

# Buenas prácticas para el diseño y la calificación de los exámenes

Fuente: [http://crlt.umich.edu/P8\\_0](http://crlt.umich.edu/P8_0)

Adapted from CRLT Occasional Paper #24: M.E. Piontek (2008)

Center for Research on Learning and Teaching. Universidad de Michigan

## Contenido

|       |   |   |
|-------|---|---|
| 1     | Pautas para diseñar exámenes válidos y confiables .....                                   | 1 |
| 2     | Directrices generales para la elaboración de preguntas de opción múltiple y de ensayo ... | 3 |
| 2.1.1 | Preguntas de opción múltiple .....  | 3 |
| 2.1.2 | Pautas para escribir preguntas de opción múltiple .....                                   | 4 |
| 2.2   | Preguntas de desarrollo .....   | 5 |
|       | Pautas para la redacción de preguntas de ensayo.....                                      | 5 |
|       | Pautas para la calificación de las preguntas de ensayo .....                              | 6 |

La función más obvia de los métodos de evaluación (como exámenes, cuestionarios, trabajos y presentaciones) es permitir que los profesores emitan juicios sobre la calidad del aprendizaje de los estudiantes (es decir, asignar calificaciones). Sin embargo, el método de evaluación también puede tener un impacto directo en la calidad del aprendizaje de los estudiantes. Los estudiantes asumen que el enfoque de los exámenes y las tareas reflejan las metas educativas más valoradas por los docentes, y orientan su método de estudio en función del sistema de evaluación (McKeachie y Svinicki, 2006). Los sistemas generales de calificación también pueden tener un impacto. Por ejemplo, una estricta curva de campana (es decir, distribución normal de las calificaciones) tiene el potencial de amortiguar la motivación y la cooperación en el aula, mientras que un sistema que recompensa estrictamente la competencia entre estudiantes (es decir, calificación de referencia de criterio) podría percibirse como una contribución a la inflación de la calificación. Dada la importancia del sistema de evaluación de las interacciones entre el profesorado y los alumnos sobre el aprendizaje, ¿cómo pueden los docentes desarrollar exámenes que brinden datos útiles y relevantes sobre el aprendizaje de sus alumnos y al tiempo orientar a los estudiantes para dedicar su tiempo a los aspectos importantes la asignatura? ¿Cómo influyen las formas de evaluación en este proceso?

## 1 Pautas para diseñar exámenes válidos y fiables

Idealmente, un examen efectivo tiene cuatro características. Son:

- Válido (proporciona información útil sobre los conceptos para los que fue diseñado).

- Fiable (permite mediciones consistentes y discrimina entre diferentes niveles de rendimiento),
- Reconocible (durante la preparación de la asignatura, con el diseño del curso, el docente ha preparado a los estudiantes para ese tipo de evaluación), y
- Realista (sobre el tiempo y el esfuerzo necesarios para completar la tarea) (Svinicki, 1999).

Lo que es más importante, los exámenes y las tareas deben centrarse en el contenido y los comportamientos más importantes enfatizados durante el curso. ¿Cuáles son las principales ideas, problemas y habilidades que se espera que los estudiantes aprendan durante un curso / unidad / módulo en particular? Estos son los resultados de aprendizaje que se desea medir. Por ejemplo, si su resultado de aprendizaje implica la memorización, entonces debe evaluar la memorización o clasificación; Si espera que los estudiantes desarrollen capacidades de resolución de problemas, sus exámenes deben centrarse en evaluar las habilidades de aplicación y análisis de los estudiantes. Como regla general, las evaluaciones que se centran demasiado en los detalles (por ejemplo, hechos aislados, cifras, etc.) "probablemente conducirán a una mejor retención de las notas al pie por parte de los estudiantes a costa de los puntos principales" (Halpern y Hakel, 2003, p 40) Como se señala en la Tabla 1, cada tipo de examen puede ser más adecuado para medir algunos resultados de aprendizaje que otros, y cada uno tiene sus ventajas y desventajas en términos de facilidad de diseño, implementación y calificación.

