



**Universidad Autónoma de Baja California**  
Instituto de Investigación y Desarrollo Educativo



## MANUAL PARA EL ANÁLISIS GRÁFICO DE ÍTEMS

- GUÍA PARA EL PROGRAMA PARA EL ANÁLISIS GRÁFICO DE ÍTEMS



GUADALUPE DE LOS SANTOS

ENSENADA, B.C. JUNIO DE 2010

# ÍNDICE

	Página
<b>1 PRESENTACIÓN.....</b>	<b>3</b>
<b>2 ANÁLISIS GRÁFICO DE ÍTEMS.....</b>	<b>4</b>
2.1 Evaluación de la calidad de los ítems mediante el Análisis Gráfico de Ítems.....	5
2.1.1 Ítem fácil que no discrimina.....	7
2.1.2 Ítem fácil que discrimina poco.....	10
2.1.3 Ítem fácil que discrimina bien.....	11
2.1.4 Ítem con dificultad media que no discrimina.....	13
2.1.5 Ítem con dificultad media que discrimina poco.....	14
2.1.6 Ítem con dificultad media que discrimina bien.....	15
2.1.7 Ítem difícil que no discrimina.....	16
2.1.8 Ítem difícil que discrimina poco.....	17
2.1.9 Ítem difícil que discrimina bien.....	18
2.2 Criterios para eliminar o corregir ítems .....	19
<b>3 PROGRAMA PARA EL ANÁLISIS GRÁFICO DE ÍTEMS.....</b>	<b>20</b>
3.1 Requerimientos.....	20
3.2 Instalación.....	21
3.3 Creación e importación de datos.....	21
3.4 Procesamiento de datos.....	22
3.5 Programa para el análisis gráfico de ítems .....	23
3.5.1 La interface del PAGI.....	23
3.5.2 Entrada de archivos.....	25
<b>4 GLOSARIO.....</b>	<b>30</b>
<b>5 REFERENCIAS .....</b>	<b>32</b>

# 1. PRESENTACIÓN

La evaluación es un proceso que siempre ha generado interés en los diferentes niveles educativos. Dentro del ámbito escolar, los profesores en servicio son quienes regularmente se encargan de evaluar el aprendizaje de los alumnos en el aula.

El profesor siempre ha interactuado con los alumnos para establecer una evaluación del aprendizaje de cada uno de ellos, en función del avance que presenten. Así, los docentes se han visto en la necesidad de elaborar instrumentos que detecten las necesidades de los alumnos para atenderlas, para observar su aprendizaje y realimentar los objetivos instruccionales a fin de mejorar el desempeño académico, o para certificar el logro educativo. Así, los profesores que desarrollan sus instrumentos de evaluación deben poner un énfasis tanto en las conexiones entre la evaluación y las actividades de enseñanza, como en el aspecto técnico para mejorar la calidad técnica de los ítems que integran esos instrumentos.

Para lograr esto último, la Teoría Clásica de los Tests (TCT) ha sido una guía para los constructores de pruebas. Esta teoría considera que tras aplicar y calificar una prueba, se obtiene el puntaje que logró cada examinado, el cual es conocido como puntuación observada. Sin embargo, el nivel real de dominio o habilidad que tiene el examinado no lo conocemos; porque la puntuación que observamos está influenciada por diversos errores de medición. La TCT nos ayuda a reducir los errores que cometemos al construir un examen, de modo que sea posible conocer el nivel de habilidad real de los estudiantes. Para ello, puede utilizarse, entre otros, el método del Análisis Gráfico de Ítems (AGI).

técnica del AGI y con el apoyo de la aplicación denominada Programa para el Análisis Gráfico de Ítems (PAGI).

## 2. ANÁLISIS GRÁFICO DE ÍTEMS

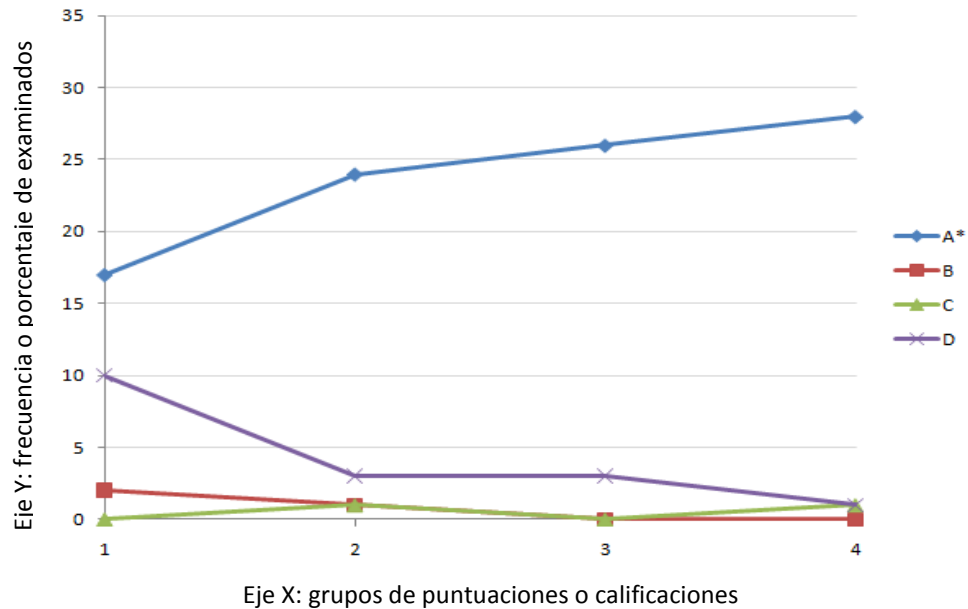
El Análisis Gráfico de Ítems (AGI) surgió de la propuesta de Batenburg y Laros (2002). El método consiste en mostrar visualmente, mediante una gráfica de líneas, la relación entre la puntuación total en la prueba y la frecuencia, proporción o porcentaje de respuestas de los examinados<sup>1</sup> que eligieron la opción correcta y las opciones falsas en un ítem de opción múltiple de una prueba. El AGI proporciona información esencial y fácilmente interpretable acerca de características técnicas del ítem como son su dificultad, su poder de discriminación y el nivel de adivinación.

El desarrollo del método AGI tuvo lugar en el contexto de proyectos para desarrollar pruebas a gran escala, específicamente en el proyecto del Sistema de Evaluación de la Educación Básica (SAEB) en Brasil. El SAEB se originó en 1997 para evaluar la ejecución de los estudiantes brasileños de educación básica en contenidos de matemáticas, física, e idioma portugués, entre otros (Van Batenburg y Laros, 2002).

La importancia de este método radica en ser utilizado para identificar ítems que presentan fallas y que deben ser excluidos de las pruebas por no tener los mínimos requerimientos técnicos psicométricos. Con el AGI los ítems de mala calidad son fáciles de detectar, porque en ellos se observa



Antes de empezar a describir el Análisis Gráfico de Ítems es necesario identificar los elementos necesarios para emplear esta técnica. Para ello, la figura 1 presenta las partes que se requieren para el análisis.



**Figura 1.** Componentes para el Análisis Gráfico de Ítems

En el método del AGI, el eje Y muestra la frecuencia o porcentaje de examinados que eligen las diferentes opciones. En el caso concreto de la figura 1 se muestra la frecuencia de estudiantes que eligieron cada opción. Mientras tanto, en el eje X se muestran los totales de aciertos o calificaciones que obtuvieron todos los estudiantes en la prueba, las cuales fueron divididas en 4 grupos iguales que representan los siguientes niveles de habilidad:

- Bajo (1)
- Medio bajo (2)
- Medio alto (3)
- Alto (4)

Los grupos de habilidad se despliegan en el eje X de izquierda a derecha para ubicar al extremo derecho al grupo 4, el de las calificaciones más altas, y que por ello es considerado el de mayor habilidad.

A la derecha de la figura 1 se encuentra una sección de código que identifica a cada una de las curvas de respuestas que dieron los examinados ante en las opciones que ofreció el ítem, mismas que están referidas como A, B, C y D. La curva de la respuesta correcta se identifica con un asterisco en la parte superior derecha de la letra correspondiente; en este caso se encuentra en la opción A. Las líneas que se observan en el eje de coordenadas son las curvas de respuestas al ítem. De ellas, una pertenece a la respuesta correcta y tres corresponden a los distractores, en este caso identificados con las letras B, C y D.

En la figura 1 se muestra un ítem de buena calidad, porque la frecuencia de examinados que eligen la respuesta correcta **A** incrementa a medida que aumenta la habilidad. Por otro lado, las demás curvas de respuestas al ítem se refieren a la frecuencia de examinados que eligen los distractores **B, C y D**, misma que en cada caso tiende a disminuir a medida que aumenta la habilidad. Ambas situaciones constituyen lo que es ideal en un buen ítem, por lo que este caso ilustra con claridad el tipo de información que aporta el análisis gráfico de ítems para mejorar los exámenes que elabora el profesor.

Después de conocer estos elementos generales del análisis gráfico de ítems, enseguida observaremos varios casos de ítems que ilustran diferentes tendencias y características de las respuestas de los examinados, mismas que hacen que un ítem pueda ser considerado como bueno o malo; fácil o difícil; con alto o bajo poder discriminativo; o con distractores efectivos o ineficaces.

### **2.1.1 Ítem fácil que no discrimina**

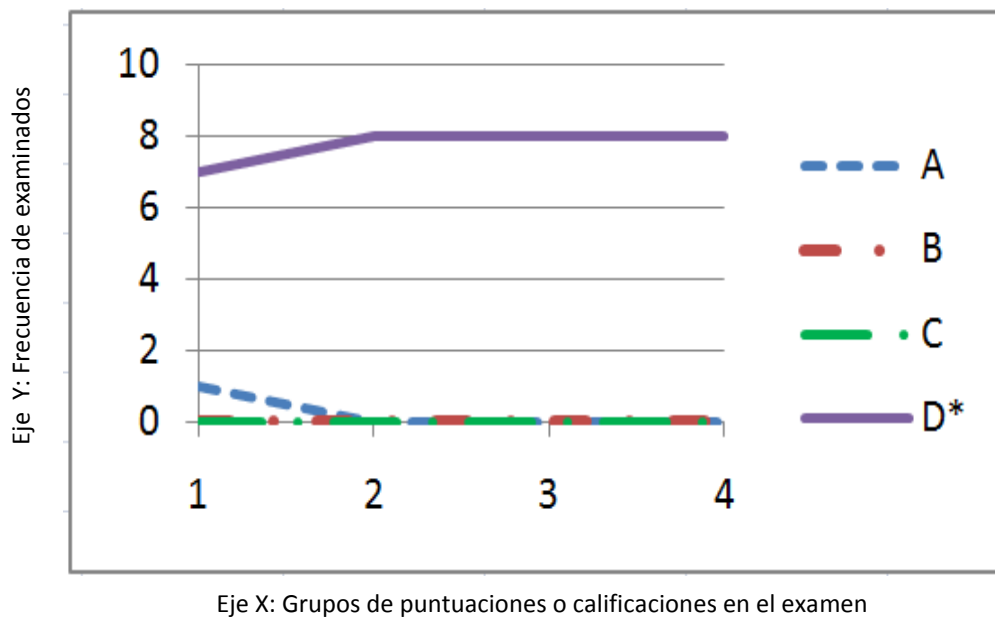
En la figura 2 aparece un ítem fácil, porque la curva de la respuesta correcta en el ítem muestra que el número de examinados que eligieron la opción correcta D es alto, tanto en los grupos de mayor habilidad (3 y 4), como en los de menor habilidad (1 y 2). Esto quiere decir que el ítem es muy fácil de contestar o que la respuesta correcta es evidente.

