

## EL USO DE LOS MAPAS CONCEPTUALES EN LA CORRECCION DE PREGUNTAS ABIERTAS EN BIOLOGIA

NIEDA OTERINO, J., DIAZ MUJICA, M<sup>a</sup> V., GARCIA BARQUERO, P., ORTEGA NUÑEZ, P., BONILLA MANGAS, I., AGUIRRE DE CARCER, I.  
Universidad Autónoma de Madrid.

---

### SUMMARY

The use of open-ended questions to assess student knowledge on a particular subject is discussed. Metric characteristics of those questions are analyzed in relation to their desirable characteristics for formative and summative evaluation. A procedure for rating open-ended questions using «concept maps» is described providing data on its content validity and reliability. Ways of using the procedure in formative evaluation are also described. As an application the procedure was used to analyze the reliability of Biology exercises (Universidad Autónoma Madrid Entrance Examinations).

---

La pregunta abierta es, sin duda, el tipo de pregunta utilizado con mayor frecuencia en las pruebas de evaluación de conocimientos. El estímulo consiste en una palabra o frase que designa genéricamente una sección de los contenidos estudiados; el examinado, en un tiempo limitado, debe redactar una respuesta que muestre sus conocimientos acerca de esos contenidos. La calificación de este tipo de preguntas presenta problemas de fiabilidad: diferentes examinadores suelen adjudicar calificaciones distintas a un mismo ejercicio; un solo examinador asigna calificaciones distintas al mismo ejercicio en diferentes ocasiones; las calificaciones son menos fiables cuando el enunciado de la pregunta permite una interpretación amplia (Coffman, 1971).

La puesta a punto y difusión de otras técnicas de evaluación con menos deficiencias desde el punto de vista métrico (p.e. las preguntas de elección múltiple), no ha logrado disminuir su uso generalizado. A ello, posiblemente haya contribuido la dificultad y el tiempo de preparación requerido para construir pruebas con distinto formato, y la utilización de pruebas con preguntas abiertas en los escasos exámenes «externos» que aún perduran en el sistema educativo (p.e. exámenes selectividad). En el panorama actual nada indica que esta situación vaya a cambiar por lo que estimamos útil el estudio de procedimientos que mejoren sus características métricas.

### 1. LAS PREGUNTAS ABIERTAS

Las condiciones de uso de este tipo de preguntas en nuestro sistema educativo no han sido estudiadas con el detalle necesario, por lo que basándonos en nuestra

experiencia aventuramos la siguiente hipótesis de trabajo. La capacidad de síntesis, la identificación de los conceptos y principios relevantes, y la capacidad de expresión escrita son las dimensiones principales que evaluamos al juzgar las respuestas proporcionadas a este tipo de preguntas.

En ocasiones, se identifica, erróneamente, las preguntas abiertas con las preguntas de ensayo al poseer ambos formatos similares. Sin embargo, los puntos de coincidencia sólo son superficiales: por prueba de ensayo se conoce «a aquellas preguntas que requieren que la persona que se examina construya una respuesta mediante una serie de frases, de tal modo que no se pueda enumerar de antemano cuáles de ellas o qué conjunto de frases constituyen una respuesta correcta» (Stalnaker, 1951).

A nuestro juicio, en las preguntas abiertas se puede identificar un conjunto de proposiciones que definen la contestación correcta, aunque ésta pueda redactarse empleando frases muy variadas.

Las características de las pruebas de evaluación deben estar estrechamente relacionadas con los usos que se pretendan hacer de la información que se obtenga. Tradicionalmente se distingue entre un uso formativo y un uso sumativo de esa información. Cuando utilizamos los resultados de una prueba para informar al alumno de los errores y aciertos contenidos en su respuesta o examinamos los resultados de un grupo de estudiantes buscando los contenidos para los que la instrucción proporcionada ha sido ineficaz, utilizamos los resultados de un modo formativo. Si la información es utilizada

para clasificar a los estudiantes en diversos grados de conocimientos haciendo corresponder a cada grupo una calificación o nota, o bien tratamos de formarnos un juicio sobre la calidad del programa de enseñanza examinando los resultados finales del grupo, la evaluación que realizamos es sumativa. En ambos casos, el procedimiento de evaluación debe ser válido y fiable. En el caso de la evaluación formativa (aquella que deberíamos utilizar a lo largo del curso académico), además, es imprescindible proporcionar una información precisa de los aciertos y deficiencias encontradas en cada respuesta.

