

Guía para la elaboración de pruebas de evaluación educativa

Ministerio
de Educación y
Formación Profesional



Guía para la elaboración de pruebas de evaluación educativa



MINISTERIO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL

Secretaría de Estado de Educación

Dirección General de Evaluación y Cooperación Territorial

Instituto Nacional de Evaluación Educativa

Madrid 2020

Catálogo de publicaciones del Ministerio: <https://sede.educacion.gob.es/publiventa/inicio.action>
Catálogo general de publicaciones oficiales: <https://cpage.mpr.gob.es>

Guía para la elaboración de pruebas de evaluación educativa



MINISTERIO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL

Secretaría de Estado de Educación

Dirección General de Evaluación y Cooperación Territorial

Instituto Nacional de Evaluación Educativa <http://educacionyfp.gob.es/inee>

Edita:

© SECRETARÍA GENERAL TÉCNICA

Subdirección General de Atención al Ciudadano, Documentación y Publicaciones

Edición: 2020

NIPO línea: 847-20-096-7

NIPO IBD: 847-20-097-2

ISBN: 978-84-369-5959-8

Diseño de la portada e infografías: Iris Triviño Urbano. Para el diseño de la portada se han empleado vectores e iconos creados por Freepik.com y Flaticon.com.

Diseño y maquetación: Alea González Sánchez

Presentación

La evaluación educativa es un proceso esencial para realizar con garantías el diagnóstico y mejora de cualquier sistema educativo. Su diseño, aplicación y análisis supone la concurrencia de diversos campos del conocimiento, además de los directamente vinculados a los currículos escolares, cada uno con sus propios objetivos y peculiaridades pedagógicas.

La necesaria participación de estos conocimientos técnicos y conceptuales (estadística, sociología, epistemología, pedagogía...) en la evaluación educativa ha contribuido a que las investigaciones que los tienen como objeto se hayan acrecentado en los últimos años, propiciando una amplia bibliografía referida a la teoría y práctica de la evaluación educativa, entre la que ocupa un papel central la elaboración de las pruebas cognitivas, punto de intersección entre los equipos técnicos que las elaboran para las Administraciones u organismos y su genuino destinatario, que no es otro que el alumnado.

Desde el área de Evaluaciones nacionales del Instituto Nacional de Evaluación Educativa hemos elaborado la presente *Guía para la elaboración de pruebas de evaluación educativa* con el fin de aportar nuestra experiencia en el diseño de este tipo de pruebas. El equipo está dirigido por Carmen Tovar Sánchez e integrado por Ruth Martín Escanilla, Juan Mesonero Gómez, Óscar Urra Ríos e Iris Triviño Urbano, que además ha realizado las infografías. Alea González Sánchez, se encargó de la maquetación. Esther Andújar Duarte, también participó en la redacción durante su periodo de prácticas de máster en este instituto. Además, hemos contado con la colaboración del resto de componentes de otras áreas del Instituto.

Hemos concebido la publicación como una obra de decidida intención práctica y la hemos elaborado con el fin de ponerla a disposición de quien desee indagar en la evaluación educativa a través de la investigación, de la docencia o de su actividad profesional en las distintas administraciones. Si con ello podemos aportar algo útil a la comunidad educativa, daremos por bien empleado este trabajo.

Índice

Introducción a la evaluación educativa. El marco de evaluación.....	7
La evaluación educativa.....	7
El marco de evaluación.....	8
¿Qué evaluar?.....	8
¿A qué población se dirige la evaluación?.....	8
¿En qué momento evaluar? ¿Cuál será el horario de aplicación?.....	9
Instrumentos de evaluación.....	9
Análisis y presentación de los resultados.....	9
Análisis de las competencias.....	9
➤ Dimensiones de la competencia.....	10
➤ Tabla de especificaciones.....	12
Pruebas de evaluación: cuestionarios cognitivos.....	15
Características generales de las pruebas de evaluación.....	15
Definición y objetivo.....	15
Elementos principales.....	16
➤ Estímulos.....	16
➤ Preguntas e ítems.....	17
➤ Unidades de evaluación	19
➤ Guía de codificación.....	19
Tipos de preguntas y utilidad.....	21
➤ Preguntas de respuesta cerrada.....	21
• Características.....	21
• Subtipos.....	22
• Corrección y codificación de las preguntas de respuesta cerrada.....	28
➤ Preguntas de respuesta semiconstruida.....	30
• Características.....	30
• Corrección y codificación de las preguntas de respuesta semiconstruida..	33
➤ Preguntas de respuesta abierta o construida.....	34
• Características.....	34
• Corrección y codificación de las preguntas de respuesta construida.....	34
• Corrección y codificación de las preguntas de respuesta abierta corta.....	36
• Corrección y codificación de las preguntas de respuesta abierta larga....	40