Tabla 1: Ventajas y desventajas de los tipos de preguntas comúnmente utilizados

| <b>Tipo de ítem</b>                    | <b>Ventajas</b>   | <b>Desventajas</b>   |
|--|---|--|
| <b>Verdadero/<br/>Falso</b>            | Es posible plantear muchas preguntas en un tiempo relativamente corto. Moderadamente fácil de preparar. Fácil de calificar. Evaluación automática.  | Algunas preguntas serán fáciles de adivinar, incluso si, el estudiante no domina la materia .  |
| <b>Opción múltiple</b>                 | Se puede utilizar para evaluar una amplia gama de contenido en un breve período. Preguntas bien diseñadas pueden medir habilidades cognitivas de orden superior. Se corrige automáticamente. Fácil corrección | Escribir buenas preguntas es difícil y lleva mucho tiempo. Es posible evaluar habilidades cognitivas de orden superior, pero la mayoría de los ítems evalúan sólo el conocimiento. Algunas respuestas correctas pueden ser adivinadas. |
| <b>Emparejamiento<br/>o asociación</b> | Las preguntas que requieren asociar conceptos e ideas, pueden escribirse rápidamente. Se puede evaluar una amplia gama de contenidos. La calificación puede hacerse de  | Este tipo de pregunta no permite evaluar habilidades cognitivas de orden superior.   |

|  |  |   |
|--|--|---|
|  | forma eficiente, de manera automática  |   |
| <b>Preguntas cortas o de completar huecos</b>  | Pueden realizarse muchas preguntas en un corto período de tiempo. Relativamente eficiente para calificar. Moderadamente fácil de escribir. | La limitación tiene que ver con que solo es aplicable a preguntas que pueden responderse o completarse en muy pocas palabras  |
| <b>Preguntas de respuesta larga o trabajos</b> | Sirven para medir habilidades cognitivas de orden superior. Son relativamente fáciles de preparar.   | Requieren mucho tiempo para la corrección. Es complicado fijar criterios de corrección objetivos, si bien se pueden utilizar rúbricas. Permite evaluar partes limitadas del contenido de la materia, pues en un examen solo podrá preguntarse por partes limitadas del programa |

## 2 Directrices generales para la elaboración de preguntas de opción múltiple

En las secciones siguientes se recogen una serie de directrices generales para la elaboración de preguntas de opción múltiple y de ensayo, que suelen utilizarse en los exámenes de nivel universitario, ya que permiten medir fácilmente las habilidades cognitivas de orden superior (por ejemplo, la aplicación, la justificación, la inferencia, el análisis y la evaluación). Sin embargo, los profesores suelen tener dificultades para elaborar, aplicar y calificar este tipo de preguntas (McMillan, 2001; Worthen, et al., 1993).

### 2.1 Preguntas de opción múltiple

Tienen una serie de ventajas. En primer lugar, pueden medir diversos tipos de conocimientos, incluida la comprensión de la terminología, los hechos, los principios, los métodos y los procedimientos, así como la capacidad de los estudiantes para aplicar, interpretar y justificar. **Cuando se diseñan cuidadosamente, las preguntas de opción múltiple también pueden evaluar las habilidades de pensamiento de orden superior.**

Las preguntas de opción múltiple **son menos ambiguas que los ítems de respuesta corta**, por lo que proporcionan una evaluación más centrada en los conocimientos del estudiante. Los ítems de opción múltiple son preferibles a las preguntas de verdadero y falso por varios motivos: en los ítems de verdadero y falso, los estudiantes pueden puntuar por saber que una afirmación es incorrecta, sin saber qué es lo correcto. Los ítems de opción múltiple ofrecen una mayor fiabilidad que los ítems de verdadero y falso, ya que la oportunidad de adivinar se reduce con el mayor número de opciones. Finalmente, un profesor puede diagnosticar un malentendido o una redacción imprecisa de la pregunta analizando las opciones incorrectas elegidas por los estudiantes.

Una desventaja de los ítems de opción múltiple es que requieren desarrollar opciones incorrectas (distractores), pero plausibles, lo cual es difícil de lograr. Además, las preguntas de opción múltiple no permiten a los docentes medir la capacidad de los estudiantes para organizar y presentar ideas. Por último, como es mucho más fácil crear ítems de opción múltiple que prueben la memoria y el reconocimiento en lugar de un pensamiento de orden superior, los exámenes de opción múltiple corren el riesgo de no evaluar el aprendizaje profundo que muchos instructores consideran importante (Groenlandia y Linn, 1990; McMillan, 2001).

