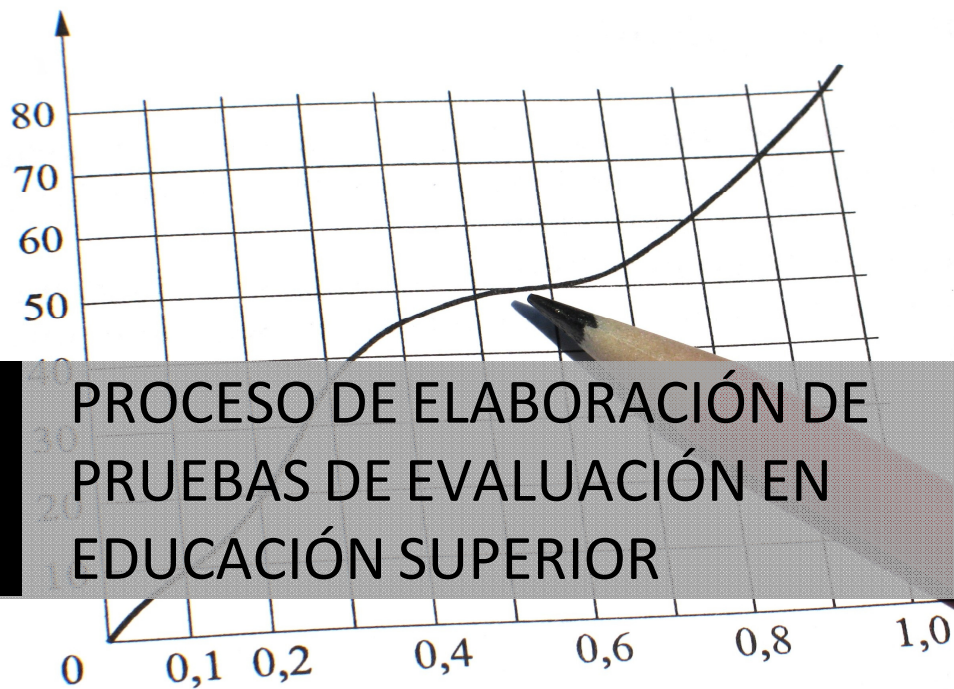


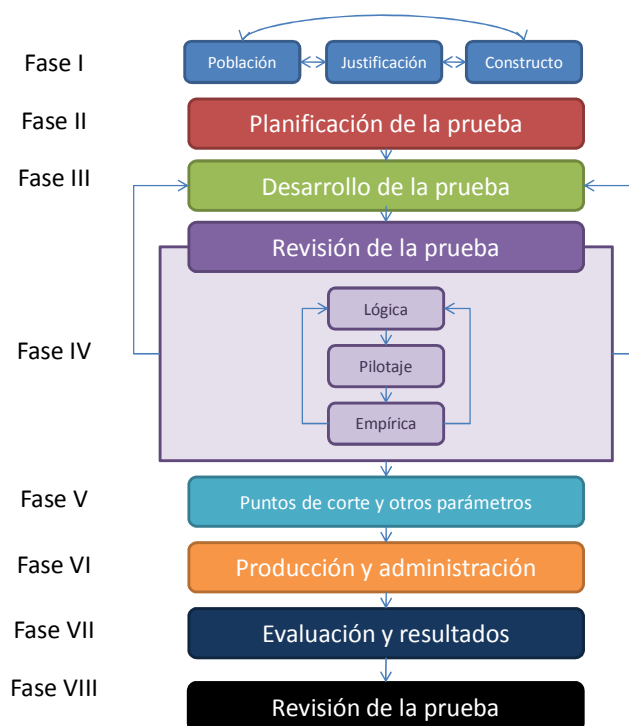
febrero de 2009



PROCESO DE ELABORACIÓN DE PRUEBAS DE EVALUACIÓN EN EDUCACIÓN SUPERIOR

El proceso de evaluación y desarrollo de pruebas

Figura 1: Diagrama de Flujo del proceso general de evaluación y desarrollo de pruebas



Fase I: Justificación de la prueba

En esta fase se deben abordar tres aspectos que redundarán en el buen desarrollo de fases posteriores. Estos tres aspectos están íntimamente relacionados.