Al examinar las respuestas a preguntas abiertas, generalmente utilizamos un procedimiento de calificación global. Mediante la lectura de la contestación redactada por el estudiante, tratamos de percibir el grado en que se ajusta a lo que consideramos una representación cabal de los conocimientos impartidos, incluibles bajo el título de la pregunta, y detectar expresiones que reflejen ideas erróneas. El resultado de esta actividad es un juicio que expresamos mediante una calificación y, quizás, algunos signos que identifican expresiones erróneas o faltas de ortografía. La posibilidad de utilizar esta información de un modo formativo es bastante limitada, ya que no comunica al examinando sus aciertos y deficiencias. Si el evaluador domina los contenidos podemos confiar en la validez del procedimiento, sin embargo la fiabilidad del mismo es discutible.

El procedimiento de corrección de preguntas abiertas que planteamos en este trabajo, está basado en una representación del contenido conceptual expuesto en el ejercicio y en la valoración de ese contenido utilizando un baremo prefijado. Se trata de una sistematización del procedimiento más o menos difuso que utilizamos los profesores. Esta sistematización proporciona al procedimiento de corrección todos los atributos que hemos enunciado como deseables para los distintos fines que puede perseguir este tipo de pruebas.

En lo que sigue describiremos el método de corrección con cierto detalle, examinaremos los datos disponibles sobre su validez y fiabilidad, del programa, acerca de sus conocimientos y de la eficacia del programa, respectivamente. Por último, a modo de aplicación, expondremos brevemente los resultados que obtuvimos al examinar con este método la fiabilidad de las calificaciones que se otorgaban a los ejercicios de biología en las pruebas de acceso en la Universidad Autónoma de Madrid.

## 2. PROCEDIMIENTO DE CORRECCION

El contenido informativo de las ciencias experimentales, y en concreto de la biología, posee una estructura conceptual. Esta estructura está formada por una serie de conceptos y unas relaciones entre dichos conceptos. Cualquier especialista en la disciplina puede identificar, ante un tema específico:

- a) Los conceptos biológicos pertinentes

- b) La jerarquía existente entre esos conceptos.
- c) Las relaciones existentes entre los conceptos.

Estas características nos permiten construir una representación conceptual de la contestación óptima a la pregunta abierta formulada, teniendo en cuenta el nivel de conocimientos exigibles a cada curso y el tiempo de que dispone el estudiante para elaborar la respuesta. Para representar el contenido de esa respuesta utilizamos un «mapa conceptual».

Un «mapa conceptual» es una representación bidimensional de una disciplina o de parte de una disciplina (Stewart et al 1979). Su carácter bidimensional permite representar las relaciones proposicionales (principios) entre los conceptos.

### 2.1. Construcción de un mapa conceptual para la pregunta abierta

Para obtener una respuesta óptima a la pregunta «La reproducción alternante en vegetales» que utilizamos para extraer los datos de este estudio, los miembros del equipo especialistas en biología, redactaron individualmente un texto que representaba una contestación óptima, en tres cuartos de hora, para el nivel de C.O.U. Los distintos textos fueron discutidos y reformados hasta alcanzar un acuerdo. Siguiendo las indicaciones de Novak (Novak, 1980), a partir de este texto:

- a) Identificamos los conceptos científicos relevantes.
- b) Jerarquizamos y ordenamos los conceptos desde los más generales a los más específicos.
- c) Utilizamos líneas para conectar conceptos, escribiendo una expresión que resumiera la relación existente entre cada par de conceptos relacionados.
- d) Intentamos que el mapa resultante fuese visualmente eficaz, es decir, que mostrara los detalles de los conceptos y sus relaciones de un modo consistente y eficiente.

El mapa conceptual elaborado para calificar las contestaciones de los estudiantes a nivel de C.O.U. está dibujado en la Figura 1