Por otro lado las opciones A, B y C, conocidas como distractores, en general no resultan atractivas para los examinados. Los distractores B y C no fueron elegidos por ningún examinado, ni en los grupos de alta habilidad (3 y 4), ni en los de baja habilidad (1 y 2); y la

opción A fue elegida por un solo examinado del grupo 1; es decir, que los distractores no resultaron atractivos para los examinados con calificaciones bajas en el examen, como debería haber sido.

Como consecuencia de lo anterior, el ítem no permite discriminar entre los examinados que dominan el contenido que evalúa el ítem y los examinados que no lo dominan. Esto puede observarse porque la pendiente o inclinación de la curva de la respuesta correcta es mínima y por ello tiende a ser horizontal.

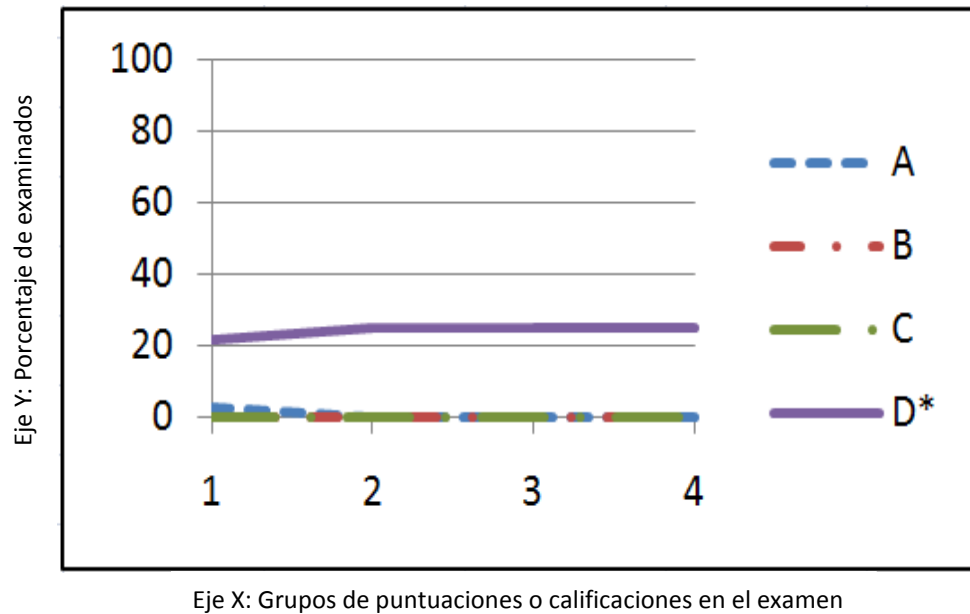
En síntesis, el ítem es muy fácil pero no permite discriminar entre los que saben y quienes no saben lo que el ítem evalúa, por lo que puede considerarse inadecuado.



**Figura 2.** Ilustración de un ítem fácil que no discrimina (Eje Y = N° de examinados)

En la figura 3 aparece el mismo ítem, pero ahora el eje Y muestra el porcentaje de examinados que eligieron las opciones, en vez del número de examinados que lo hicieron, como en la figura 2.





**Figura 3.** Ilustración de un ítem fácil que no discrimina (Eje Y = % de examinados)

La figura 3 permite observar de manera panorámica el total de respuestas que dieron el 100% de los examinados a todas las opciones del ítem, por lo cual se pueden ver con menos claridad y detalle las curvas de respuestas al ítem. Al igual que en la figura 2, se observa que este ítem es demasiado fácil y que existe poca diferencia entre los examinados que saben poco y los examinados de los grupos de habilidad alta, pues casi todos responden correctamente el ítem. Además, se ve que los distractores son poco atractivos para los examinados. La respuesta correcta resulta evidente. Para este tipo de ítem se sugiere revisar la redacción de los distractores a fin de hacerlos plausibles para quienes pertenecen a los grupos de habilidad baja (1 y 2). En otras palabras, este ítem es fácil pero malo; pues no discrimina entre los examinados.

La diferencia entre las figuras 2 y 3 radica en el cambio en el eje Y, del número o frecuencia de examinados, por el porcentaje de ellos. Este cambio es importante y necesario para que no se distorsione la perspectiva general del conjunto de respuestas que dieron los examinados ante el ítem. Es decir, sin importar el número de examinados que responden un ítem, siempre serán el 100%; por lo que se trata de un número fijo. En cambio, en la figura 2, el máximo número de examinados que aparece en el eje Y depende de la frecuencia con que se dieron las respuestas en cada opción del ítem, la cual puede variar de un ítem a otro.

Lo anterior significa que la figura 3 presenta las respuestas a todas las opciones que dieron el 100% de los examinados. La figura 2 presenta solo la parte de la gráfica, en la figura 3, a partir de la cual se dio la máxima frecuencia de examinados que respondieron alguna opción, usualmente la respuesta correcta cuando un ítem tiene calidad. Arriba de este punto no hay información. Por así decirlo, la figura 2 es como un zoom o ampliación que hacemos a la parte de la figura 3 que presenta la información más interesante sobre las respuestas que se dieron en el ítem, a fin de que podamos verla con más claridad.

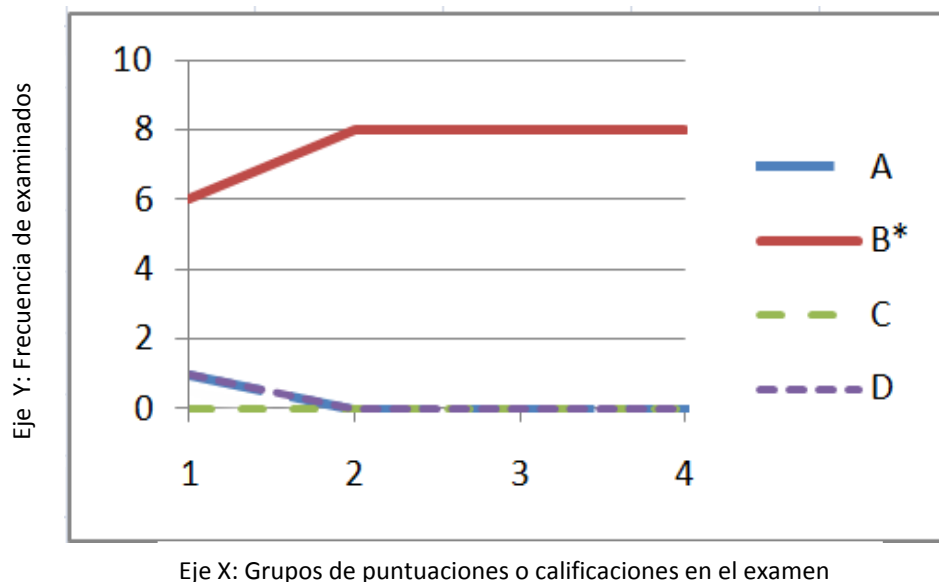
### **2.1.2 Ítem fácil que discrimina poco**

La figura 4 muestra un ítem fácil que discrimina poco, porque la curva de la respuesta correcta en el ítem muestra que el número de examinados que eligieron la opción correcta B, es alto en los grupos de habilidad media y alta (2, 3 y 4), mientras que en el grupo de habilidad más baja (1) es menor. Esto quiere decir que el ítem es muy fácil de contestar o que la respuesta es obvia.

En cuanto a los distractores, sus curvas de respuesta muestran que resultaron poco atractivos para los examinados. El distractor C no fue elegido por ningún examinado, ni en los grupos de alta habilidad (3 y 4), ni en los de baja habilidad (1 y 2). Las opciones A y D fueron elegidas, en cada caso, solo en una ocasión por un examinado del grupo más bajo (1); esto quiere decir que en general los distractores no resultaron atractivos para los examinados con poca habilidad en el examen, como debería ser.

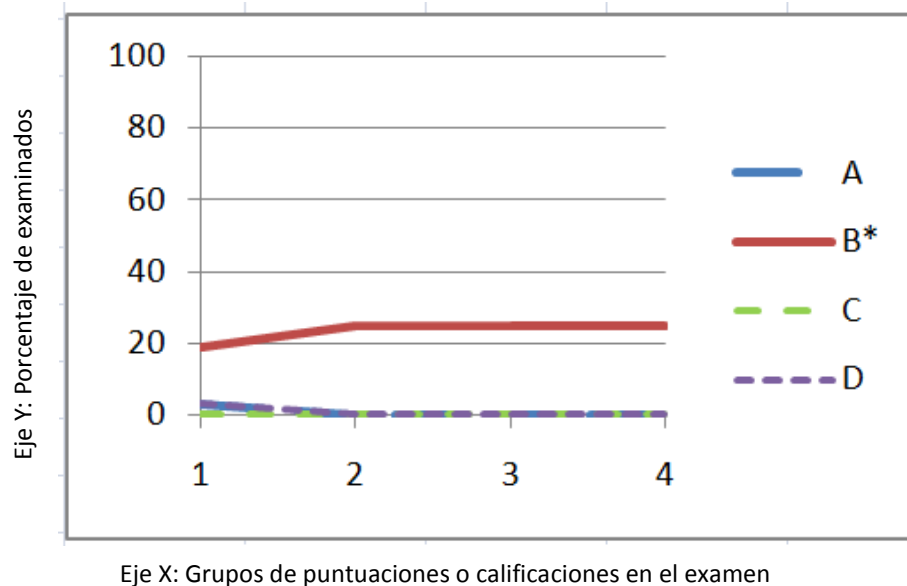
Como consecuencia de lo anterior, el ítem apenas es capaz de separar ligeramente a los examinados que conocen el contenido que evalúa el ítem, de los examinados que lo conocen menos. Esto puede notarse porque la inclinación de la curva de la respuesta correcta B es buena, pero solo entre los grupos de habilidad 1 y 2, pero luego tiende a ser horizontal a medida que incrementa la habilidad.

En resumen, el ítem es fácil y discrimina entre los examinados un poco mejor que el ítem ilustrado en las figuras 2 y 3, por lo que apenas puede ser considerado adecuado.