Elaboración de pruebas de evaluación.....	49
Ciclo de elaboración de pruebas.....	49
➤ Construcción de las unidades de evaluación.....	49
➤ Revisión de calidad de los estímulos e ítems asociados.....	49
➤ Configuración de las pruebas de evaluación.....	50
➤ Pruebas piloto.....	50
➤ Estudio principal.....	51
Recomendaciones generales para la elaboración de pruebas.....	53
➤ Formato de la prueba.....	53
➤ Características generales.....	53
➤ Elaboración de ítems.....	54
• El tronco o vástago.....	55
• Las opciones (clave y distractores). Ítems de opción múltiple.....	56
➤ Ítems de respuesta cerrada de opción múltiple compleja.....	59
➤ Ítems de respuesta cerrada de emparejamiento o relación.....	59
➤ Ítems de respuesta cerrada de rellenar huecos.....	61
➤ Ítems de respuesta semiconstruida.....	61
➤ Ítems de respuesta abierta o construida	63
Cuestionarios de contexto.....	68
Elaboración de los cuestionarios de contexto.....	68
Diseño de los cuestionarios de contexto.....	68
Composición de los cuestionarios de contexto.....	69
➤ El Índice social, económico y cultural en los cuestionarios de contexto.....	72
Características de los cuestionarios de contexto.....	72
Bibliografía.....	73

Introducción a la evaluación educativa. El marco de evaluación

La evaluación educativa

La evaluación ocupa en la actualidad una posición protagonista dentro del ámbito educativo gracias, principalmente, a la toma de conciencia sobre su importancia por parte de los diferentes agentes educativos y al alcance que los informes de resultados han tenido durante las últimas décadas. Cualquier proceso educativo necesita de evaluación ya que, tal y como aparece recogido en la legislación vigente en nuestro país, esta constituye un elemento fundamental para la mejora educativa y para la rendición de cuentas de las Administraciones correspondientes con respecto a la situación y desarrollo del sistema educativo. La evaluación es una herramienta de seguimiento y valoración tanto de los resultados como del propio proceso de enseñanza-aprendizaje que, además, no se circunscribe de manera exclusiva al aula y a quienes en ella operan, esto es, alumnado y profesorado, sino que comprende asimismo otros ámbitos y agentes de la actividad educativa entre los que se encuentran equipos directivos, centros, currículo, inspección y Administraciones.

Pero, ¿qué implica exactamente la “evaluación educativa”? Si se quiere comprender su importancia y las razones por las que su ámbito de actuación se ha visto ampliado, es necesario clarificar en primer lugar este concepto. Se entiende la evaluación como una actividad o proceso sistemático de identificación, recogida o tratamiento de datos sobre elementos o hechos educativos, con el objetivo de valorarlos primero y, sobre dicha valoración, tomar decisiones (García Ramos, 1989). El resultado de una evaluación nunca debe ser, simplemente, una calificación numérica, sino que de ella debe emanar un juicio de valor que permita a los agentes educativos correspondientes orientar sus acciones hacia la mejora de la calidad del proceso de enseñanza-aprendizaje. La evaluación, al prescribir realmente los objetivos de la educación, determina en gran medida lo que los alumnos aprenden y cómo lo aprenden, lo que los profesores enseñan y cómo lo enseñan, los contenidos y los métodos; en otras palabras, el producto y el proceso de la educación, ya que la actividad educativa del alumnado y el profesorado está en algún grado canalizada por la evaluación (Orden Hoz, 1989).

Todos los agentes que intervienen en el proceso educativo son susceptibles de ser evaluados: alumnado, docentes, equipos directivos, centros educativos, Administraciones educativas, etc. Además, se pueden distinguir diferentes tipos de evaluación educativa según el fin que esta persiga. Sin embargo, el propósito de esta introducción no es hacer un repaso de las numerosas formas y tipos de evaluación y de los diversos instrumentos empleados para realizarlas, sino centrarse en las evaluaciones externas o heteroevaluaciones a gran escala (es decir, aquellas que implican a un gran conjunto de alumnado y centros educativos y no a la que realiza el profesorado en su quehacer diario dentro del aula) y que se llevan a cabo tomando como fuente fundamental de información al alumnado, pero también la información proporcionada por las familias, los docentes y los equipos directivos. Estas evaluaciones se orientan a la evaluación de los sistemas educativos, bien de ámbito internacional, nacional o regional.

En España, se han venido realizando evaluaciones de diagnóstico a nivel autonómico y estatal desde hace ya más de dos décadas. En la actualidad, estas se orientan a la evaluación de las competencias alcanzadas por el alumnado en diferentes momentos de la escolarización obligatoria. Estas evaluaciones, recogidas en la legislación vigente, tienen un carácter formativo y orientador, como corresponde a su función diagnóstica y, como se ha señalado anteriormente, su finalidad última es la recogida de información valiosa que oriente la toma de decisiones. Asimismo, existen organismos internacionales como la OCDE (Organización para la Cooperación y el Desarrollo Económicos) y la IEA (Asociación Internacional para la Evaluación del Rendimiento Educativo) que promueven estudios de evaluación internacional y a gran escala para evaluar el nivel de aprendizaje que se alcanza en un país o región.