- **Población:** se trata de identificar a la población que va a ser evaluada. Es importante especificar las características de esta población.
- **Justificación:** La justificación o propósito de la prueba responde a dos preguntas: a) cómo será usada e interpretada la prueba, b) qué se podrá inferir de los resultados y a qué contextos se podrá realizar esta inferencia.
- **Constructo:** se trata de especificar el dominio de la prueba, es decir qué se va a evaluar.

Documentos de trabajo en ANEXO:

IA: Ficha de especificación de las características de la población a evaluar.

IB: Ficha de justificación de la prueba.

IC: Ficha de especificación de dominio.

En Educación Superior y otros niveles educativos, y especialmente para las pruebas de clase, no es necesario detenerse en exceso en la población y la justificación de la prueba. Sin embargo, podrían ser dos elementos que a pesar de estar más o menos claro a priori, podrían ser motivo de reflexión.

Fase II: Planificación de la prueba

En esta fase se realizan todas las acciones concretas que llevan a la especificación de la prueba: qué apariencia van a tener las tareas o preguntas, detalles de administración, detalles sobre la puntuación, peso concreto de las preguntas en la prueba,...Para este caso sería conveniente diseñar el proceso de evaluación. Muchos detalles pueden surgir más adelante, siendo posible tener que volver a esta fase para la revisión de ciertos aspectos.

Muchas de estas cuestiones están resueltas con la definición de los criterios de evaluación, algunas no tanto.

Fase III: Desarrollo de la prueba

Es la fase donde se van a escribir los ítems concretos. Para ello hay una serie de reglas que se deben tener en cuenta.

Documentos de trabajo en ANEXO:

IIA: Tabla de especificaciones

IIIA: Especificaciones de ítem.

Fase IV: Revisión de la prueba

En esta fase se contemplan tres aspectos, uno de los cuales requiere de tanta complejidad como el proceso mismo de evaluación, únicamente que se realiza a escala menor, el pilotaje.

Se debería comenzar con una valoración del grado de adecuación del contenido de la prueba. Habitualmente se denomina a este aspecto como el de obtención de evidencias de la validez de contenido y se realiza con un grupo de jueces expertos. El contenido se debe valorar respecto a su relevancia y representatividad. En definitiva, se trata de una revisión lógica del test.

Una vez revisada la prueba y realizadas las correcciones pertinentes, se desarrolla el pilotaje de la misma. En el pilotaje entran en juego la mayor parte de aspectos que se tendrán en cuenta a la hora de realizar la prueba definitiva. La muestra del pilotaje no necesariamente ha de ser representativa, pero si lo es, asegura de manera más precisa cómo puede funcionar la prueba con una probabilidad más ajustada a la situación definitiva.

De este pilotaje se obtendrán datos concretos sobre el proceso mismo de la prueba y especialmente del funcionamiento de los ítems, aspecto fundamental desde un acercamiento desde la Teoría de Respuesta al Ítem (TRI). A partir de estos datos se pueden realizar distintos análisis que permiten realizar la revisión empírica.

Esta fase, y especialmente el pilotaje, requiere de una gran coordinación docente y tiempo de trabajo del que a veces no se dispone o por razones obvias, no puede realizarse un pilotaje en condiciones con los mismos estudiantes a los que luego van a realizar la prueba definitiva. Se recomienda al menos quedarse en el nivel de revisión lógica por un grupo de docentes que estén familiarizados con el dominio de contenido.

Documentos de trabajo en ANEXO:

IVA: Ficha para la revisión lógica de los ítems.

Fase V: Establecimiento de puntos de corte y otros parámetros

Esta fase está íntimamente relacionada con la revisión empírica y lógica de la anterior fase. Se trata de establecer estándares (desde una valoración de jueces) o puntos de corte (con datos empíricos), especialmente en pruebas de referencia criterial.

Esta fase requiere de unos conocimientos sobre evaluación educativa específicos y no siempre se dan las condiciones necesarias para aplicarlos de la forma correcta en el proceso de enseñanza/aprendizaje. Cuanto menos un único profesor o si los puntos de corte ha de depender de datos empíricos. No obstante, en este último caso se pueden aprovechar pruebas realizadas por alumnado de cursos pasados para establecer puntos de corte en una prueba similar que se realice en un momento determinado.