**Figura 4.** Ilustración de un ítem fácil que discrimina poco. (Eje Y = N° de examinados)

En la figura 5 aparece el mismo ítem, pero ahora el eje Y muestra el porcentaje de examinados que eligieron las opciones, en vez del número de examinados que lo hicieron, como en la figura 4.



**Figura 5.** Ilustración de un ítem fácil que discrimina poco. (Eje Y = % de examinados)

### 2.1.3 Ítem fácil que discrimina bien

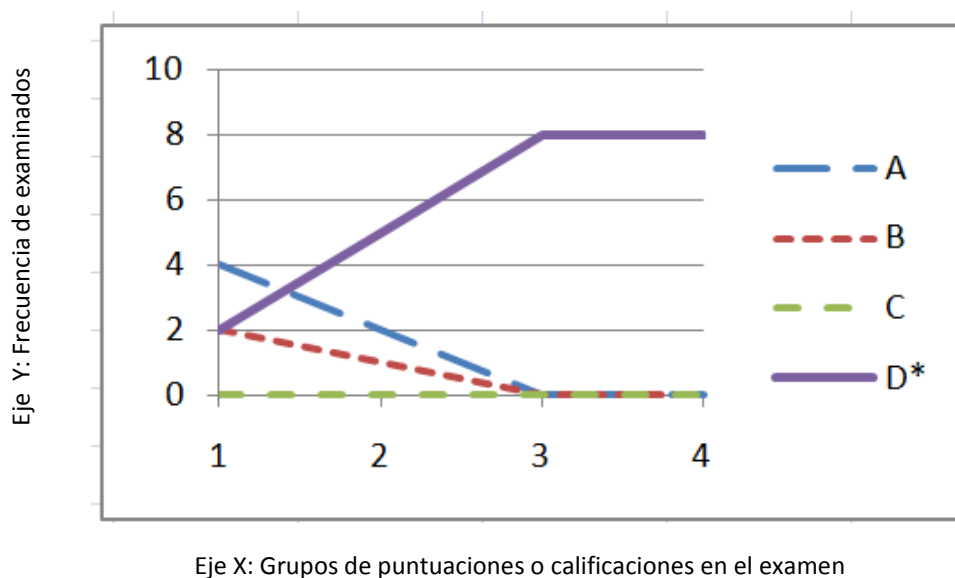
La figura 6 ilustra un ítem fácil que discrimina apropiadamente entre los examinados, porque la curva de la respuesta correcta en el ítem indica que el número de examinados que

eligieron la opción correcta D es alto en los grupos de habilidad 3 y 4, y empieza a incrementar desde el grupo de habilidad 2; pero en el grupo de habilidad 1 es menor. Esto quiere decir que el ítem es fácil solo para examinados con habilidad media y alta.

Respecto a los distractores, las curvas de respuestas de A y B indican que resultaron atractivos para algunos examinados de los grupos 1 y 2. El distractor C no resultó atractivo para ningún examinado. Esto quiere decir que los distractores fueron atractivos para examinados que no dominan el contenido evaluado, como debe ser el caso.

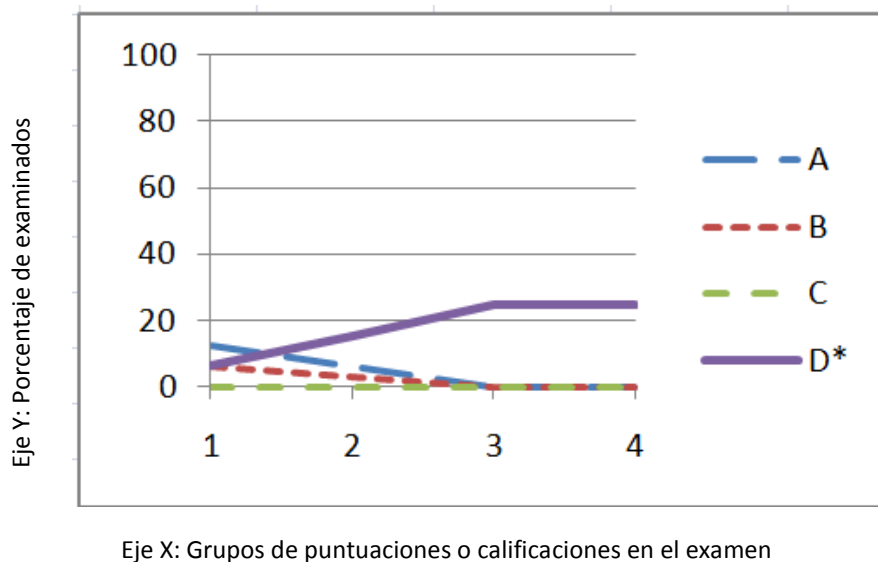
Por lo tanto el ítem es capaz de discriminar entre los examinados que dominan el contenido de los examinados que no lo dominan. Esto puede conocerse porque la inclinación de la curva de respuesta correcta incrementa a medida que aumenta el grupo de habilidad.

En resumen, el ítem resulta relativamente fácil y tiene una buena discriminación pues permite diferenciar a examinados que dominan el contenido de los que no lo dominan.



**Figura 6.** Ilustración de un ítem fácil con discriminación alta. (Eje Y = N° de examinados)

En la figura 7 aparece el mismo ítem, pero ahora el eje Y muestra el porcentaje de examinados que eligieron las opciones, en vez del número de examinados que lo hicieron, como en la figura 6.



**Figura 7.** Ilustración de un ítem fácil con discriminación alta. (Eje Y = % de examinados)

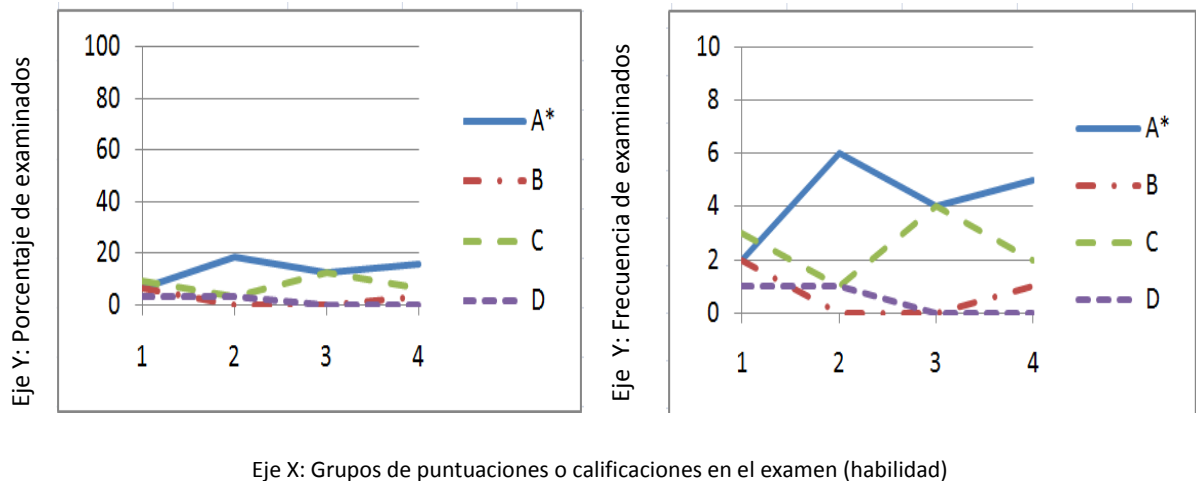
### 2.1.4 Ítem con dificultad media que no discrimina

La figura 8 presenta juntas las gráficas con frecuencia y porcentaje de examinados, y muestra un ítem con dificultad media que no discrimina, porque la curva de la respuesta correcta en el ítem deja ver que el número de examinados que eligieron la opción correcta A es un poco más de la mitad, pero fluctúa entre los grupos de habilidad; solo en el grupo de habilidad 2 incrementa y en los demás grupos se mantiene con altibajos. Esto quiere decir que el ítem es difícil de contestar para algunos grupos y para otros no; o que resulta confuso para los examinados con diferente habilidad.

Por otro lado, los distractores B y D parecen funcionar bien, aunque la opción B fue elegida por un examinado del grupo de habilidad 4. Además, el distractor C resultó atractivo para algunos de los examinados de mayor habilidad (grupos 3 y 4). Esto quiere decir, que esos distractores resultaron complejos para los examinados de los grupos de alta habilidad.

Como consecuencia de lo anterior, el ítem no permite discriminar con claridad entre los examinados que dominan el contenido que evalúa el ítem, y los examinados que no lo dominan. Esto puede distinguirse porque la pendiente o inclinación de la curva incrementa y decrecienta de manera irregular entre los grupos de diferente habilidad (1, 2, 3 y 4).

En resumen, el ítem resulta con dificultad media pero no permite discriminar entre los examinados que dominan el contenido de los que no lo dominan, por lo que puede considerarse como inadecuado. Una acción que podría mejorar el ítem sería redactar la opción C de manera diferente, a fin de que no fuera atractiva para los grupos de habilidad alta (3 y 4).



**Figura 8.** Ilustración de un ítem con dificultad media que no discrimina

### 2.1.5 Ítem con dificultad media que discrimina poco

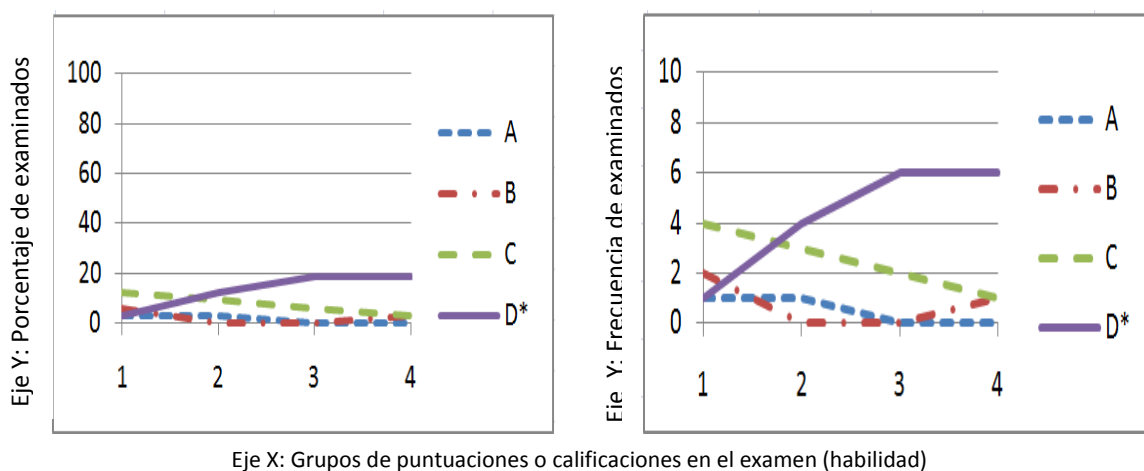
La figura 9 muestra un ítem con dificultad media que discrimina poco, porque la curva de la respuesta correcta en el ítem revela que los examinados que eligieron la opción correcta D se encuentran en todos los grupos de habilidad, aunque principalmente en los grupos 3 y 4, pero el distractor C es atractivo tanto para examinados que no dominan el contenido como para los que lo dominan. El ítem resulta de dificultad media porque los examinados se confunden entre la respuesta correcta D y el distractor C.