Respecto a los datos extraídos por las evaluaciones a gran escala se utilizan, principalmente, para guiar y encauzar las políticas educativas, por tanto, sus resultados también pueden tener efectos directos dentro del aula, siempre que se asuman las limitaciones que, como potencial guía curricular, tienen las pruebas estandarizadas de las que se componen este tipo de estudios a gran escala (Shepard, 2006).

Las evaluaciones habitualmente toman como unidad de información el rendimiento del alumnado y como instrumentos de evaluación emplean pruebas cognitivas complementadas con cuestionarios de contexto.

El marco de evaluación

El punto de partida de toda evaluación educativa es el diseño de un marco de evaluación que responda a todas las preguntas relacionadas con su planificación y organización. Entre estas cuestiones el marco de evaluación da respuesta a los siguientes apartados:

¿Qué evaluar?

En la actualidad la evaluación no se orienta a la valoración de los contenidos concretos que el alumnado conoce, sino que, con un enfoque competencial, se centra en la capacidad de emplear estos conocimientos y habilidades para la resolución de tareas específicas; por ello, se deben definir claramente las competencias, destrezas, capacidades y habilidades que se pretende diagnosticar, sus dimensiones y los indicadores de logro que servirán para diseñar los instrumentos de evaluación (Villa y Poblete, 2007). La descripción de estas competencias debe remitirse al marco legal existente.

Además de la definición y el análisis de las competencias que se quieren evaluar, debe determinarse claramente si la evaluación se orienta al diagnóstico del sistema educativo en su conjunto (y, por lo tanto, a la concreción de la distribución de la población en los distintos niveles de desarrollo competencial que se definan), o si se buscan resultados de rendimiento por centros educativos o incluso a nivel individualizado. Debe resaltarse que toda evaluación lleva implícita la idea de mejora, entendiendo por mejora la reducción de las diferencias entre los resultados obtenidos y los resultados deseados, por lo que esta idea debe estar presente desde el inicio de su planificación.

¿A qué población se dirige la evaluación?

La respuesta a esta pregunta está muy relacionada con el objetivo perseguido por la evaluación. En este sentido se plantean principalmente dos tipos de modelos. En primer lugar, el más extendido es el de la evaluación que se dirige a un curso o nivel educativo concreto. Así, será susceptible de participar en la evaluación todo el alumnado que en el momento de su realización esté cursando un determinado nivel educativo, independientemente de su edad. Por ejemplo, las pruebas internacionales TIMSS (Estudio de las Tendencias en Matemáticas y Ciencias) y PIRLS (Estudio Internacional de Alfabetización Lectora) se dirigen a todo el alumnado que está cursando 4.º de Educación Primaria en el momento de realización del trabajo de campo. Un segundo modelo incluye en la población objetivo al alumnado que tiene una determinada edad en el momento de la evaluación; es el caso de PISA (Programa internacional para la Evaluación de Estudiantes), que evalúa al alumnado que cumple 16 años en el año natural en el que se realiza la prueba.

Por otra parte, la evaluación se puede realizar de manera "censal" o de manera "muestral". En el primer caso, todo el alumnado que cumpla los requisitos para estar incluido en la población objetivo deberá participar en la evaluación, mientras que en el segundo caso solo participará una muestra tomada mediante métodos estadísticos de forma que sea representativa de la población. Dependiendo del objetivo de la evaluación se deberá realizar de un modo u otro, aunque, si el objetivo es el diagnóstico del sistema educativo en su conjunto, el análisis de los datos derivados

de una muestra representativa permite estimar los verdaderos valores poblacionales con un margen de error acotado y tan pequeño como se desee, una vez fijado el nivel de confianza, y con costes asociados mucho menores que si se realizara una evaluación censal. La única justificación que podría tener una evaluación de carácter censal sería la evaluación individualizada del alumnado en una escala externa estandarizada. Esta evaluación individualizada del alumnado suele tener un carácter informativo sobre el nivel de adquisición de las competencias evaluadas y es siempre complementaria a la evaluación continua e interna que realiza el profesorado.

¿En qué momento evaluar? ¿Cuál será el horario de aplicación?

Las respuestas a estas preguntas se refieren tanto al nivel educativo como al momento dentro del curso escolar en el que se realice la evaluación. En cuanto al horario de aplicación, se deberá tener en cuenta la distribución horaria en los centros educativos para interferir lo mínimo posible durante los días de aplicación.