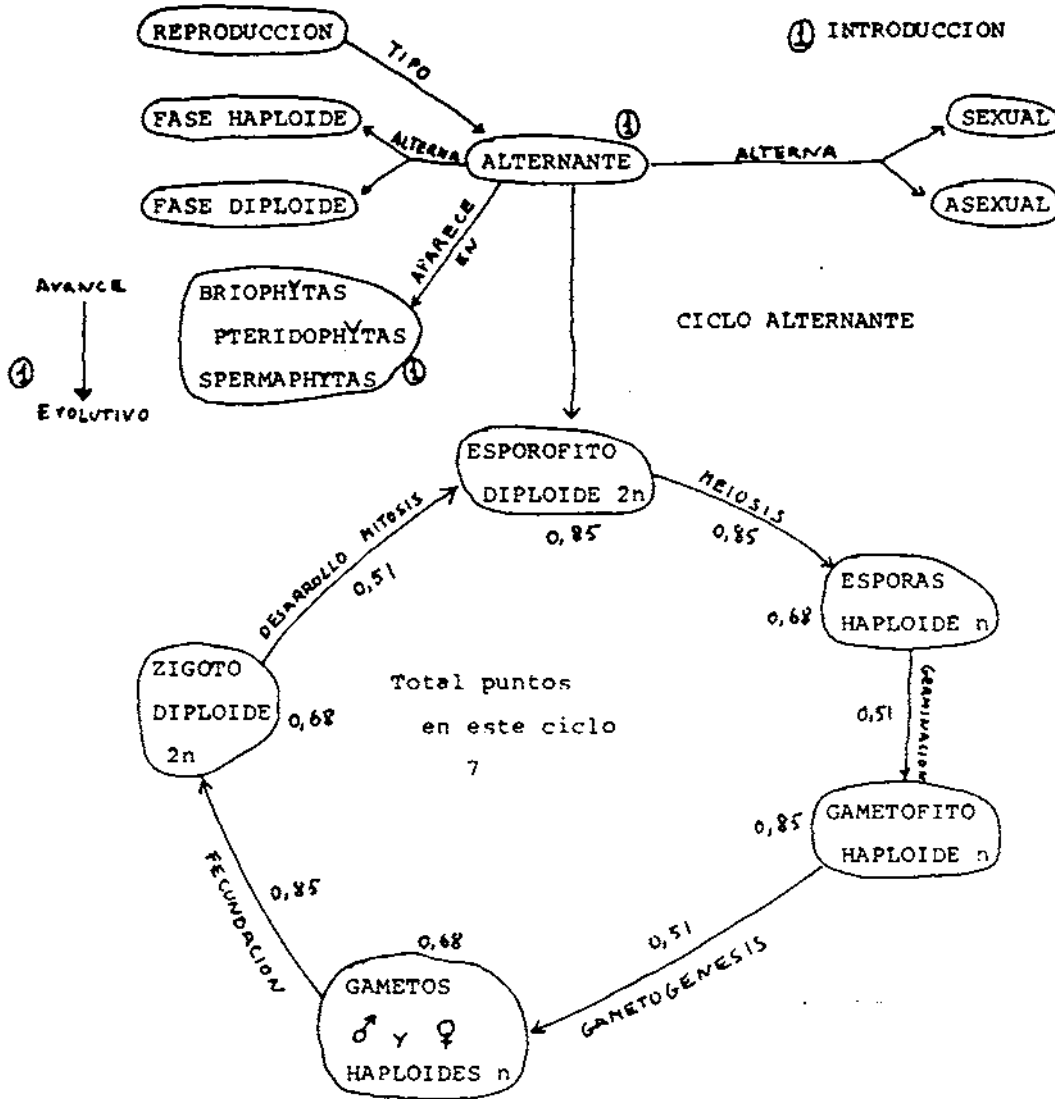
### 2.2. Baremación del mapa conceptual

En una contestación a una pregunta abierta no todos los contenidos pertinentes tienen la misma importancia. Al baremar el mapa conceptual asignando distintas puntuaciones parciales a cada uno de los contenidos pretendíamos reflejar la importancia relativa de cada uno de ellos en el contexto de la pregunta.

Para ello, utilizando una escala de 1 a 5, los especialistas en la materia asignaron puntuaciones a cada uno de los conceptos y relaciones expresadas en el mapa. Las puntuaciones otorgadas a cada elemento del mapa fueron promediadas y transformadas a una escala en la que la puntuación máxima fuera diez. Los distintos números de la figura 1 corresponden a las puntuaciones parciales de cada elemento en la escala de ba-

Figura 1

MAPA CONCEPTUAL: REPRODUCCION ALTERNANTE EN VEGETALES (COU)



remación aditiva, utilizada posteriormente, para calificar las contestación de los estudiantes.

Los conceptos y relaciones más importantes en esta pregunta en el nivel de C.O.U. pensamos que son los siguientes.

Conceptos	Puntuación	Relaciones	Puntuación
Esporófito	0.85	Meiosis	0.85
Gametofito	0.85	Fecundación	0.85
Espora	0.68	Germinación	0.51
Gameto	0.68	Gametogénesis	0.51
Zigoto	0.68	Mitosis	0.51

El núcleo fundamental de la pregunta recibiría una puntuación máxima de siete puntos; la definición correcta de reproducción alternante, un punto; la definición de los grupos vegetales afectados, un punto; y la consideración de su significado evolutivo, un punto adicional.