Por otro lado, el distractor A funciona bien pues es atractivo para los examinados que dominan poco el contenido, pero no lo es a medida que incrementa la habilidad. El distractor B parece funcionar bien, pero fue elegido por un examinado del grupo de habilidad 4.

Como consecuencia de lo anterior, el ítem permite discriminar entre los examinados que saben el contenido que evalúa el ítem de los examinados que no dominan el contenido.

Esto puede distinguirse porque la pendiente o inclinación de la curva de la respuesta correcta incrementa a medida que aumenta la habilidad.

En resumen el ítem resulta de dificultad media con una discriminación que es aceptable, por lo que puede ser considerado como mínimamente adecuado. Para mejorar el ítem se podría recomendar una redacción diferente del distractor C de manera que no sea atractivo para los examinados que dominan el contenido.



**Figura 9.** Ilustración de un ítem con dificultad media que discrimina poco

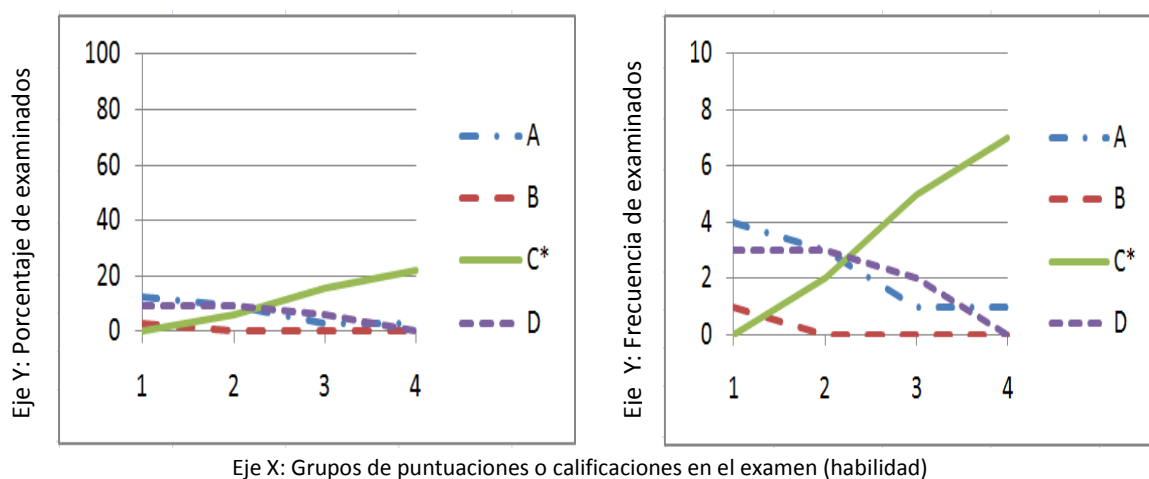
### 2.1.6 Ítem con dificultad media que discrimina bien

La figura 10 muestra un ítem de dificultad media que discrimina bien, porque la curva de la respuesta correcta en el ítem permite observar que el número de examinados que eligieron la opción correcta C aumenta a medida que incrementa la habilidad. Esto quiere decir que el ítem es fácil pero solo para examinados que dominan el contenido.

Por otro lado, los distractores B y D parecen funcionar bien, porque sus curvas de respuestas decrecientan a medida que aumenta la habilidad y distraen a examinados que no dominan el contenido. El distractor A también funciona bien, aunque confunde a un examinado del grupo de habilidad más alta. Esto da como resultado que los distractores funcionen en general correctamente porque distraen a examinados que dominan poco el contenido.

Como consecuencia de lo anterior, el ítem permite observar diferencias entre los examinados que dominan el contenido que evalúa el ítem de los que no lo dominan. Esto puede observarse porque la inclinación de la curva de la respuesta correcta se hace más pronunciada a medida que aumenta la habilidad.

En síntesis, el ítem resulta ser con dificultad media que discrimina bien entre los examinados que dominan el contenido de los que no lo dominan, por lo que es considerado como bueno.



**Figura 10.** Ilustración de un ítem con dificultad media que discrimina bien

### 2.1.7 Ítem difícil que no discrimina

La figura 11 muestra un ítem difícil que no discrimina, porque la curva de la respuesta correcta en el ítem muestra que el número de examinados que eligieron la opción correcta C es relativamente bajo y fluctúa entre los grupos de habilidad; solo en el grupo de habilidad 3 incrementa y en los demás grupos se mantiene con altibajos. Esto quiere decir que el ítem es difícil de contestar para algunos grupos y para otros no, o que resulta confuso para los examinados con diferente habilidad.

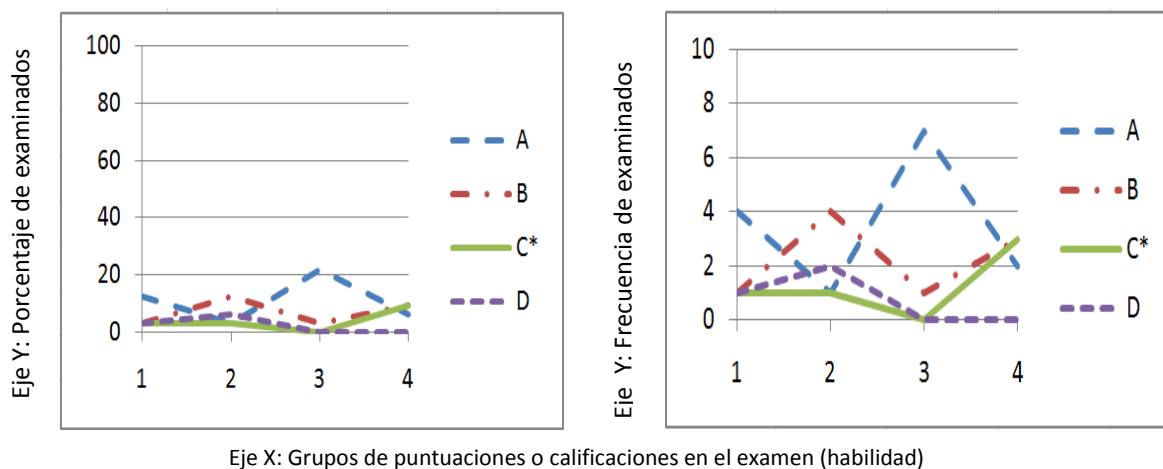
Por otro lado, el distractor D parece funcionar bien, porque disminuye la frecuencia o porcentaje de los examinados que lo eligen a medida que aumenta la habilidad. El distractor A funciona bien para el grupo de habilidad más baja (1), pero distrae considerablemente a los grupos de habilidad alta (3 y 4). La frecuencia o porcentaje de quienes eligen el distractor B



fluctúa dentro de los 4 grupos de habilidad, ocasionando que no disminuya a medida que aumenta la habilidad. Esto quiere decir, que los distractores resultaron complejos para los examinados de los grupos de habilidad alta.

Como consecuencia de lo anterior, el ítem no permite discriminar entre los examinados que dominan el contenido que evalúa el ítem de los examinados que no lo dominan. Esto puede distinguirse porque la pendiente o inclinación de la curva de la respuesta correcta fluctúa objetablemente entre los grupos de habilidad.

En síntesis, el ítem resulta ser difícil y no discrimina entre los examinados que dominan el contenido de los que no lo dominan, por lo que puede considerarse como inadecuado. Una acción que podría mejorar el ítem sería redactar de manera diferente los distractores A y B, a fin de que no fueran atractivos para los grupos de habilidad alta (3 y 4).



Eje X: Grupos de puntuaciones o calificaciones en el examen (habilidad)  
**Figura 11.** Ilustración de un ítem difícil que no discrimina

### 2.1.8 Ítem difícil que discrimina poco

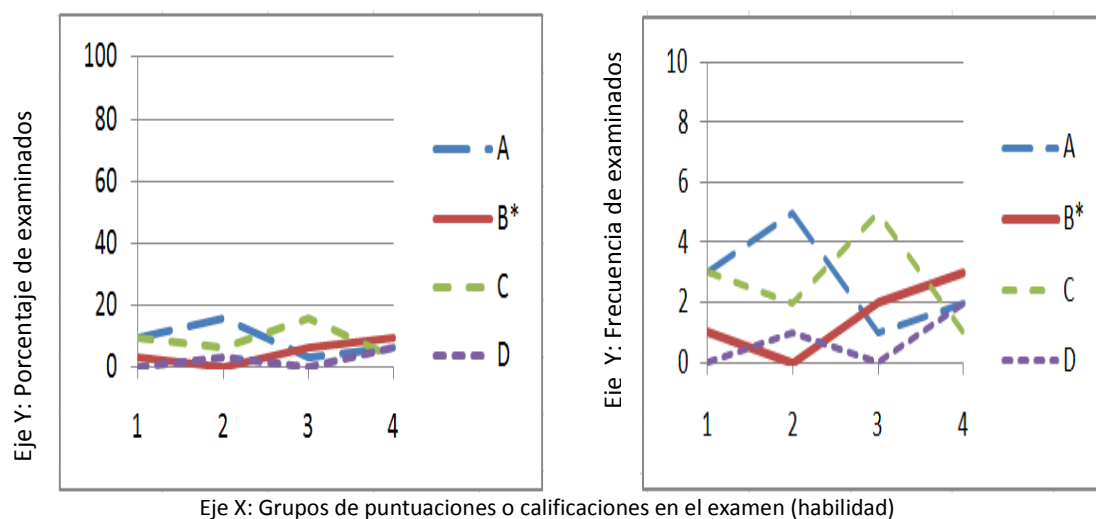
En la figura 12 se presenta un ítem difícil que tiene poca discriminación, porque la curva de la respuesta correcta muestra que el número de examinados que eligieron la respuesta correcta B incrementa poco a medida que incrementa la habilidad.

En cuanto a los distractores, por lo general resultaron atractivos tanto para los examinados que dominan el contenido (grupos 3 y 4), como para examinados que no lo dominan (grupos 1 y 2). La curva del distractor A incrementa en el grupo de habilidad 2 para

disminuir hasta el grupo de habilidad 3 pero vuelve a incrementar en el grupo de habilidad 4. Los distractores C y D presentan fluctuaciones entre los grupos de habilidad. Esto quiere decir, que los distractores resultaron por igual atractivos para examinados que dominan el contenido (lo que no debe suceder) y para examinados que no lo dominan (como debe ser).

En base a lo anterior, podemos decir que el ítem permite separar de manera mínima entre los examinados que comprenden el contenido que evalúa el ítem, de los que no lo comprenden. Esto puede observarse porque la pendiente de la curva de la respuesta correcta B incrementa en los grupos de habilidad alta (3 y 4).