Instrumentos de evaluación

La realización de pruebas de evaluación a gran escala supone una alteración de la vida cotidiana de los centros educativos. Por ello, el trabajo de campo relacionado con este tipo de evaluaciones requiere unos instrumentos que, además de ser adecuados para la medición de la competencia a la que se orientan, se puedan aplicar en los espacios disponibles en los centros y en un tiempo limitado. Por ello, las evaluaciones se realizan normalmente mediante pruebas escritas, en formato papel o, cada vez con mayor frecuencia, en formato digital. La aplicación suele realizarse en uno o más días, dependiendo de las competencias que se evalúen. Dentro de los instrumentos de evaluación se incluyen todos los materiales que se utilizarán como fuente de información, entre los que se encuentran las pruebas cognitivas y los cuestionarios de contexto.

Análisis y presentación de los resultados

En este apartado se incluye la planificación del modo en el que se obtendrán las puntuaciones y/o niveles de rendimiento, el nivel de desagregación que se realizará de los mismos y la forma en la que se presentarán y comunicarán a la comunidad educativa. En ese sentido, se deben definir los tipos de informes que se elaborarán.

Análisis de las competencias

El marco de evaluación debe definir con claridad los referentes de evaluación de cada una de las competencias, lo que permitirá elaborar los instrumentos de evaluación de forma adecuada. Entre otras cuestiones se definirá de forma precisa la competencia, las dimensiones de la misma y los criterios de evaluación o indicadores de logro del aprendizaje¹. Asimismo, se suele incluir la tabla o matriz de especificaciones, de la que aquí se hablará en el siguiente apartado. Además, es habitual incorporar ejemplos de preguntas de evaluación de la competencia de distintos niveles² y su situación dentro del marco de la evaluación.

1 Existen diferentes terminologías. La legislación actual incluye criterios de aprendizaje y estándares de aprendizaje. Otra posible denominación es “meta de aprendizaje”.

2 El nivel de las preguntas se estima mediante la prueba piloto. Si en el momento de la definición del marco de la evaluación no se ha realizado una prueba piloto, se incluye el nivel estimado por los expertos que diseñan el marco de evaluación.

➤ Dimensiones de la competencia

El análisis de las competencias incluye las distintas dimensiones en las que esta se desglosa para su evaluación. Entre ellas se incluyen los bloques de contenidos, dominios de contenido (TIMSS) o temas (denominación en PISA); procesos cognitivos (existen distintas taxonomías para su clasificación, la más conocida es la de Bloom (Bloom, 1956); los contextos; las destrezas y algunas otras dimensiones específicas de cada competencia, como son, por ejemplo, los tipos de texto, en competencia lingüística. La definición de las dimensiones y subdimensiones se incluye en el marco de la evaluación acompañada de una propuesta de su peso relativo asociado, teniendo en cuenta el curso o nivel al que se dirige la prueba de evaluación y el desarrollo cognitivo del alumnado.

Esta propuesta sirve para orientar la construcción de la prueba de evaluación en tanto que determina, de modo orientativo, el número de preguntas e ítems que deben asociarse a cada cruce de estas subdimensiones, que se muestra en la tabla de especificaciones, evitando que se produzca un exceso o defecto de ítems asociados a alguna subdimensión. Así se logra una prueba equilibrada y adaptada a la planificación prevista en el marco de la evaluación.

Para evaluar las matemáticas, el marco de evaluación de TIMSS 2019 (Mullis, I.V.S. y Martin, M.O., 2017) establece las dimensiones de contenidos y las dimensiones cognitivas que se recogen en la Figura 1.1. A su vez, las dimensiones de contenido y las dimensiones cognitivas se desglosan en diferentes subdimensiones, que aparecen descritas en el correspondiente marco de evaluación. Además, se incluyen los porcentajes aproximados que deben tenerse en cuenta como ponderación o peso en la elaboración de la prueba.

Figura 1.1. Dimensiones de contenido y dimensiones cognitivas en TIMSS 2019. Matemáticas

TIMSS 2019. 4.º de EP. Matemáticas			
Dimensiones de contenido		Dimensiones cognitivas	
Números	50 %	Conocimiento	40 %
Mediciones y Geometría	30 %	Aplicación	40 %
Estadística	20 %	Razonamiento	20 %

La Figura 1.2 muestra los procesos cognitivos y sus diferentes niveles de progresión descritos en el *Marco de la evaluación final de Educación Primaria* (INEE, 2015b). Estos niveles aparecen definidos, y a cada uno de ellos se le asocian una serie de verbos que facilitan su posterior relación con los indicadores de logro que sirven de referencia a la evaluación.