### 2.1.1 Pautas para escribir preguntas de opción múltiple

Estos son algunos consejos sobre la redacción de enunciados, las opciones de respuesta correctas y los distractores (McMillan, 2001, p. 150; Piontek, 2008):

- Los enunciados plantean claramente la pregunta.
  - ¿El enunciado está redactado de la manera más clara, directa y sencilla posible?
  - ¿El objeto de la pregunta se describe completamente en el enunciado?
  - ¿El enunciado está redactado en positivo sin incluir negaciones, para evitar la posibilidad de que los estudiantes pasen por alto términos como "no", "no" o "menos"?
  - ¿Proporciona el enunciado sólo información relevante para el problema, sin extensiones innecesarias?
- Las opciones de respuesta incluyen la respuesta correcta y los distractores ( opciones incorrectas). Las preguntas de opción múltiple suelen tener al menos tres distractores.
  - ¿Son los distractores creíbles para los estudiantes que no saben la respuesta correcta?
  - ¿Hay una sola respuesta correcta?
  - ¿Son todas las respuestas posibles similares en cuanto a estructura gramatical, longitud y complejidad?
  - ¿Son cortas las opciones de respuesta?
  - ¿Se evitan las opciones complejas? ¿Se colocan las opciones en un orden lógico?
  - ¿Se distribuyen las respuestas correctas por igual entre todas las opciones? (Por ejemplo, ¿es la respuesta "A" correcta aproximadamente el mismo número de veces que las opciones "B", "C" o "D")?. (Esto no sería aplicable si el orden de las opciones de respuesta fuese aleatorio)

Ejemplo de buenas preguntas de opción múltiple que evalúan las habilidades de pensamiento de orden superior. El ejemplo procede de un examen de farmacología (Park, 2008):

*El paciente WC fue ingresado por quemaduras de tercer grado en más del 75% de su cuerpo. El médico que lo atiende le pide que empiece a dar a este paciente una terapia de antibióticos. ¿Cuál de las siguientes es la mejor razón por la que WC necesitaría profilaxis antibiótica?*

- a. *Sus lesiones por quemaduras han roto la inmunidad innata que previene la invasión microbiana.*
- b. *Sus lesiones han inhibido su inmunidad celular.*
- c. *Sus lesiones han perjudicado la producción de anticuerpos.*
- d. *Sus lesiones han estimulado la médula ósea, activando así el sistema inmunológico*

Una segunda pregunta se basa en la primera al describir los resultados de los análisis del paciente dos días después, pidiendo a los estudiantes que desarrollen una explicación para los resultados de laboratorio posteriores. (Ver Piontek, 2008 para la pregunta completa.)

## 2.2 Preguntas de desarrollo

Este tipo de preguntas permite evaluar el pensamiento complejo al exigir a los estudiantes que organicen, integren e interpreten informaciones, construyan argumentos, den explicaciones, evalúen el mérito de las ideas y lleven a cabo otros tipos de razonamiento (Cashin, 1987; Gronlund & Linn, 1990; McMillan, 2001; Thorndike, 1997; Worthen, et al., 1993). Las **preguntas cortas** son buenas para evaluar el conocimiento y la comprensión básicos y generalmente requieren una breve respuesta por escrito (por ejemplo, "*Establezca dos hipótesis acerca de por qué las aves migran. Resuma la evidencia que apoya cada hipótesis*" [Worthen, et al., 1993, p. 277].) **Las preguntas de ensayo de respuesta larga** permiten a los estudiantes construir una variedad de estrategias, procesos, interpretaciones y explicaciones para una pregunta, como las siguientes:

*Los redactores de la Constitución se esforzaron por crear un gobierno nacional eficaz que equilibrara la tensión entre el gobierno de la mayoría y los derechos de las minorías. ¿Qué aspectos de la política estadounidense favorecen el gobierno de la mayoría? ¿Qué aspectos protegen los derechos de los que no son mayoría? Basándose en el material de sus lecturas y conferencias, ¿los redactores de la Constitución lograron equilibrar esta tensión? ¿Por qué o por qué no? (Shipan, 2008).*

Además de medir el pensamiento y el razonamiento complejos, las ventajas de los ensayos consiste en que potencian y actúan como motivación para desarrollar mejores hábitos de estudio y proporcionar a los estudiantes flexibilidad en sus respuestas. Los profesores pueden evaluar la capacidad de los estudiantes para comunicar y expresar su razonamiento. Normalmente, la preparación de este tipo de preguntas requieren menos tiempo para redactarlas que las preguntas de opción múltiple que miden el razonamiento.