En síntesis, el ítem resulta ser difícil y discrimina poco, por lo que puede ser considerado apenas adecuado. Para mejorar el ítem se recomienda redactar nuevamente los distractores A, C y D a fin de hacerlos menos atractivos para los que saben; o, en su caso, redactar la respuesta correcta de manera más contundente.



Eje X: Grupos de puntuaciones o calificaciones en el examen (habilidad)  
**Figura 12.** Ilustración de un ítem difícil que discrimina bien

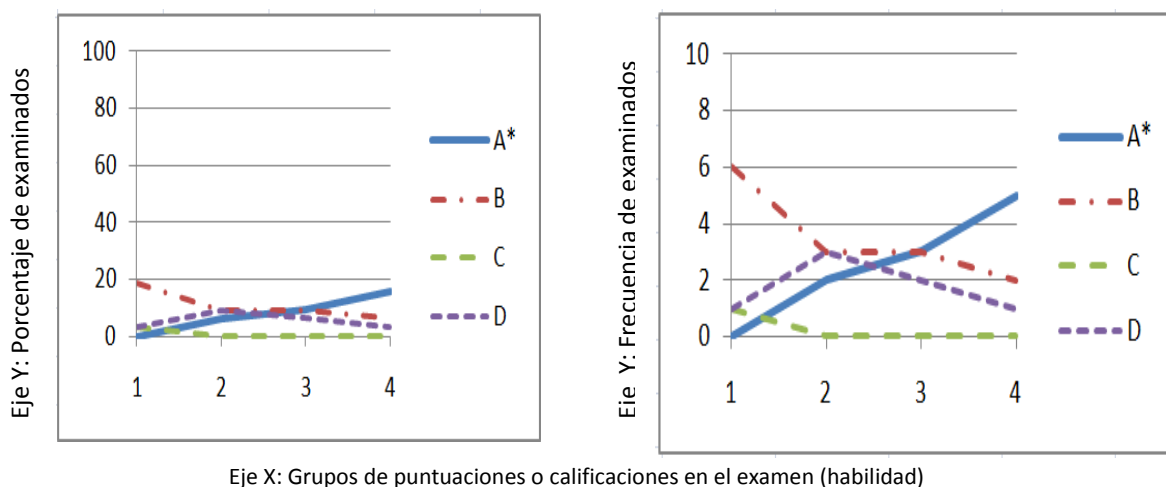
### 2.1.9 Ítem difícil que discrimina bien

La figura 13 presenta un ítem difícil que discrimina alto, porque la curva de la respuesta correcta en el ítem muestra que el número de examinados que eligieron la respuesta correcta incrementa a medida que aumenta la habilidad.

En cuanto a los distractores, la opción C demostró ser poco atractiva porque solo la eligió un solo examinado. El distractor B fue atractivo para examinados de los grupos de habilidad baja (1 y 2), pero también resultó atractivo para 5 examinados de los grupos de habilidad alta (3 y 4). El distractor D cumple con la función de ser atractivo para examinados que no dominan el contenido; sin embargo, fue elegido por 3 examinados de los grupos de habilidad alta (3 y 4). Esto quiere decir que los distractores resultaron atractivos para los examinados de los diferentes grupos de grupos de habilidad, aunque para el grupo de habilidad 4 resultaron atractivos de manera mínima.

En consecuencia, el ítem permite discriminar con claridad entre los examinados que dominan el contenido de los examinados que no lo dominan. Esto puede comprobarse porque la tendencia de la curva de la respuesta correcta incrementa a medida que aumenta la habilidad.

En resumen el ítem es difícil pero discrimina alto, por lo que puede ser considerado un ítem adecuado. Los examinados que dominan el contenido lo responden correctamente.



Eje X: Grupos de puntuaciones o calificaciones en el examen (habilidad)  
**Figura 13.** Ilustración de un ítem difícil que discrimina bien

## 2.2 Criterios para eliminar o corregir ítems

Los casos que se comentaron en esta sección, nos permiten apreciar que el AGI aporta mucha información significativa sobre la calidad de un ítem. Así mismo, nos permite conocer

las características que presentan los ítems que resultan inapropiados para evaluar el aprendizaje. En consecuencia, es posible proporcionar criterios que puedan utilizarse como una guía general para razonar sobre la posible eliminación o en su caso corrección de ítems defectuosos. Enseguida se presentan cuatro criterios principales que proponen Batenburg y Laros (2002) para determinar la eliminación o corrección de ítems:

- Tomar en cuenta el número de violaciones al supuesto de que la opción correcta incrementa con un aumento del puntaje total.
- Tomar en cuenta el número de violaciones al supuesto de que las opciones falsas decrezcan con un incremento del puntaje total.
- Tomar en cuenta el número de intersecciones entre las opciones correctas y falsas tras el inicio del intervalo de discriminación.
- La ausencia de poder discriminativo o una baja pendiente de la respuesta correcta.

No obstante lo anterior, considérese que al eliminar un ítem malo de una prueba debe ser reemplazado por otro ítem que cubra la misma parte del contenido aprendizaje.

### 3. PROGRAMA PARA EL ANÁLISIS GRÁFICO DE ÍTEMS

El Programa para el Análisis Gráfico de Ítems (PAGI), es un programa psicométrico gratuito y de código abierto, que fue desarrollado para efectuar el análisis gráfico de ítems basado en la Teoría Clásica de los Tests. PAGI es capaz de realizar estimaciones de la dificultad, la discriminación y el análisis de los distractores en los ítems de opción múltiple o de respuesta alterna de una prueba. Como hemos visto, estos análisis se pueden ejecutar e interpretar de manera sencilla con la ayuda de gráficos.

#### 3.1 Requerimientos

PAGI requiere la instalación previa de cualquier versión de Windows XP o Windows 7 y que se cuente con la máquina virtual de Java, versión 1.6 o superior. PAGI puede leer archivos de Excel elaborados en versiones de Microsoft Excel 2003 y 2007.

### 3.2 Instalación

El archivo de instalación PAGI.jar, puede ser instalado desde una carpeta zip enviada por correo electrónico o desde la carpeta donde fue copiado previamente. Al ejecutar este archivo, el módulo de análisis psicométrico se configura hasta estar listo para realizar el análisis gráfico de ítems.

### 3.3 Creación e importación de datos

Independientemente de cómo se registren las respuestas de los examinados en la prueba (manualmente o mediante un lector óptico), para que PAGI pueda analizar dichos resultados es necesario que estén en una tabla de doble entrada procesada en Excel. Es decir, PAGI analiza matrices de datos con un formato donde las columnas representan los ítems de la prueba y las filas o renglones a los examinados que respondieron los ítems del instrumento. De esta manera, cada celda de la tabla representa la respuesta que dio un examinado a uno de los ítems del examen. Lo anterior puede ser observado en el ejemplo que se presenta en la tabla 1. Esta tabla es conocida como tabla de datos brutos, porque en ella los datos de los examinados derivados de las respuestas que dieron en el examen no han sufrido modificación alguna.

**Tabla 1.** Datos brutos que representan las respuestas de los examinados ante el conjunto de ítems de la prueba

Número de ítem	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10
Clave	A	C	B	A	D	B	C	A	D	B
examinado										
1	A	C	B	A	D	B	C	A	D	B
2	#	C	C	A	D	B	C	A	D	A
3	A	C	D	C	A	B	D	A	C	B
4	D	C	B	A	B	B	C	C	D	B
5	A	B	B	#	D	D	C	D	#	#
6	C	C	B	A	D	A	A	A	A	B
7	D	#	B	A	D	B	C	A	D	B
8	B	C	B	D	C	B	C	A	C	B
9	A	C	B	A	A	B	C	B	A	C
10	A	C	B	A	D	A	C	A	D	B
11	A	B	D	A	D	#	#	#	#	#

Nótese que en la tabla, la primera fila corresponde a la clave de respuestas correctas para cada ítem del examen.

La tabla 1 presenta las opciones que eligieron los examinados en cada ítem de una prueba, mismas que pueden ser representadas con las letras A, B, C y D o con los números 1, 2, 3, y 4, cuando el ítem presenta cuatro opciones de respuesta. Por otro lado, en algunas celdas aparece el símbolo de #, lo cual significa que en ese ítem este examinado no contestó o no seleccionó ninguna de las opciones; estos datos son conocidos como datos perdidos. Por otro lado, cuando el examinado conteste otra opción diferente a alguna de las predeterminadas, su respuesta será identificada con el símbolo \$; es decir, si el examinado elige otro carácter o número que no están marcados como opciones de respuesta se utiliza el símbolo anterior.

Nótese que es importante poner en la tabla los símbolos # y \$ cuando corresponda, para que PAGI pueda procesar adecuadamente nuestra base de datos.

### **3.4 Procesamiento de datos**

Una vez que ha sido estructurada la base de datos en la tabla, tiene que ser depurada; es decir, se debe detectar a estudiantes e ítems inmedibles. Como ya se mencionó, es posible identificar a estudiantes que durante la aplicación de la prueba, no respondieron con seriedad. Estos estudiantes tienen que ser descartados para evitar una confusión en el análisis de los ítems de la prueba.

Para realizar el procesamiento de los datos, en este ejemplo se tomarán en cuenta los 10 primeros datos de la tabla 1, puesto que el examinado número 11 no respondió suficientes ítems y por ello fue depurado de la base de datos. Lo siguiente es convertir la tabla de datos brutos en una tabla de codificación de 1 y 0; es decir asignar el número 1 a los que contestaron correctamente y un 0 a los que tuvieron error en el ítem. El siguiente paso es calificar el examen y ordenar los puntajes de mayor calificación a menor, de acuerdo a las puntuaciones totales que obtuvieron los examinados. Esto es realizado por el programa PAGI.