Figura 1.2. Ejemplo de definición de los procesos cognitivos para la competencia matemática. *Evaluación final de Educación Primaria*

Procesos		Descripción	Acciones
Conocer y reproducir	Acceso e identificación	Acciones de recordar y reconocer los términos, los hechos, los conceptos elementales del conocimiento matemático y de reproducir algoritmos.	Nombrar, definir, encontrar, mostrar, imitar, listar, contar, recordar, reconocer, localizar, reproducir, relatar.
	Comprensión	Acciones para captar el sentido y la intencionalidad de textos de lenguaje matemático y de códigos relacionales e interpretarlos para resolver problemas.	Explicar, ilustrar, extraer, resumir, completar, traducir a otros términos, aplicar rutinas, seleccionar, escoger.
Aplicar y analizar	Aplicación	Aptitud para seleccionar, transferir y aplicar información para resolver problemas con cierto grado de abstracción y la de intervenir con acierto en situaciones nuevas.	Clasificar, resolver problemas sencillos, construir, aplicar, escoger, realizar, desarrollar, entrevistar, organizar, enlazar.
	Análisis	Posibilidad de examinar y fragmentar la información para resolver problemas, encontrar causas y motivos, realizar inferencias y encontrar evidencias que apoyen generalizaciones.	Comparar, contrastar, demostrar, experimentar, planificar, resolver, analizar, simplificar, relacionar, inferir, concluir.
Razonar y reflexionar	Síntesis y creación	Acciones de recoger información y relacionarla de distintas formas, establecer nuevos patrones y descubrir soluciones alternativas.	Combinar, diseñar, imaginar, inventar, planificar, predecir, proponer, adaptar, estimar.
	Juicio y valoración	Capacidades para formular juicios con criterio propio, cuestionar tópicos y exponer y sustentar opiniones fundamentadas.	Criticar, concluir, determinar, juzgar, recomendar, reformular, establecer criterios y/o límites.

En el análisis de las competencias se debe tener en cuenta el contexto en el que se redactarán las preguntas. Por ejemplo, la elección apropiada de estrategias matemáticas y representaciones depende normalmente del contexto en el que se plantea un problema. Según el *Marco Analítico de Evaluación de PISA 2018* (OCDE, 2017), es importante que se use una amplia variedad de contextos en la evaluación para alcanzar la gama más amplia posible de intereses individuales y casuísticas de las personas que conforman nuestra sociedad del siglo XXI. Concretamente para competencia matemática, PISA 2018 estableció cuatro categorías de contexto:

- Personal: engloba problemas o desafíos a los que podría enfrentarse el alumnado relacionados con él mismo, su familia o su grupo de amigos. Este tipo de contextos que se consideran personales incluyen la preparación de comida, hacer la compra, juegos, salud personal, deportes, viajes, horario y finanzas personales...
- Ocupacional: incluye problemas centrados en el mundo del trabajo, desde medidas y escalas, contabilidad, control de calidad, diseño y arquitectura, hasta cualquier ámbito del trabajo, siempre sin olvidar que los ítems que se planteen al alumnado deben ser accesibles para su edad.
- Social: plantea problemas que afectan a la sociedad y a la comunidad del alumnado. Algunos ejemplos son el uso de transporte público, estadísticas nacionales, economía, políticas públicas, gobierno, sistemas de votación...

- Científico: referido a problemas que implican el uso de las matemáticas relacionados con la Ciencia y la Tecnología. Por ejemplo: climatología, tiempo meteorológico, medicina, ciencias espaciales, genética, medidas y el mundo de las matemáticas en sí mismo.

En el mismo marco mencionado anteriormente, se establece que el porcentaje de ítems en una prueba de cada contexto debe ser de un 25 %.

➤ Tabla de especificaciones

A partir de las dimensiones definidas, se establece una relación entre estas que puede mostrarse en forma de matriz o tabla; son las denominadas “matrices” o “tablas de especificaciones”. En ellas se ponen en relación las distintas dimensiones de la competencia analizadas con los indicadores de logro o criterios de evaluación. La tabla de especificaciones suele incluir los porcentajes asignados a cada subdimensión de la competencia con el fin de servir de guía útil para la elaboración de las pruebas.

Por ejemplo, la tabla de especificaciones que se muestra en la Figura 1.3 contiene una propuesta del peso porcentual de los procesos cognitivos y los bloques de contenido asociados. Debe indicarse que existen tablas de especificaciones multidimensionales (matrices de especificaciones): la que se muestra en la figura relaciona únicamente dos dimensiones.