Las principales desventajas de los ensayos tienen que ver con la cantidad de tiempo que los instructores deben dedicar a leer y calificar las respuestas de los estudiantes, y la dificultad de elaborar y utilizar de manera cuidadosa criterios/rúbricas, para asegurar la fiabilidad de la calificación. Los ensayos pueden evaluar sólo una cantidad limitada de contenido en un examen debido al tiempo que se requiere para que los estudiantes respondan a cada pregunta de ensayo. Como consecuencia, los ensayos o preguntas largas no proporcionan una buena muestra de los conocimientos sobre el conjunto del programa de la asignatura (Gronlund & Linn, 1990; McMillan, 2001).

### 2.2.1 Pautas para la redacción de preguntas de ensayo

Conviene tener en cuenta lo siguiente: (Gronlund & Linn, 1990; McMillan, 2001; Worthen, et al., 1993):

- Restringir el uso de las preguntas de desarrollo a evaluar los resultados de aprendizaje que son difíciles de medir utilizando otros formatos. Por ejemplo, para evaluar memorización de conocimientos, las preguntas de verdadero y falso, las que requieren rellenar huecos en blanco o las de opción múltiple son más adecuadas.
- Identificar las aptitudes y conocimientos específicos que se evaluarán. [Piontek \(2008\)](#) proporciona varios ejemplos de preguntas que evalúan diferentes tipos de habilidades de razonamiento, tales como:
  - Generalizaciones: *Enuncie un conjunto de principios que puedan explicar los siguientes eventos*
  - Síntesis: *Escribir un informe bien organizado que muestre...*
  - Evaluación: *Describe las fortalezas y debilidades de...*
- Escribir la pregunta claramente para que los estudiantes no sientan que están adivinando "lo que el instructor quiere que haga".
- Indicar la cantidad de tiempo y esfuerzo que los estudiantes deben dedicar preparar el ensayo trabajo.
- En caso de que el trabajo o ensayo incluya varias preguntas, evitar dar a los estudiantes opciones sobre qué preguntas del ensayo deben responder. Esta elección disminuye la validez y la fiabilidad del examen porque cada estudiante estaría realizando un examen diferente, al centrarse o no en determinadas preguntas.
- Considerar la posibilidad de desglosar el tema general del ensayo en varias preguntas concretas (en lugar de una pregunta amplia) esto permitirá evaluar diferentes habilidades y conocimientos de los estudiantes.
- Asegurarse de que haya suficiente tiempo para responder a las preguntas.

### 2.2.2 Pautas para la calificación de las preguntas de ensayo

Debería tenerse en cuenta lo siguiente (Gronlund & Linn, 1990; McMillan, 2001; Wiggins, 1998; Worthen, et al., 1993; *Escribir y calificar las preguntas de ensayo*, 1990):

- Describa lo más claramente posible que constituiría una respuesta correcta.
- Seleccione un método de puntuación apropiado en función de los criterios. Esto puede hacerse mediante una rúbrica o protocolo de evaluación que indica los criterios de puntuación y los puntos que se asignarán a cada criterio. A continuación, se incluye un ejemplo de rúbrica para una pregunta de examen de historia podría tener el siguiente aspecto:

| Criterio de evaluación  | Nivel de desempeño |                 |                 |
|-------------------------|--------------------|-----------------|-----------------|
|                         | <i>0 puntos</i>    | <i>1 puntos</i> | <i>2 puntos</i> |
| Número de referencias a | 0-2 referencias    | 3-5 referencias | 6+ referencias  |

|   |   |  |  |
|---|---|--|--|
| fuentes o publicaciones                 |   |  |  |
| Precisión histórica                     | Muchas inexactitudes  | Pocas inexactitudes                                    | No hay inexactitudes   |
| Argumentación histórica                 | No hay argumentos o utiliza una argumentación con escaso fundamento | La argumentación es vaga y poco apoyada por evidencias | La argumentación es clara y está bien apoyada por la evidencia |
| Corrección gramatical                   | Muchos errores de gramática y ortografía                            | Pocos errores (1 a 2) de gramática o de ortografía     | Ningún error de gramática o de ortografía                      |
| <b>Total de puntos (de 8 posibles):</b> |   |  |  |