**Tabla 2.** Codificación en 0 y 1 de los datos brutos, ordenados de mayor puntuación a menor

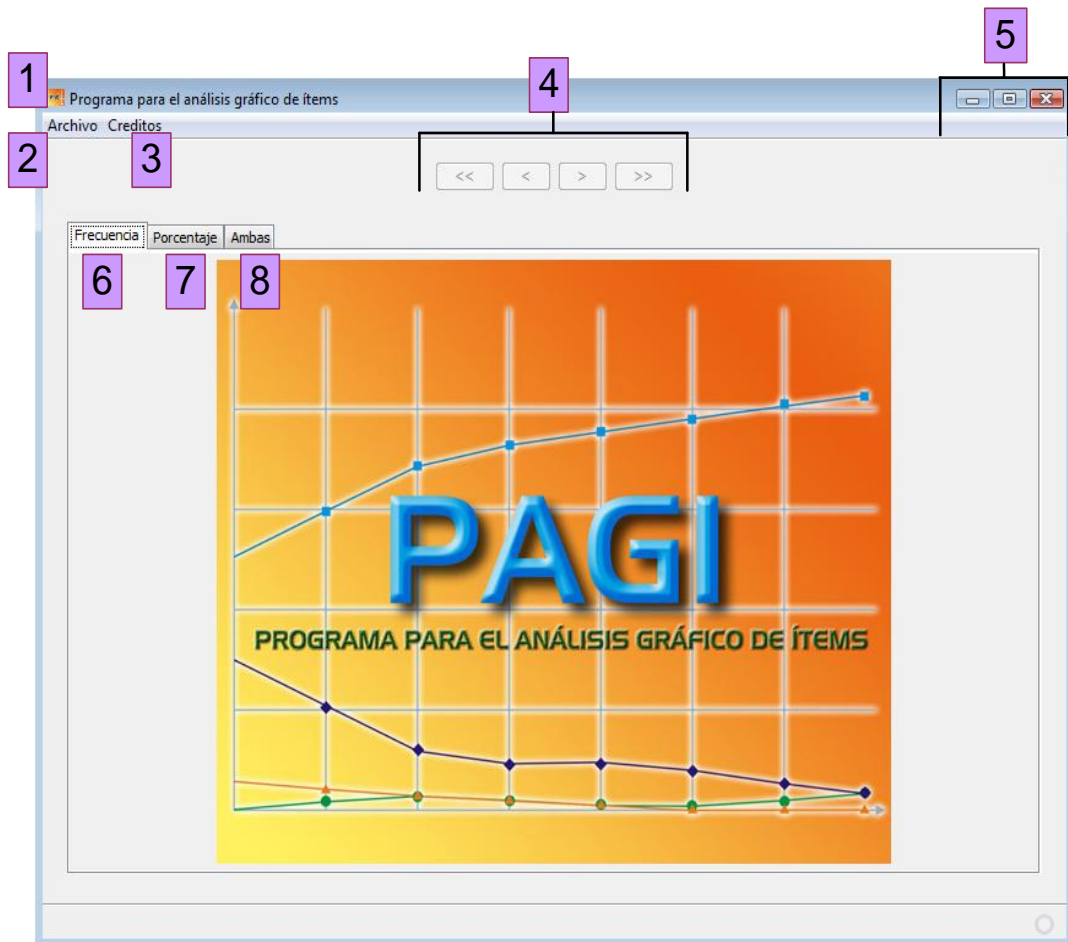
Número de Ítem	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	TOTAL
Clave Examinado	A	C	B	A	D	B	C	A	D	B	
1	1	1	1	1	1	1	1	1	1	1	10
7	1	0	1	1	1	1	1	1	1	1	9
10	1	1	1	1	1	0	1	1	1	1	9
2	0	1	0	1	1	1	1	1	1	0	7
4	0	1	1	1	0	1	1	0	0	1	6
6	0	1	1	1	1	0	0	1	0	1	6
8	0	1	1	0	0	1	1	1	0	1	6
9	1	1	1	1	0	1	1	0	0	0	6
3	1	1	0	0	0	1	0	1	0	1	5
5	1	0	1	0	1	0	1	0	0	0	4

Hasta el momento el programa PAGI ha logrado realizar una codificación de datos en ceros y unos; ha realizado la sumatoria de todos los aciertos; ha ordenado de mayor a menor las puntuaciones en la prueba y ha dividido las calificaciones de los examinados en cuatro grupos de habilidad, según los aciertos que obtuvieron.

### 3.5 Programa para el análisis gráfico de ítems

#### 3.5.1 La interface del PAGI

Para realizar el análisis gráfico de ítems por medio de la aplicación PAGI, es necesario conocer la estructura de la ventana principal, misma que se presenta en la figura 3.1. Como puede observarse, los componentes que la integran se encuentran enumerados de manera consecutiva y en cada caso se describe su función más abajo.



**Figura 3.1** Ventana principal del Programa para análisis gráfico de ítems

1. Título de la ventana principal.
2. Menú Archivo que ofrece dos opciones: una que permite seleccionar el archivo que se va a analizar y otra para salir del programa.
3. Menú de Ayuda, que presenta información general sobre el programa PAGI.
4. Controles que sirven para navegar entre los ítems del examen. Por ejemplo, al dar clic en el signo > se muestra el siguiente ítem y así sucesivamente.
5. Controles que sirven para minimizar, maximizar o cerrar la aplicación.
6. Control que permite observar la gráfica que presenta PAGI, con el eje Y en la modalidad de frecuencia de examinados.
7. Control que permite observar la gráfica que presenta PAGI, con el eje Y en la modalidad de porcentaje de examinados.
8. Control que permite observar simultáneamente ambas gráficas de la aplicación PAGI.



### 3.5.2 Entrada de archivos

En la figura 3.2 se muestra la entrada de los datos, en la parte superior izquierda del menú Archivo se muestra una opción señalada como “Abrir archivo”. Por medio de ella se puede buscar el archivo donde se encuentre la base de datos en Excel que contiene los datos brutos del examen.

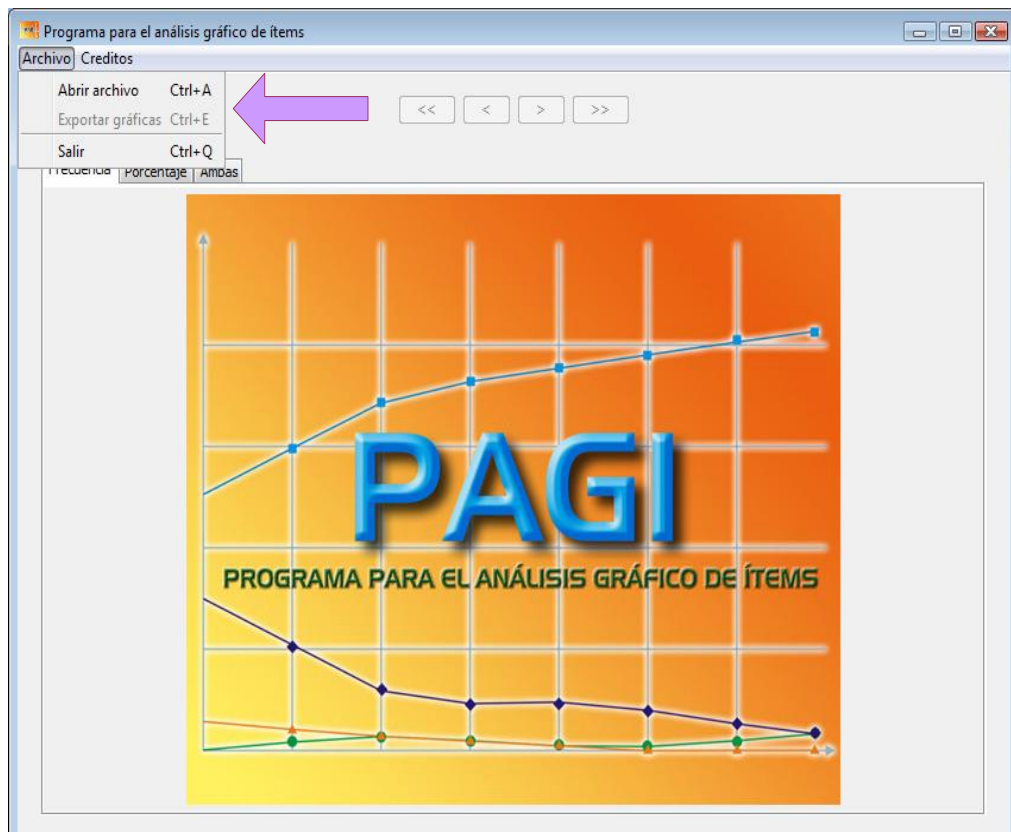
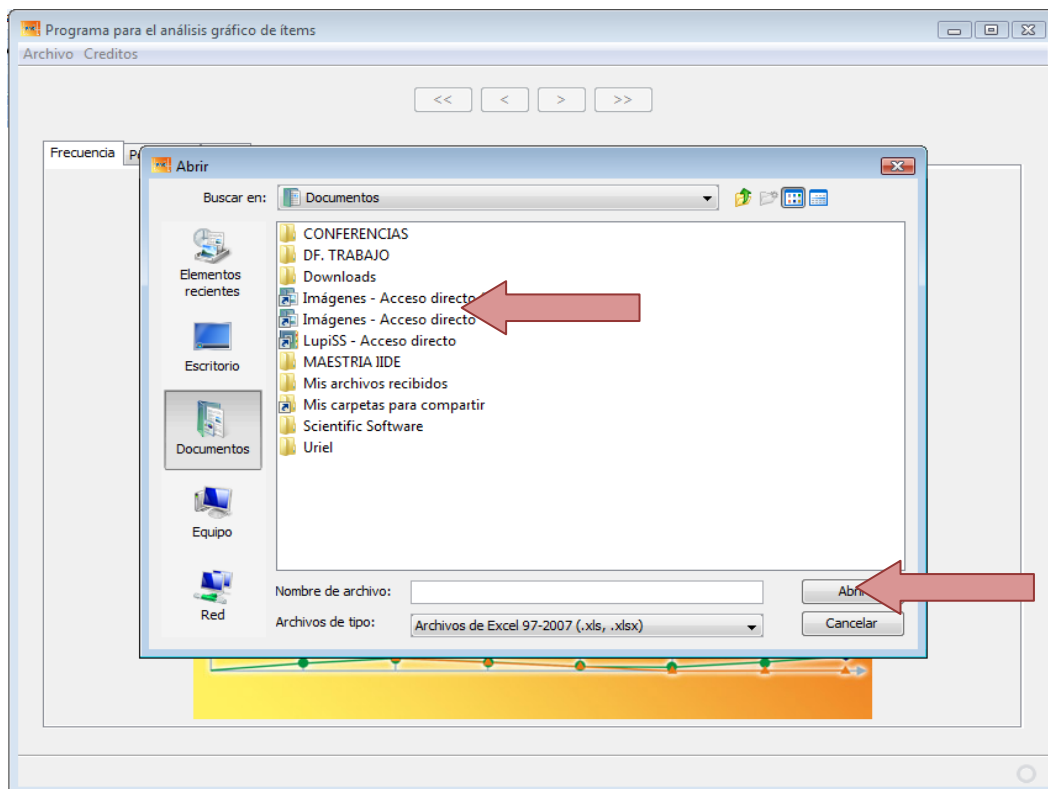