Figura 1.3. Modelo de tabla de especificaciones de una competencia

			Temas				%	Total
			Tema 1	Tema 2	Tema 3	Tema 4		
Procesos	Proceso 1	Subproceso 1					XX %	XX %
		Subproceso 2					XX %	
	Proceso 2	Subproceso 1					XX %	XX %
		Subproceso 2					XX %	
	Proceso 3	Subproceso 1					XX %	XX %
		Subproceso 2					XX %	
Ponderación total			XX %	XX %	XX %	XX %	100 %	

En el cruce de dos de las subdimensiones se incluyen los criterios de evaluación o los indicadores de logro asociados a ambas. Los indicadores de logro deben de ser evaluables, observables y medibles y permiten graduar el rendimiento o desempeño alcanzado en la competencia. Por ejemplo, la Figura 1.4 muestra una de las celdas de la Tabla de especificaciones de la competencia matemática en el *Marco de la evaluación final de Educación Primaria* (INEE, 2015b).

Figura 1.4. Ejemplo de celda de la tabla de especificaciones de la competencia matemática en el *Marco de la evaluación final de Educación Primaria* (2015)

		Tema: Números
Proceso: conocer y reproducir	Acceso e identificación	Identifica múltiplos y divisores, utilizando las tablas de multiplicar.
		Calcula cuadrados, cubos y potencias de base 10.
		Redondea números decimales a la décima, centésima o milésima más cercana. [...]
	Comprensión	Calcula el mínimo común múltiplo y el máximo común divisor.
		Utiliza los números negativos en contextos reales.
		Utiliza los números ordinales en contextos reales.
		Realiza operaciones con números decimales: suma, resta, multiplicación y división. [...]



Elaboración de pruebas de evaluación

Marco de evaluación

- ¿Qué evaluar?
- ¿A qué población?
Censal o muestral.
- ¿Cuándo? ¿En qué horario?
- ¿Qué instrumentos de evaluación?
- Definición de las competencias.
- Análisis de datos e informe de resultados.

Cuestionarios cognitivos

- ¿Qué tipo de preguntas?
- Respuesta cerrada.
- Respuesta semiconstruida.
- Respuesta construida o abierta.

Cuestionarios de contexto

- ¿A quién irán dirigidos?
- Alumnado.
- Familias.
- Profesorado.
- Tutores y tutoras.
- Equipos directivos.

¿Cómo evaluar competencias?

- Exige el diseño de instrumentos en los que el estudiante demuestre con evidencias que puede realizar las tareas de la competencia exigida.
- Requiere desenvolverse en diferentes contextos que exigen al alumnado activar e integrar conocimientos, normas, técnicas y procedimientos, habilidades, destrezas, actitudes y valores.

Pruebas de evaluación por competencias

Formadas por **preguntas e ítems** que requieren la activación de la competencia con la implicación de **procesos cognitivos de grado superior** en los que el alumnado debe comprender, analizar, extraer información, sintetizar, diseñar, justificar o extraer conclusiones.

COMPETENCIA*

Combinación de **conocimientos, capacidades y actitudes** adecuadas al contexto.

Conocimientos

Hechos y cifras, conceptos, ideas y teorías que ya están establecidos y apoyan la comprensión de un área o tema concreto.

Capacidades

Habilidad para realizar procesos y utilizar los conocimientos existentes para obtener resultados.

Actitudes

Mentalidad y disposición para actuar o reaccionar ante ideas, personas o situaciones.

DEFINICIÓN DE LA COMPETENCIA

Temas, bloques o dimensiones de contenidos

Procesos cognitivos

Destrezas

Contextos

Resultados del aprendizaje

(criterios de evaluación, indicadores de aprendizaje, metas educativas...)

Tabla de especificaciones

Incluye los pesos otorgados a cada una de las subdimensiones de la competencia.

		DIMENSIONES				Total
		Dimensión 1		Dimensión 2		
		Subdim. 1	Subdim. 2	Subdim. 3	Subdim. 4	
PROCESOS	Proceso 1					35 %
	Proceso 2					25 %
	Proceso 3					40 %
	Ponderación total	20 %	35 %	15 %	30 %	100 %

Pruebas de evaluación: cuestionarios cognitivos

Las pruebas de evaluación son los instrumentos empleados habitualmente para realizar las evaluaciones de carácter externo. Estas pruebas consisten en cuestionarios cognitivos, que son instrumentos de valoración constituidos por preguntas o ítems que, para su respuesta correcta, requieren la activación de la competencia con la implicación de procesos cognitivos de grado superior en los que el alumnado debe comprender, analizar, extraer información, sintetizar, diseñar, justificar o extraer conclusiones.