Fase VI: Producción y administración de la prueba

Se trata de llegar a un acuerdo sobre recomendaciones de formato de la prueba, aspectos relacionados con la producción de la prueba y con la administración de la misma (instrucciones concretas de administración, obtención de resultados, de corrección, por ejemplo).

Fase VII-VIII: Evaluación de resultados, informes, evaluación de la prueba

En esta fase se trata de una vez obtenidos los resultados analizarlos y emitir los informes correspondientes, atendiendo a normas éticas y respondiendo al propósito con el que se realizaba la evaluación.

Asimismo, se trata de hacer una revisión de aquellos aspectos de la prueba que no han funcionado bien.

Referencias bibliográficas

Shrock, S. Y Coscarelli, W. (2000). *Criterion-Referenced Test Development*. Silver Spring, MD: ISPI.

Chatterji, M. (2003). *Designing and Using Tools for Educational Assessment*. USA: Pearson Education Inc.

Westgaard, O. (1999). *Test that Work*. San Francisco, CA: Jossey-Bass/Pfeiffer.

Muñiz, J. (1997). *Introducción a la teoría de respuesta a los Ítems*. Madrid: Pirámide.

Olea, J., Ponsoda, V., Prieto, G. (1999). *Tests informatizados: fundamentos y aplicaciones*. Madrid: Pirámide.

Jornet, J. Y Suárez, J. (1995). Evaluación referida al criterio, en García Hoz, V. (Cord). *Tratado de Educación personalizada*. Vol. *Problemas y métodos de investigación en Educación Personalizada*. Barcelona: Rialp.

Jornet, J., Suárez, J., González Such, J. Y Belloch, C. (1995). *Estrategias de elaboración de pruebas criteriosales en Educación Superior*. UNED.

Anexos

IA: FICHA DE ESPECIFICACIÓN DE CARACTERÍSTICAS DE LA POBLACIÓN

En esta ficha se trata de delimitar las características de la población que va a pasar la prueba:

Características demográficas de la población (sexo, edad, tipo de escuela en la que están matriculados,...)

Población de referencia (N; ej.: son un total de 42.000 alumnos con estas características)

Población a la que afectará la prueba (censal, muestra (especificar n de muestra))

Criterios de exclusión para ciertos colectivos o variaciones a la hora de elaborar pruebas
(ej.: los alumnos con necesidades educativas especiales no realizarán la prueba,...)

IB: FICHA DE JUSTIFICACIÓN DE LA PRUEBA

En esta ficha se trata de especificar los propósitos de la evaluación que se va a realizar, respondiendo a los siguientes aspectos, además de los normativos (leyes, decretos, órdenes, resoluciones,...) que existan sobre el tema y que amparen justifiquen dicha evaluación.

¿ Por qué evaluamos?

1. Identificar las necesidades de los estudiantes antes de comenzar su instrucción
2. Establecer metas educativas
3. Planificar o mejorar estrategias de enseñanza
4. Emplazar a algunos estudiantes en grupos de enseñanza especial
5. Seguir el “crecimiento” del estudiante (en competencias)
6. Comunicar expectativas de los estudiantes, motivar estudiantes
7. Diagnosticar las fortalezas y debilidades del estudiante
8. Proporcionar entrenamiento y retroalimentación a los estudiantes
9. Proporcionar una base para graduar/marcar rendimiento
 - a. Acercamiento a criterio
 - b. Acercamiento normativo
 - c. Combinación de los dos
10. Comunicar el progreso de los estudiantes a sus familias
11. Tomar decisiones para la promoción o no de los estudiantes
12. Generar datos para evaluación del profesorado
13. Generar datos para el equipo de profesores/planificación de departamentos
14. Generar datos para evaluar programas (ej.: programas curriculares, etc
15. Otros:

¿Quién es más probable que utilice la información?

Profesores/Maestros

Estudiantes

Padres

Equipo de profesores

Equipos directivos de las escuelas

Otros (Explicar)

Otras cuestiones

IIA: MODELO TABLA DE ESPECIFICACIONES PARA TEST CON RESPUESTAS BASADAS EN ÍTEMS ESCRITOS ESTRUCTURADOS

Título de la unidad:

Contenido	Nivel Taxonómico			Total Pesos: n (%)
	<i>Conocimiento de hechos</i>	<i>Aplicación</i>	<i>Alto nivel</i>	
El estudiante...				
1.				
1.1.				
1.2.				
1.3.				
2.				
2.1.				
2.2.				
2.3.				
2.4.				
Total ítems				
Peso (%)				

Se puede cambiar el nivel taxonómico por aquellas categorías de la taxonomía elegida que se desee.