Figura 3.2 Entrada de datos del PAGI

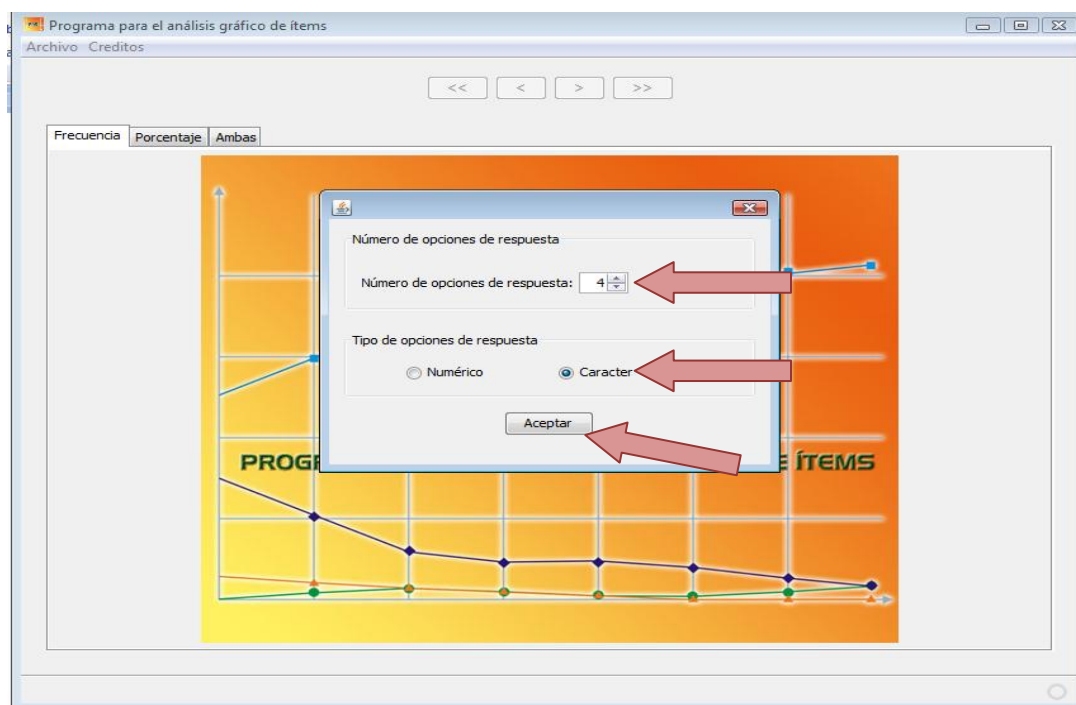
A continuación se abre una ventana que permite realizar la búsqueda del archivo, como se muestra en la figura 3.3.



**Figura 3.3** Búsqueda de archivo

Una vez localizado, se selecciona el archivo en formato .xls o .xlsx, según la versión de Microsoft Office Excel 2003 o 2007 que se haya utilizado para crear el archivo.

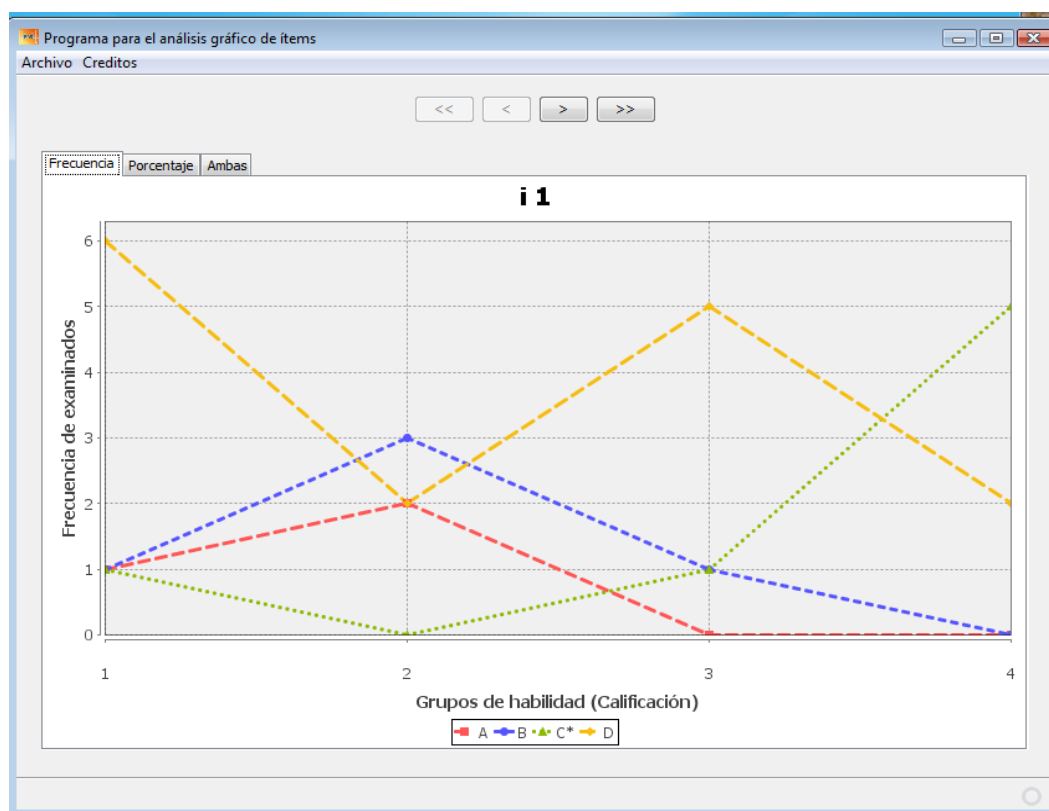
Posteriormente, se da clic en el control Abrir, como se muestra en la figura 3.4. Si el archivo tiene el formato adecuado abrirá una ventana como la que se muestra enseguida:



**Figura 3.4** Ventana de configuración detallada

En la figura 3.4 se solicita el número de opciones de respuesta que tiene cada ítem, que generalmente son 4 o 5. A continuación se indica el tipo de opciones de respuesta que se ofrecieron al examinado, mismo que puede ser “Caracter” en caso de que las opciones de respuesta hayan sido letras (A, B, C, D), o “Numérico” si fueron números (1, 2, 3, 4). Una vez seleccionados ambos elementos de configuración se da clic en “aceptar”.

Inmediatamente después aparece nuevamente la ventana principal de PAGI (ver figura 3.5), pero ahora ya se pueden visualizar las gráficas necesarias para efectuar el análisis gráfico de cada ítem. Por omisión, PAGI presenta primero la gráfica con frecuencia.

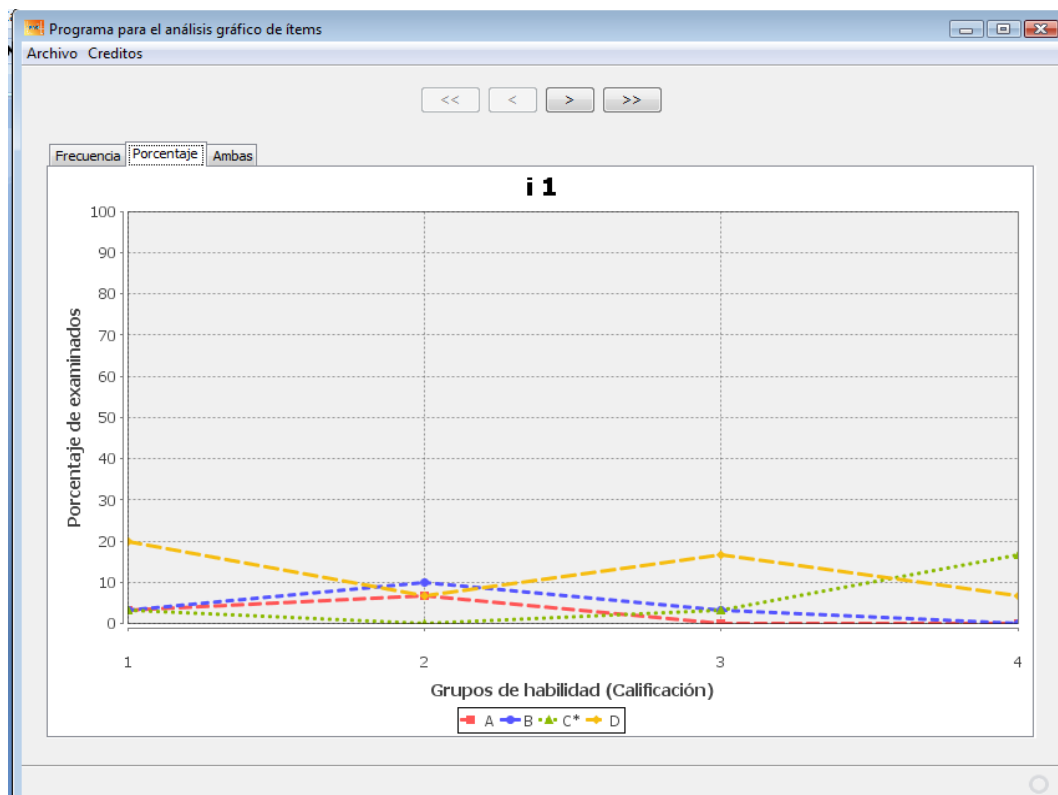


**Figura 3.5** Gráfica con frecuencia de examinado

En la parte superior de la figura 3.5 aparecen los controles << < > >> que sirven, como ya se dijo, para navegar entre los ítems del examen. Al dar clic en el signo > se muestra el siguiente ítem y así sucesivamente los demás. PAGI es flexible y permite al usuario regresar o avanzar a observar los ítems anteriores o posteriores por medio de estos controles. Las

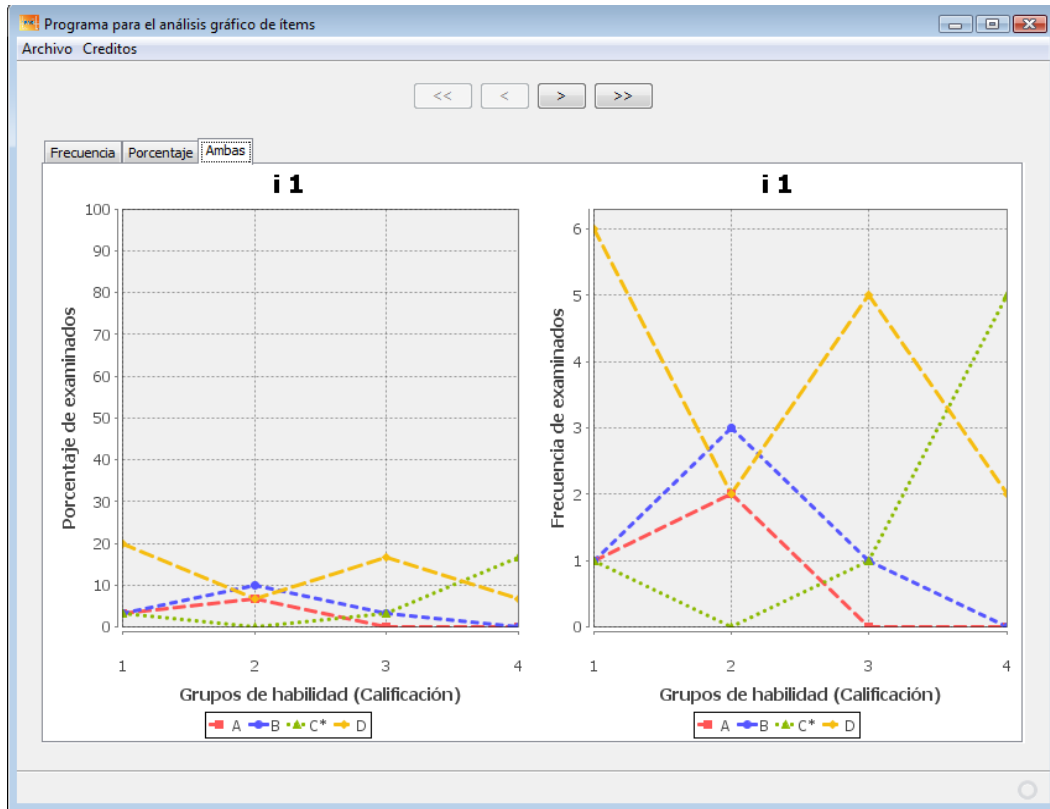
flechas dobles que se muestran permiten al usuario ir de manera directa al ítem final o al inicial.

