

¿Cómo puedo mejorar mis exámenes? Algunas ideas.

Nuria Guasch García
Óscar Soler Canela
Regent's College London

1. Introducción

En la mayoría de los contextos educativos donde se imparten clases de E/LE, los docentes tienen que diseñar instrumentos de evaluación. Entre estos instrumentos se encuentran los exámenes. Con frecuencia, los profesores no saben qué procedimientos les pueden ayudar a evaluar y mejorar sus exámenes. En este taller se intenta dar herramientas a los profesores para investigar la calidad de sus exámenes.

2. Consideraciones generales

No todos los contextos educativos son iguales. Sin embargo, hay algunas consideraciones generales que se pueden aplicar a todos ellos independientemente de sus particularidades.

La consideración más general de la que podemos partir, y que a su vez es fundamental, es el documento donde se encuentran las especificaciones de examen. Las especificaciones son un documento fijado por cada centro de lenguas, y donde se especifica detalladamente cómo debe ser un examen.

En estos centros se suele cargar con la responsabilidad de hacer un examen al profesor. Uno de los problemas que esto puede acarrear, es que el profesor encargado no tenga tiempo y se dedique a mirar exámenes anteriores como modelo a seguir, limitando así el abanico de contenidos que se pueden incluir en el nuevo examen (ya que a lo mejor en ese modelo se han dejado partes del programa sin evaluar); También existe el peligro de reproducir vicios incluidos en exámenes anteriores. Para evitar estos y más problemas, existen las especificaciones, que es un documento de unas dos o tres páginas en el que se explican, por ejemplo:

- los tipos de ítems que deben entrar en el examen,
- cómo se debe evaluar,
- cuántos textos tiene que haber en la comprensión lectora,
- los contenidos que hay que cubrir
- a quién va dirigido el examen

- etc.

Es un documento cuya labor consiste en facilitar la labor del profesor a la hora de hacer un examen.

Una vez leídas las especificaciones, sí que podemos centrarnos en la redacción o en el análisis de los ítems de nuestros exámenes.

3. Tipos de ítems

Un ítem es cualquier parte de un examen a la que se le puede dar cierta nota.

Hay diferentes formas de clasificar los ítems según el tipo de respuesta que se espera de los estudiantes:

a) Ítems de respuesta seleccionada:

Los estudiantes tienen que seleccionar unas opciones que se les dan. Estos ítems se suelen usar para evaluar destrezas de comprensión. Algunos ítems de este tipo son los ítems de emparejamiento, donde hay que relacionar dos palabras, o frases, etc, o los ítems de selección múltiple, en los que hay que marcar una de las opciones dadas.

b) Ítems de respuesta construida: son esos ítems en los que el alumno debe producir él mismo la respuesta, sin opciones previas dadas. Se suele usar para evaluar las destrezas productivas. Dos ejemplos serían los ejercicios de completación en los que no se da ninguna opción, y los ítems de respuesta corta, donde los estudiantes tienen que escribir una palabra, dos o una frase como respuesta a la pregunta hecha.

c) Ítems de respuesta personal: estos ítems son más flexibles, y entre ellos se encuentran el portafolio y la autoevaluación.

4. ¿Qué hacemos para analizar los ítems?

En la sección anterior hemos visto distintos tipos de ítems. En este taller nos vamos a centrar únicamente en cómo analizar los ítems de respuesta seleccionada, y, en concreto, ítems de respuesta múltiple.

Existen dos tipos de procedimientos que se pueden seguir para analizar un ítem:

- a. Procedimientos de juicio.
- b. Procedimientos empíricos

a. Procedimientos de juicio

Son los procedimientos que hacemos las personas. A través de ellos nos preguntamos si lo que se mide en el ítem es lo que realmente se debería estar midiendo, si el ítem está bien formulado, si las opciones para la respuesta son adecuadas, etc.

Cuando nos referíamos a las especificaciones de un examen, comentábamos cómo en algunas ocasiones se copiaban o seguían modelos anteriores. En algunos casos existe la tendencia a copiar exámenes sin preguntarse qué es lo que estamos evaluando. Esta pregunta es fundamental, ya que dependiendo de lo que se quiera evaluar, los ítems que tenemos pueden servir o no para esa función. Un ítem puede ser muy bueno, pero no darnos información sobre lo que el examen está evaluando, por tanto, para ese tipo de examen no será útil.

Al mismo tiempo, hay que tener en cuenta la opinión de otros evaluadores, ya que para el redactor de un ítem no es siempre fácil mantener la distancia, y lo que al redactor le parece que está evaluando un ítem no siempre es obvio para los demás evaluadores.