IIB: MODELO DE ESPECIFICACIÓN DE LA EVALUACIÓN

Propósito de la evaluación:

Población:

Constructo:

Dominio:

Indicador general (objetivo de aprendizaje)

1.

Indicadores específicos (competencias)

1.1.

1.2.

1.3.

1.4.

Indicador general (objetivo de aprendizaje)

2.

Indicadores específicos (competencias)

2.1.

2.2.

2.3.

2.4.

Método de valoración y pesos deseados

Indicadores:

Método y pesos:

Indicadores:

Método y pesos:

Método de puntaje

Indicadores:

Método y pesos:

Indicadores:

Método y pesos:

Valoración individual o grupal

Quién evalúa:

Tabla de especificaciones (Ficha IIA)

IIIC: MODELO DE FICHA DE ESPECIFICACIÓN DE ÍTEMS

Objetivo:

Especificación de habilidades:

Estructura del ítem:

Características del estímulo:

Características opcionales:

Ejemplo de ítem

Redacción (por escrito u oralmente):

Estímulo:

Respuesta clave

IVA: FICHA PARA LA VALORACIÓN DE LA CONGRUENCIA ÍTEM-OBJETIVO

(Trabajo para individual para cada juez)

Valore de 1 a 5 (1: muy baja, 2: baja, 3: media, 4: alta, 5: muy alta) la congruencia del ítem con su objetivo de referencia. La congruencia expresa si el ítem mide el objetivo para el cual fue creado.

Nº ítem	Valor	Nº ítem	Valor

(Trabajo del conjunto de jueces)

Matriz de valoración de los jueces.

Ítem															Total	Media	Desv. Típ
Juez																	
Total																	
Media ítem																	
Desv. Típ																	

IVB: FICHA PARA LA REVISIÓN LÓGICA DE LOS ÍTEMS

Criterios de revisión general:

<i>El ítem...</i>	<i>Ítems que NO cumplen con el criterio</i>
1. Asocia el contenido, conducta, condición y nivel taxonómico con la competencia a la que se refiere.	
2. Está redactado correctamente desde el punto de vista gramatical y ortográfico (incluyendo signos de puntuación)	
3. Tiene una respuesta correcta o la mejor respuesta claramente establecida.	
4. Evita el lenguaje confuso	
5. Está libre de pistas sobre su respuesta correcta.	
6. Está redactado en afirmativo. Si está en negativo se destacan las palabras necesarias.	

Criterios para ítems de asociación

	<i>Ítems que NO cumplen con el criterio</i>
1. Las instrucciones para realizar la asociación están claras.	
2. Los grupos del ítem son homogéneos.	
3. Hay más de 2 y menos de 15 ítems en un conjunto de asociación.	
4. Los ítems están en la izquierda y las respuestas en la derecha.	
5. Hay más opciones de respuesta que ítems.	
6. Los ítems y sus respuestas están ordenados siguiendo un sistema lógico	

Criterios para ítems de respuesta breve

<i>El ítem...</i>	<i>Ítems que NO cumplen con el criterio</i>
1. Los espacios en blanco están situados al final de la frase (a menos que se requiera otro lugar).	
2. Los espacios en blanco están únicamente establecidos para la palabra clave.	
3. Se evitan múltiples espacios en blanco.	
4. Los espacios en blanco son aproximadamente de la misma longitud.	
5. Las frases han sido parafraseadas de un texto (no son exactamente iguales al material de instrucción, libro de texto, etc).	

Criterios para ítems de verdadero/falso

<i>El ítem...</i>	<i>Ítems que NO cumplen con el criterio</i>
1. La redacción de la frase es claramente verdadera o claramente falsa.	
2. Se evitan dobles negaciones.	
3. Las frases verdaderas y falsas son de igual longitud (para el conjunto de ítems verdadero/falso).	
4. Hay aproximadamente el mismo número de frases verdaderas que falsas (para el conjunto de ítems verdadero/falso).	
5. Las frases verdaderas y falsas están distribuidas aleatoriamente en el test (para el conjunto de ítems verdadero/falso).	
6. Las frases han sido parafraseadas de un texto (no son exactamente iguales al material de instrucción, libro de texto, etc).	

