

Cómo citar la publicación:
García Aretio, L. (12/05/2020). Diseño de una prueba de evaluación. *Contextos universitarios mediados*. (ISSN: 2340-552X), <https://aretio.hypotheses.org/4113>.

Diseño de una prueba de evaluación

Lorenzo García Aretio
UNED

En la evaluación de los aprendizajes muchos estudiantes y, también bastantes profesores pueden llegar a *dudar de la rigurosidad y justicia* de las decisiones finales basadas en dudosos procesos de evaluación, dada la ligereza de la proposición de algunas pruebas y, a veces, la arbitrariedad en la valoración que llevaríaN a que evaluadores diferentes de una misma actividad, pregunta o producto, ofreciesen valoraciones muy distantes en su resultado. Esas dudas pueden provenir de:

- no tener claros los *criterios*,
- el *desconocimiento* de estos criterios parte de los estudiantes,
- el *escaso rigor* y claridad en la redacción de las preguntas o actividades,
- la *falta de claridad* en saber qué proceso mental se quiere en realidad evaluar y,
- en qué *condiciones* se ha de realizar esa prueba.

Criterios en pruebas de evaluación

En función de estos peligros, les traslado unas propuestas que entiendo podrían mejorar la formulación de las pruebas que se proponen a los estudiantes (García Aretio, 1994, 2001).



Habría de tenerse inicialmente muy claro cómo responder a las cuestiones ¿Qué, quién, cuándo y cómo debemos evaluar? Se remite a la lectura previa y necesaria de esa entrada. Vamos a los restantes criterios.

- **Consideración de los objetivos**

En toda prueba se debe ajustar la redacción de las preguntas a todos y cada uno de los objetivos previamente requeridos, que debe entenderse que están enfocados al logro de las respectivas competencias. Habría de evaluarse, por tanto, lo que se pretendía que aprendiesen los estudiantes. Por eso, el contenido de la pregunta ha de ser siempre relevante y estar de acuerdo con los objetivos. Por tanto, debe prescindirse de minucias y de la erudición sin relevancia aunque el número de *ítems* tenga que ser elevado.

- **Proceso mental que se pretende evaluar**

Incluir en la prueba preguntas que evidencien el tipo de proceso mental que se pretende que muestre el estudiante (*describir, comparar, relacionar, analizar, sintetizar, evaluar, crear...*). Cuidar la redacción para que quede meridianamente claro qué es lo que se está requiriendo en la correspondiente pregunta.

- **Enunciado de las cuestiones**

Huir de enunciados o pies similares a los epígrafes que aparecen en el material estudiado, dando cierta originalidad a la pregunta, organizando la cuestión de manera diferente, para que el estudiante tenga que seleccionar adecuadamente. Resulta importante tratar de dificultar en lo que se pueda el *plagio* o intercomunicación sencilla, evitando preguntas simples o de fácil respuesta que, en todo caso, cualquier buscador podría facilitar en segundos. Aunque ha de cuidarse el ceñir las cuestiones o actividades estrictamente a lo que debe haber aprendido el estudiante, según objetivos deseados.

- **Tiempo concedido al estudiante para responder o resolver**

Calcular adecuadamente el tiempo medio requerido para responder a la prueba propuesta o a resolver la tarea. El estudiante debe tener garantía de poder responder/resolver satisfactoriamente las pruebas en el tiempo concedido.

- **Univocidad de la pregunta**

Cuidar la formulación de las preguntas, de forma que todos puedan entender lo mismo. La redacción de las cuestiones será clara y precisa. Habrán de evitarse redacciones que lleven al equívoco, preguntas con doble sentido, capciosas, confusas, etc., que desvirtúan con frecuencia la valoración adecuada de cada estudiante, según quien le pudiera corregir la prueba.

- **Instrucciones para cumplimentar la prueba**

Determinar con claridad las instrucciones que ha de seguir el protagonista de la evaluación para la correcta realización de la prueba. Dar orientaciones sobre:

- la *distribución del tiempo* en pruebas que se compongan de más de un tipo de instrumento o técnica de evaluación;
- *criterios* de valoración de las partes de la actividad, o prueba, o de las preguntas;
- *fórmula* que se aplicará, ponderación, si es el caso, penalización de errores, etc.;
- *instrucciones* concretas sobre la forma de responder, por ejemplo si son de bolígrafo y papel, rodear, cruzar, tachar, marcar, señalar, etc.;
- *otras* consideraciones adaptadas al curso en cuestión que eviten calificaciones que no respondan a la realidad de lo que sabe cada uno.

- **Estructura de la prueba**

Tipo de cuestiones según respuesta requerida (dedicará alguna entrada a este punto): pruebas de ensayo, respuesta libre, elaboración, respuesta breve y tasada, pruebas objetivas en sus múltiples posibilidades, etc.

Deberá atenderse, en el caso de instrumentos o técnicas diferentes en la misma prueba, a agruparlas con el fin de no generar confusión. Igualmente, se considerará el número de preguntas que se pretenden plantear, ponderación de las mismas en la valoración final de la prueba...

Si se han incluido pruebas objetivas, deberán tenerse en cuenta otras consideraciones que desarrollaré en otra entrada.

- **Cuidar las condiciones de la realización**

En pruebas presenciales, el silencio en el lugar de la prueba es requerido por la mayoría de los estudiantes que tienen problemas para concentrarse cuando existe un fondo de conversación o ruidos molestos. No se deberán atender las solicitudes individuales sobre interpretación de alguna cuestión.