Para seguir estos procedimientos, nos podemos hacer muchas preguntas, por ejemplo:

- ¿Qué está evaluando el ítem?
- ¿El ítem evalúa la inteligencia?
- ¿Hay más de una respuesta correcta (o ninguna respuesta) correcta?
- ¿Hay distractores que no distraen?
- ¿Hay problemas de registro?

b. Procedimientos empíricos.

Hay dos procedimientos estadísticos fundamentales que se siguen para analizar los ítems siguiendo estos procedimientos de forma empírica:

- A. Calcular el índice de facilidad
 - B. Calcular el índice de discriminación
- A. El índice de facilidad nos muestra si un ítem es demasiado fácil difícil y representa el número de alumnos que lo contesta correctamente. Para calcularlo se divide el número total de estudiantes que han contestado correctamente un ítem, y se divide entre el número total de alumnos que han hecho el examen. Esto nos da inmediatamente una idea de lo fácil (o difícil) que es el ítem.

El índice de facilidad abarca los porcentajes de 0 a 100. Un valor del 100% para un ítem significa que es muy fácil, ya que todos los alumnos lo han contestado bien; por el contrario, un valor del 0% representa un ítem muy difícil, al que nadie ha contestado bien. Eso significa que este ítem no será capaz de darnos mucha información sobre los candidatos que están haciendo el examen.

- B. El Índice de Discriminación: para algunos exámenes, nos interesa que los ítems separen a los estudiantes que más saben de los que menos saben. El I.D. muestra cómo discrimina un ítem, es decir, cómo distingue entre los alumnos de distintos grados de capacidad, y por tanto, si realmente nos ayuda a hacer esa diferenciación.

Para calcularlo, primero se ordena a los estudiantes según su puntuación.

Después se divide a los estudiantes en 3 grupos: alto, medio y bajo. Es imprescindible que el número de estudiantes del grupo alto y del grupo bajo sean iguales.

Finalmente se calcula el índice de facilidad del grupo alto y del grupo bajo, y se restan ambos: $I.D. = I.F. \text{ grupo alto} - I.F. \text{ grupo bajo}$

El índice de discriminación cubre también las puntuaciones del 0 al 100. El mayor índice de discriminación posible es de 100% y se consigue si todos los alumnos del grupo más alto dan la respuesta correcta, y ninguno del grupo bajo responde correctamente.

Los evaluadores piensan que lo mejor es que el índice de discriminación de un ítem se acerque al 50%. Si está por encima del 80% se suele considerar como demasiado fácil y si está por debajo del 20% demasiado difícil. En lugar de expresarlo en porcentajes, es frecuente encontrarlo en decimales. En el siguiente cuadro podemos encontrar una breve descripción de los ítems según su valor:

Si el I.D es/está entre...

>0,40 (muy buenos, funcionan muy bien)

0,30-0,39 (buenos pero mejorables)

0,20-0,29 (regulares, hay que mejorarlos)

<0,19 (malos)

Si el I.D. es negativo hay que revisar el ítem porque quiere decir que lo están contestando bien los estudiantes del grupo bajo, y sin embargo los del grupo alto/bueno no.

5. Conclusión

Antes de empezar a redactar un examen, es fundamental tener especificaciones de examen y seguirlas, especialmente si se trabaja en un centro con muchos profesores. Las especificaciones de examen son la manera de conseguir que todos los profesores produzcan exámenes en la misma dirección, y no siguiendo su propio criterio, ya que esto último podría resultar en un examen que no evalúe lo que el centro necesita.

Una vez producido el examen es necesario hablar y discutir con los profesores lo que se ha hecho para ver que no haya errores, y seguir los procedimientos de juicio necesarios antes de pilotar el examen. Una vez pasada esta fase, se debería pilotar los ítems con estudiantes del mismo nivel que el de los estudiantes reales a los que va a ir dirigido el examen.

Tras pilotar el examen, tenemos herramientas estadísticas objetivas (los procedimientos empíricos) que nos pueden ayudar a revisar los ítems, y ver si los distractores están cumpliendo la función que deben. En el caso de que falte acuerdo sobre un ítem, estos datos estadísticos nos pueden ayudar a tomar una decisión y resolver el desacuerdo.

6. Bibliografía

- Alderson, J.C., Claphan, C., y Wall, D. (1998) *Exámenes de idiomas.* Cambridge University Press, Madrid.
- Brown, J.D., y T. Hudson (2002) *Criterion-referenced Language Testing.* Cambridge University Press, Cambridge.
- Brown, J.D., (2005) *Testing in Language Programs.* McGraw Hill, New York.