La figura 3.6 muestra la gráfica con porcentaje de examinados. Así, se puede observar las similitudes y diferencias entre las gráficas de frecuencia y porcentaje de examinados que seleccionaron cada opción.



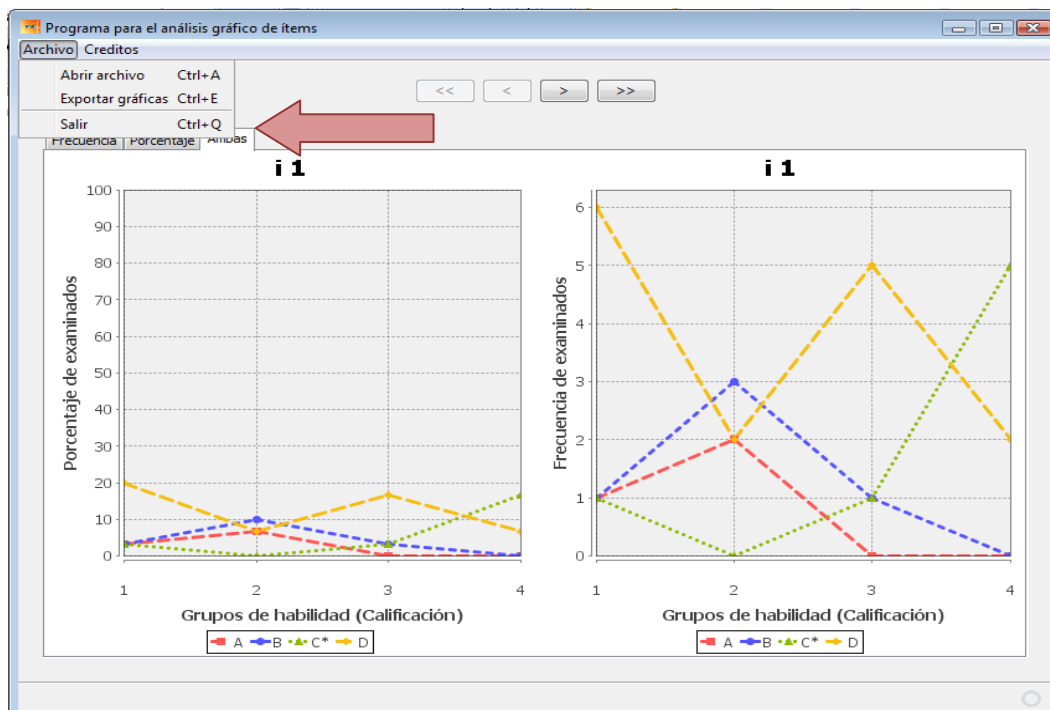
**Figura 3.6** Gráfica con porcentaje de examinados

En la última pestaña (figura 3.7) se muestran juntas ambas gráficas: la de frecuencia y la de porcentaje. Esta gráfica compuesta permite al usuario observar al mismo tiempo y comparar la visión panorámica de las curvas de respuestas que se dieron en el ítem y la visión detallada de las mismas. Lo anterior se observa en la figura 3.7 que se presenta enseguida.



**Figura 3.7** Pestaña donde se muestran ambas gráficas

Una vez que se ha concluido el análisis y la interpretación de todos los ítems del examen, mediante el análisis gráfico de ítems, en el menú de “Archivo” se encuentra la opción Salir para finalizar el análisis, como lo muestra la figura 3.8



**Figura 3.8** Ruta para salir del programa

## 4. GLOSARIO

**Ítem.** Los ítems son comúnmente identificados como preguntas o reactivos. Por ejemplo, una prueba de lectura consiste en cinco pasajes de texto y cuatro preguntas deben ser contestadas sobre cada pasaje. Se concibe a las veinte preguntas como veinte ítems.

**Dificultad o índice de dificultad.** También conocido como Valor  $p$  del ítem. Es importante tener en cuenta la dificultad de los ítems ya que en las respuestas aportadas por los estudiantes no se proporciona su nivel de habilidad. Para ello, la inclusión o exclusión de ítems está basada en su grado de dificultad. Se llama valor  $p$  porque se trata de la proporción o probabilidad de acertar el ítem.

Para ítems binarios (los calificados 0/1; es decir, correcto o incorrecto), el valor  $p$  de un ítem es la proporción de respuestas correctas en la población examinada, y se obtiene dividiendo el número de examinados que tuvieron correcto el ítem entre el total de quienes lo respondieron. Se considera el valor  $p$  como una propiedad del ítem respecto a una sola población. Por ejemplo, se desarrolla un ítem para una prueba de cuarto grado de español que será aplicada. Asumimos que el ítem es fácil porque tiene un valor  $p$  de 0.8, lo que quiere decir que el contenido es comprendido con facilidad. Incluso puede ser contestado por la población de segundo grado, donde su valor  $p$  es de 0.25. Así, se considera el valor  $p$  en referencia a alguna población.

Nota 1. Los valores  $p$  son valores que pertenecen a ítems en alguna población y se analizan en una muestra. La muestra no es igual que la población. Si se computa el valor  $p$  de un ítem en dos muestras independientes, se encontrará dos valores diferentes. El valor  $p$  encontrado en la muestra es considerado como una estimación del valor  $p$  en la población, para ello la exactitud depende del tamaño de la muestra.

Nota 2. El valor  $p$  se interpreta como medida de dificultad; sin embargo, mientras más alto sea el valor  $p$  el ítem es más fácil. Por ello es conocido también como el índice de facilidad.

**Índice de discriminación del ítem.** El índice de discriminación de un ítem permite separar los niveles de habilidad alta de los niveles bajos con base en las respuestas al ítem. En términos psicométricos define cuál es la calidad psicométrica de una prueba que tiene este ítem en particular. Por ejemplo, se usa un ítem binario difícil en una prueba, por ello se observa que el ítem discrimina a los estudiantes que tienen el ítem correcto de los que no. Pero el ítem binario tiene sólo dos categorías

(correcto e incorrecto), si el ítem separa a los buenos de los otros, no separa a los estudiantes con habilidad media de los débiles.

La discriminación es una propiedad local. Su poder discriminativo en el ítem se representa con un solo número. Para ello, en la TCT existen varios índices de discriminación:

1. La diferencia entre el índice de dificultad del reactivo para el grupo alto, formado por los examinados que obtuvieron las calificaciones más altas en la prueba, digamos el 27% de ellos, y el índice de dificultad del ítem para el grupo bajo, formado por los examinados que obtuvieron las calificaciones más bajas en la prueba, digamos el 27% de ellos (discriminación mediante grupos contrastados).
2. La correlación entre la puntuación en el ítem y la puntuación total en la prueba (la correlación ítem – test).
3. La correlación entre la puntuación en el ítem y la puntuación total en la prueba con ese ítem excluido (correlación ítem-resto).
4. En particular para los ítems de opción múltiple: la correlación entre la puntuación en la prueba y cada uno de los distractores (la correlación opción – total).

**Correlaciones entre los distractores y los puntajes en la prueba.** La correlación biserial puntual está compuesta por un coeficiente de correlación que debe ser positivo; esto quiere decir que los estudiantes que seleccionaron la opción correcta deben obtener los puntajes totales más altos. En el caso de los distractores, su correlación con la puntuación total en la prueba debe ser negativa, es decir, los estudiantes que seleccionan las opciones incorrectas son quienes obtienen un puntaje total bajo en la prueba. Este hecho, diferencia a los que saben de los que no (Van Batenburg y Laros, 2002). Para calcular la correlación se necesitan dos series de puntajes. Por ejemplo: al calcular la correlación ítem-test. Una puntuación es el puntaje obtenido por los examinados en la prueba y la otra puntuación es la obtenida por los examinados en el ítem.

La última es igual a uno, si la respuesta es correcta; y es cero si la respuesta es incorrecta. Para calcular la correlación entre un distractor y el puntaje en la prueba, se recodifican las respuestas dadas en la prueba por los examinados. Suponiendo que el ítem bajo estudio es uno de opción múltiple con cuatro opciones de respuesta (A, B, C y D), y la opción B es la respuesta correcta; esto significa que se da un puntaje en el ítem de uno a cada examinado que escogió B, y un puntaje de cero a los demás. Para calcular la correlación entre el puntaje en la prueba y el distractor A, se debe crear una variable binaria nueva, y asignar un puntaje de uno a cada examinado que escogió A y

poner un cero a los demás. La correlación que se busca es la correlación entre una nueva variable y la puntuación obtenida por los examinados en la prueba. Por eso, se da a conocer la necesidad de almacenar las observaciones originales o puntajes brutos. Si uno guarda sólo los puntajes en el ítem (los ceros y unos), no es posible calcular la correlación entre el distractor y el puntaje en el test, pues es imposible saber cuál de los distractores fue escogido a partir del conocimiento de que la respuesta no es correcta.

## 5. REFERENCIAS

- Batenburg, T. y Laros, J. (2002). Data Screening and Graphical Analysis of Items *Universidad de Groningen, en Holanda y Universidad de Brasilia, Brasil.*
- Batenburg, T. y Laros, J. (2002). Graphical Item Analysis. *Education Research and Evaluation (8) 3*
- Contreras, L. A. (2001). Procedimiento para Asegurar la Validez de Contenido de una Prueba Criterial: el Caso de un Examen de Español para la Educación Primaria en Baja California. *Memoria del VI Congreso Nacional de Investigación Educativa.* Manzanillo, México. COMIE. 2001.
- Encinas, J. A., Rivera, R. E. y Contreras, L. A., (2005). Evaluación colegiada del aprendizaje en la Universidad Autónoma de Baja California: Construcción de un examen criterial de gran escala para evaluar el dominio de conceptos y procedimientos del cálculo diferencial. *Memoria del VIII Congreso Nacional de Investigación Educativa.* México. COMIE.
- Verhelst, N. (2004). Sección C. Teoría clásica de los tests, en reference Supplement to the Preliminary Pilot version of the Manual for Relating Language examinations to the Common European Framework. Strasbourg. Language Policy Division.