Sólo si se detecta algún tipo de error en algún examen, la aclaración se hará pública, con el fin de no trastornar la calidad de la prueba ni atentar contra los principios de igualdad y justicia y la función didáctica del examen. El hecho de copiar, «soplar», etc. desvirtúan igualmente estos principios. En las pruebas presenciales cuando son muchos estudiantes, un momento clave para evitar este último problema, así como la posibilidad de suplantación de personalidad es el final de los exámenes por la confusión y desorden que habitualmente se generan al proceder a su entrega.

Los requisitos de una buena prueba de evaluación

Tras los criterios señalados, las actividades, trabajos y pruebas de evaluación deberían cumplir una serie de requisitos que podrían resumirse en estos cinco: *validez, fiabilidad, sensibilidad, objetividad y representatividad.*



- **Validez**

Para que una prueba responda a este requisito ha de medir sólo lo que debe medir, aquello para lo que ha sido elaborada. Una condición previa para asegurarnos que la prueba es válida, sería cerciorarnos si se lograron los *objetivos de aprendizaje* propuestos.

Además de la validez referida a los objetivos, podría hablarse de *validez curricular*, es decir, la adaptación de la prueba a las demás variables curriculares: contenidos, actividades, niveles, recursos empleados, etc. Otra validez sería la *psicológica* que exigiría que la prueba estuviese preparada de tal manera que se ajustase a las condiciones mentales del examinando: horario, duración, lugar de realización, etc.

La *revisión previa* de la prueba por parte de otros colegas expertos en la materia puede enriquecerla debido a sus juicios en cuanto a interpretación, facilidad o dificultad de los diferentes *ítems* o elementos.

- **Fiabilidad**

Una prueba es fiable cuando mide con un grado adecuado de *precisión*. Se refiere a la *consistencia* o constancia de los datos que puedan ser extraídos de la evaluación. Es decir, en el resultado de la prueba no interviene la subjetividad del

examinador y, por tanto, calificada por el mismo docente en fechas diferentes, o por distintos jueces, el resultado sería siempre el mismo o muy aproximado.

La prueba merece confianza cuando los resultados son consistentes, no están influidos por el evaluador no depende de cuándo ni por quién sea aplicada. Una prueba es más fiable cuando:

- Sostiene un *mayor número de preguntas*. A más preguntas menos probabilidad de incidencia de la suerte o azar. Cuidado con el tiempo concedido.
- Es superior la *calidad de las preguntas*. A mayor poder discriminatorio de las preguntas y menos ambigüedad en su formulación, mayor fiabilidad.
- El *método de corrección* es el adecuado. Buenas preguntas mal corregidas disminuyen la fiabilidad.
- Obviamente también influye en la fiabilidad factores tales como: cansancio de los alumnos, hora de realización, condiciones ambientales, tiempo disponible, etc.

- **Sensibilidad**

Una prueba es sensible cuando tiene capacidad para matizar el nivel de aprendizaje de cada cual. Por ello deberá encerrar preguntas relativamente fáciles, de mediana dificultad y aquellas que detecten rendimientos máximos.

Si se seleccionan sólo cuestiones fáciles, comprobaríamos que los rendimientos son altos; pero, por supuesto, no significativos. Lo contrario ocurriría si se centra toda la valoración en aquellos puntos que por ser de nuestra particular preferencia y campo de nuestra dedicación nos llevan a insistir en ellos, pero plantean tal dificultad que, normalmente, hacen fracasar a los estudiantes.

Una prueba bien planteada debería cubrir, por principio, toda el área de conocimientos motivo de examen. Los casos extremos, no recomendables, se darían, bien cuando en una prueba todos los alumnos resuelven todo *satisfactoriamente*, o en caso contrario, cuando ningún alumno resuelve dichas cuestiones.

- **Objetividad**

Los resultados finales de la evaluación deben ser independientes de quien pudiera tasarlos. Resulta evidente que esta característica, ligada a la de la fiabilidad, rige más y mejor cuando se utiliza para evaluar pruebas objetivas. En las pruebas de ensayo o de desarrollo, la característica de objetividad contiene mayores dificultades, pero ha de ser una permanente pretensión.

- **Representatividad**

Muy relacionada con las anteriores dado que una prueba puede considerarse como representativa cuando en ella se recogen, proporcionalmente a su importancia, todos los objetivos y contenidos motivo de examen. Porque, suele ser habitual en la construcción de pruebas el incluir excesivo número de preguntas que exigen sólo evocación de contenidos informativos, e incluso, excederse en determinadas cuestiones que sólo afectan a una determinada área, unidad de estudio o parte del total de los objetivos que se pretenden evaluar.

Para evitar estos problemas reales que disminuyen notablemente la calidad de la evaluación, resulta aconsejable elaborar previamente una *tabla de especificación* que nos garantice que se van a ponderar todos los objetivos y contenidos de forma adecuada. Me referiré en otra entrada a ese instrumento.

Referencia:

- García Aretio, L. (1987). *Rendimiento académico y abandono en la educación superior a distancia*. UNED.
- García Aretio, L. (1994). *Educación a distancia hoy*. UNED.
- García Aretio, L. (2001). *La educación a distancia. De la teoría a la práctica*. Ariel.