En la evaluación competencial se debe plantear un modelo de evaluación no basado en la memorización de aprendizajes conceptuales, sino en el desarrollo y fomento de las competencias clave que el alumnado debe adquirir. La consideración de los cuestionarios cognitivos como única fuente de información del nivel de desarrollo competencial en todas sus dimensiones ha sido cuestionada, argumentando que con este instrumento se dejan fuera de la evaluación ciertas dimensiones de la competencia como son el “saber hacer” o las actitudes y valores. Sin embargo, si el diseño de los cuestionarios cognitivos es suficientemente cuidado es posible acercarse a la evaluación de estos aspectos de las competencias de manera que resulte útil para evaluar de manera fiable el nivel de desarrollo competencial. No obstante, no debe olvidarse que la evaluación externa proporciona una “foto fija” del desarrollo competencial y por lo tanto está sujeta a varias fuentes de error, una de ellas derivada del instrumento de evaluación utilizado. La evaluación externa es siempre complementaria a la evaluación continua en el aula, que utiliza además de los cuestionarios cognitivos otras muchas fuentes de información, y de hecho cada vez con mayor frecuencia se emplean en las evaluaciones externas, de forma incipiente o experimental, instrumentos de evaluación diferentes de los cuestionarios cognitivos. No obstante, esta sección se centrará únicamente en la descripción de este tipo de cuestionarios.

Características generales de las pruebas de evaluación

Definición y objetivo

Una prueba de evaluación es un cuestionario cognitivo que se realiza sobre el rendimiento de personas evaluadas, a quienes propone situaciones-problema ante las que deberán responder demostrando en qué grado son capaces de activar sus conocimientos, habilidades, actitudes o valores.

Las pruebas de evaluación deben constituirse a partir de casos o situaciones que sirvan como base para la indagación y que se remitan, en la medida de lo posible, a situaciones familiares o verosímiles para las personas evaluadas. Las preguntas formuladas a partir de cada situación-problema deben ir dirigidas a la comprobación del grado de dominio de las competencias objeto de la evaluación.

Si nos referimos a las pruebas de evaluación estandarizadas, estas se constituyen como el principal instrumento de medida para el análisis del rendimiento de un conjunto de estudiantes de un determinado nivel educativo. Estas pruebas requieren que su construcción sea idónea para el objetivo que persiguen, y su elaboración, desde su diseño hasta su validación, es un procedimiento largo y de cierta complejidad.

Elementos principales

El elemento fundamental de las pruebas de evaluación es el **ítem**, que de modo amplio podría definirse como cada una de las preguntas que conforman estas pruebas. Los ítems suelen agruparse formando **unidades de evaluación**, cada una de las cuales está formada por una serie de ítems vinculados a una presentación textual, gráfica, sonora o audiovisual denominada **estímulo**, que presenta a la persona evaluada una situación determinada. Normalmente, para cada competencia que desee evaluarse se elaboran varias unidades de evaluación.

Al igual que los ítems se agrupan para formar unidades de evaluación, también estas lo hacen para dar lugar a pruebas de evaluación. Las pruebas de evaluación se apoyan en su correspondiente **tabla de especificaciones** que responde a lo establecido en el marco de evaluación y sirve de guía al proceso de creación de la prueba, ayudando a cubrir todas las dimensiones de la competencia que se esté evaluando.

➤ Estímulos

En términos generales, los estímulos más habituales suelen ser de cuatro tipos:

- a) **Textos escritos:** anuncios, textos extraídos de los medios de comunicación, instrucciones, carteles informativos, diálogos o narraciones de hechos, descripciones de situaciones reales o simuladas, textos literarios o científicos, etc.
- b) **Imágenes:** fotografías, dibujos, mapas, esquemas, gráficos, cómics, etiquetas de productos, etc.
- c) **Audios:** narraciones, entrevistas, reportajes, etc.
- d) **Audiovisuales:** en general, vídeos.

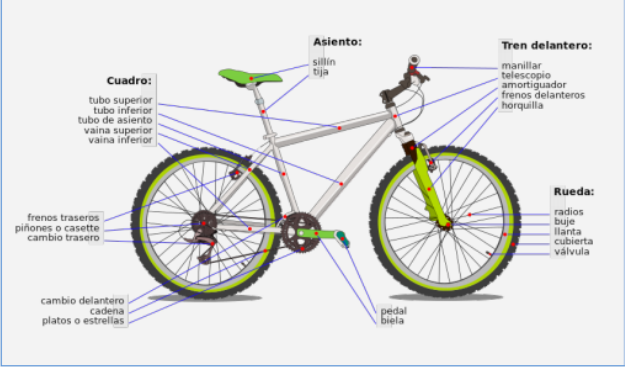
Es importante que los tipos y formatos de estímulos seleccionados para una prueba sean diversos. Además, es recomendable que cada estímulo lleve asociado un título. La Figura 2.1 muestra dos estímulos que se han empleado para la evaluación de las competencias básicas en ciencia y tecnología, en 6.º de Educación Primaria (INEE, 2015b), y la competencia lingüística en 3.º de Educación Primaria (INEE, 2015a), respectivamente, aunque podrían ser situaciones válidas para evaluar otras competencias o ser empleadas para otros niveles educativos.