2.3. La calificación de las respuestas

Como ya indicamos anteriormente esta calificación o corrección debemos realizarla de un modo congruente con el fin que perseguimos.

Si pretendemos un uso formativo, podemos multicopiar el mapa conceptual «patrón» hasta obtener tan-



Cuadro 2

Tema: Reproducción alternante en vegetales				
Tribunal	A	B	C	D
Nº ejercicios	17	30	4	5
Correlación	0.784	0.48	0.996	0.95
Tema: Propiedades de las proteínas				
Tribunal	A	B	C	D
Nº ejercicios	19	81		
Correlación	0.842	0.199		

Los resultados del cuadro 2 expresan claramente la diferente fiabilidad de las calificaciones otorgadas por los distintos tribunales de las pruebas de aptitud ante la misma pregunta. Así, la correlación con el tribunal A es bastante aceptable, mientras que la fiabilidad de las calificaciones otorgadas por el tribunal B es francamente anómala. El escaso número de ejercicios en los tribunales C y D no nos permite realizar inferencias más allá de la constatación de la alta correlación obtenida.

Indagando sobre las acusadas discrepancias existentes entre el tribunal B y el equipo de investigación averiguamos que en dicho tribunal no había ningún especialista en biología, y por el modo de calificar obser-

vado, intuimos que los ejercicios fueron evaluados por varios correctores.

Del análisis de estos resultados concluimos que se confirma la escasa fiabilidad con que se otorgan las calificaciones en estas pruebas (y por ende la falta de validez de las decisiones que se adoptan a partir de ellas) cuando en los tribunales no existen especialistas en los contenidos que evalúan.

La calificaciones poco fiables tienen una escasa validez (Isaac y Michael 1972, p. 87) por lo que resulta urgente utilizar métodos fiables en aquellas evaluaciones de conocimientos sumativas de las que se desprenden decisiones importantes para la vida de los estudiantes. A pesar de la importancia de las decisiones que se adoptan a partir de las calificaciones de la selectividad, son escasas las universidades (p.e. la Universidad Autónoma de Barcelona; Valdés del Toro, 1984; Martín Ayuso 1984) que adoptan medidas elementales, como la señalada, para facilitar la validez y fiabilidad de sus decisiones. Señalemos que a raíz de este estudio, la Universidad Autónoma de Madrid, incorporó, en las convocatorias de acceso de 1983 y 1984, una serie de medidas encaminadas a mejorar los métodos con que se evaluaban los conocimientos de los candidatos a la admisión.

REFERENCIAS BIBLIOGRAFICAS

COFFMAN, W.E., 1971. Essay Examination en R.L. Thorndike (Ed) *Educational Measurement*. (American Council on Education. Washington D.C.) p.p. 271-302.

ISAAC, S. y MICHAEL, W.B., 1972. *Handbook in Research and Evaluation. For education and the behavioral Sciences*. (R.K. Knapp. Publisher: San Diego. California).

MARTIN AYUSO, S., 1984. Errores e irregularidades en las calificaciones de la nota media del expediente académico y su importancia, en I. Aguirre de Cárcer (Ed.) *La Selectividad a Debate*, (Ediciones de la Universidad Autónoma de Madrid).

NIEDA OTERINO, J., DIAZ MUJICA, M.V., GARCIA BARQUERO, P., ORTEGA, P., BONILLA, I., AGUIRRE DE CARCER, I., 1984. La fiabilidad de las calificaciones en preguntas abiertas en Biología en I. Aguirre de Cárcer (Ed) *La Selectividad a Debate*. (Ediciones de la Universidad Autónoma de Madrid).

NOVAK, J., 1980. *Handbook for the Learning how to learn Program* (Cornell University. Depart Education).

ROBINSON, J.T., 1979. A critical look at grading and Evaluation Practices en M.B. Rowe (Ed). *What Research Says to the Science Teacher*. Vol. 2. National Science Teacher Association: Washington.

STALNAKER, J.M., 1951. The essay type of examination en E.F. Lindquist (Ed) *Educational Measurement*. (American Council on Education. Washington D.C.) p.p. 495-530.

ESTEWART, J., VANKIRK, J., ROWEL, R., 1979. Concept mapping; A tool for use in Biology teaching. *American Biology Teacher*, 41: 3, pp. 171-175.

VALDES DEL TORO, R., MARTIN AYUSO, S., 1984. Composición de los tribunales, selección de los correctores y sesgos de instrumentación en I. Aguirre de Cárcer (Ed.) *La Selectividad a Debate*. (Ediciones de la Universidad Autónoma de Madrid).