.º 2 pp. 285-293*. Universitat de Valencia, España. Recuperado de: <http://dialnet.unirioja.es/servlet/articulo?codigo=3023239>

**Serie
de Cuadernillos Técnicos**



Análisis de ítems con jMetrik