Para otros ejemplos de rúbricas, ver CRLT Occasional Paper #24 (Piontek, 2008).

- Aclarar el peso de la evaluación de factores independientes de los contenidos educativos que se están midiendo. Por ejemplo, ¿cómo se valora la corrección gramatical y sintáctica o el uso de la notación científica?
- Para aumentar la precisión de la calificación, de ser posible, corregir los exámenes sin fijarse en la identidad del estudiante y ordenarlos los exámenes de manera aleatoria (por ejemplo, barajar el montón)
- En caso de que la prueba incluya varias preguntas, utilice un proceso sistemático para calificar cada pregunta del ensayo. Se sugiere que se califiquen todas las respuestas de una pregunta una detrás de otra, en lugar de calificar todas las respuestas de todas las preguntas de un estudiante concreto. Este sistema hace más fácil recordar los criterios para calificar cada respuesta.

También se puede utilizar rúbricas para trabajos de los estudiantes, presentaciones orales, proyectos de curso y sitios web. Para otras estrategias de calificación, vea [Responding to Student Writing – Principles & Practices](#) and [Commenting Effectively on Student Writing](#)

---

#### Referencias

- Cashin, W. E. (1987). Improving essay tests. Idea Paper, No. 17. Manhattan, KS: Center for Faculty Evaluation and Development, Kansas State University.
- Gronlund, N. E., & Linn, R. L. (1990). Measurement and evaluation in teaching (6ª ed.). Nueva York: Macmillan Publishing Company.
- Halpern, D. H., & Hakel, M. D. (2003). Applying the science of learning to the university and beyond. *Change*, 35(4), 37-41.

McKeachie, W. J., & Svinicki, M. D. (2006). Assessing, testing, and evaluating: La calificación no es la función más importante. En *McKeachie's Teaching tips: Strategies, research, and theory for college and university teachers* (12ª ed., págs. 74-86). Boston: Houghton Mifflin Company.

McMillan, J. H. (2001). *Classroom assessment: Principles and practice for effective instruction* (Principios y prácticas para una instrucción eficaz). Boston: Allyn y Bacon.

Park, J. (2008, 4 de febrero). Comunicación personal. Facultad de Farmacia de la Universidad de Michigan.

Piontek, M. (2008). [\*Best practices for designing and grading exams. CRLT Occasional Paper No. 24.\*](#) Ann Arbor, MI. Centro de Investigación sobre el Aprendizaje y la Enseñanza.

Shipan, C. (2008, 4 de febrero). Comunicación personal. Departamento de Ciencias Políticas de la Universidad de Michigan.

Svinicki, M. D. (1999a). Evaluando y calificando a los estudiantes. En *Maestros y estudiantes: A sourcebook for UT- Austin faculty* (pp. 1-14). Austin, TX: Center for Teaching Effectiveness, Universidad de Texas en Austin.

Thorndike, R. M. (1997). *Measurement and evaluation in psychology and education* (Medición y evaluación en psicología y educación). Upper Saddle River, NJ: Prentice-Hall, Inc.

Wiggins, G. P. (1998). *Evaluación educativa: Diseño de evaluaciones para informar y mejorar el rendimiento de los estudiantes*. San Francisco: Jossey-Bass Publishers.

Worthen, B. R., Borg, W. R., & White, K. R. (1993). *Measurement and evaluation in the schools* (Medición y evaluación en las escuelas). New York: Longman.

Writing and grading essay questions. (1990, septiembre). Para su consideración, No. 7. Chapel Hill, NC: Centro de Enseñanza y Aprendizaje, Universidad de Carolina del Norte en Chapel Hill.