Criterios para ítems de elección múltiple

<i>El ítem...</i>	<i>Ítems que NO cumplen con el criterio</i>
1. Hay sólo una cuestión o idea en el <i>tallo</i> .	
2. Las respuestas son paralelas en forma y contenido.	
3. Las respuestas son de la misma longitud.	
4. Las opciones “Todas las anteriores” o “Ninguna de las anteriores” están usadas con coherencia y cautela.	
5. La respuesta correcta varía de posición en los diferentes ítems (para el conjunto de ítems de elección múltiple en el test).	
6. Las respuestas de los ítems están listadas en algún orden lógico.	
7. Las opciones de respuesta a un ítem son de igual dificultad.	
8. Las opciones de respuesta están listadas verticalmente.	
9. Los distractores son creíbles. Reflejan errores comunes o de incomprensión en examinados tipo.	

Criterios para ítems interpretativos

<i>El ítem...</i>	<i>Ítems que NO cumplen con el criterio</i>
1. Es probable que el material contextual sea nuevo para el examinado (para el conjunto de ítems interpretativos del test)..	
2. La información es clara y fácil de leer.	
3. La pregunta requiere altos niveles de análisis e interpretación.	
4. Los ítems dependen de la información contextual (para el conjunto de ítems interpretativos del test).	
5. La lista de preguntas es de longitud apropiada (para el conjunto de ítems interpretativos del test).	
6. La respuesta clave es razonable.	

IVC: FICHA PARA LA REVISIÓN LÓGICA DEL DOMINIO

1.- Cobertura y esencia de la prueba

- el dominio cubre adecuadamente el conocimiento, habilidades, destrezas, procesos o conductas esperadas en un aprendiz competente
- el dominio ha incluido todos los elementos que debía incluir
- en el dominio se presentan redundancias
- las fuentes de documentación tienen la suficiente credibilidad y están bien documentadas

2.- Organización

- los indicadores se organizan desde los más complejos a lo más simples o cualquier otra forma lógica.
- los indicadores más amplios están concretados en otros más claros y medibles

3.- Aplicación de instrucciones

- las instrucciones o aclaraciones están redactadas de manera clara, concreta y explícita?
- se observan verbos de acción usados convenientemente
- las condiciones de observación se hacen explícitas convenientemente
- los criterios de ejecución o realización están especificados claramente

Observaciones

IVD: FICHA PARA LA VALORACIÓN DE LA CALIDAD DE LA ESPECIFICACIÓN DE LA EVALUACIÓN Y DE LA PRUEBA

Calidad de la especificación de la evaluación.

- el propósito de la evaluación está claro
- la población se ha identificado con detalle
- el dominio se ha clarificado con indicadores generales y específicos
 - los indicadores individuales especifican claramente el contenido, conducta, condiciones y criterios de ejecución
 - los indicadores se han clasificado según una taxonomía
- hay una tabla de especificaciones que indica el número de ítems a considerar en diferentes indicadores y niveles taxonómicos
- se ha completado la especificación de la evaluación (Ficha IIB)

Chequeo general para la prueba o herramienta de evaluación

- El conjunto de la prueba se asocia a la tabla de especificaciones
- La prueba coincide con otras especificaciones realizadas en la Ficha IIB
 - Condiciones de testaje
 - Nivel de desarrollo, edad, y otras especificaciones para la población a la que va destinada
 - Propósito de la evaluación
 - Adaptaciones o modificaciones para población con características especiales
- El vocabulario y el nivel de lectura son adecuados
- El título del test es claro
- Las instrucciones son claras
- Si hay secciones, las instrucciones para cada sección están claras
- Se facilita la lectura fácil y la interpretación de los ítems (diseño con espacio y orden)
- Si existieran, los gráficos, dibujos, tablas y pasajes son claros y apropiados.
- El tiempo total para la ejecución de la prueba y el procedimiento de puntuación está claramente definido (tanto para los examinados como para los examinadores)
- La tipografía o letra es legible
- Los criterios de puntuación son precisos y razonables
- El test está libre de sesgos o de lenguaje o dibujos discriminativos