Figura 2.1. Ejemplo de estímulos

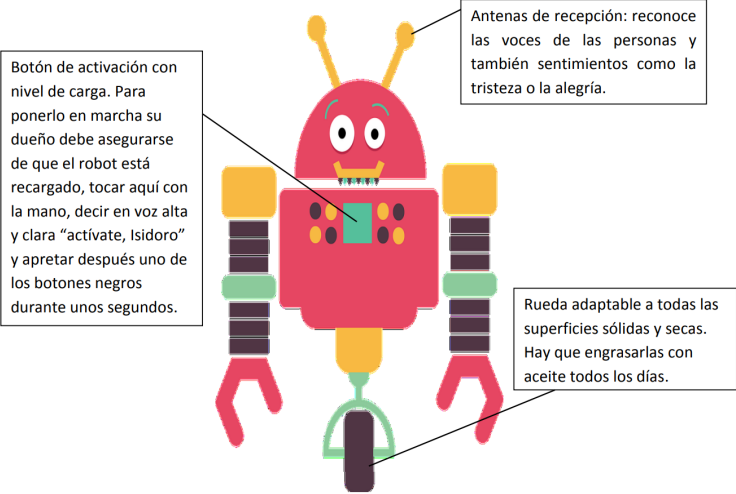
RUTA EN BICICLETA

Celia y Pedro van a hacer una ruta en bicicleta este verano con sus padres. Para tener a punto sus bicis y evitar cualquier problema, las han llevado al taller de su tío. También quieren instalar unas luces en las bicis para los paseos nocturnos.

Al entrar en el taller, han visto este póster.



Este es Isidoro, un robot que está todavía en fase de pruebas. Se alimenta de electricidad y hay que recargarlo cada dos horas. Sabe contar cuentos y cantar canciones, aunque no es capaz de mantener conversaciones. Pero lo que Isidoro puede hacer que no hace ningún otro robot es inventar y construir objetos nuevos. Infórmate bien sobre este robot y contesta a las preguntas.



Cuadro:
tubo superior
tubo inferior
tubo de asiento
vainas superior
vainas inferior

Asiento:
sillín
tija

Tren delantero:
manillar
telescopio
amortiguador
frenos delanteros
horquilla

Rueda:
radios
bujes
llanta
cubierta
válvula

frenos traseros
piñones o cassette
cambio trasero

cambio delantero
cadena
platos o estrellitas

pedal
biela

Botón de activación con nivel de carga. Para ponerlo en marcha su dueño debe asegurarse de que el robot está recargado, tocar aquí con la mano, decir en voz alta y clara "actíivate, Isidoro" y apretar después uno de los botones negros durante unos segundos.

Antenas de recepción: reconoce las voces de las personas y también sentimientos como la tristeza o la alegría.

Rueda adaptable a todas las superficies sólidas y secas. Hay que engrasarlas con aceite todos los días.

➤ Preguntas e ítems

Aunque en la mayoría de los casos ambos términos son intercambiables, a cada uno le corresponde una definición diferente y, a efectos de esta guía, es fundamental que su distinción quede clara.

Entendemos por pregunta cada uno de los enunciados a los que la persona evaluada debe responder, mientras que una posible definición de ítem sería cada una de las unidades mínimas de medida de la destreza o competencia que se persigue evaluar. Así, el ítem sería instrumento de medida de un determinado aspecto de la habilidad de los individuos referida a un aprendizaje

escolar (Haladyna, 2004). La habilidad que se pretende medir se refleja en un indicador de logro concreto que es evaluado a partir de la respuesta dada a la pregunta planteada. Según estas definiciones, en la mayoría de las cuestiones se identifica pregunta con ítem, ya que generalmente cada una de las preguntas persigue la medición de un único indicador de logro de los descritos en el marco de evaluación. Sin embargo, en algunos tipos de tareas (la producción de textos escritos o la realización de una exposición oral, el diseño de un experimento, la interpretación de un fragmento musical o la realización de una obra artística, etc.) se medirán o valorarán varios indicadores de logro a partir de la misma producción. En ese caso, una pregunta se asociará a varios ítems, dependiendo del tipo que sea. En el apartado de esta guía dedicado a los tipos de preguntas se darán más detalles sobre ello. La Figura 2.2 y la Figura 2.3 muestran ejemplos de preguntas asociadas a un ítem y a más de un ítem, respectivamente.

Figura 2.2. Ejemplo de pregunta asociada a un solo ítem

<p>En el texto aparece la palabra "coz". ¿Cuál de las siguientes palabras tiene el mismo significado?</p> <p>A. Salto</p> <p>B. Susto</p> <p>C. Pelota</p> <p>D. Patada</p>	
Indicador de logro	<p>ÍTEM 6CEXXXX. Deduce el significado de palabras y expresiones con ayuda del contexto.</p>

Figura 2.3. Ejemplo de pregunta asociada a varios ítems

<p>Con ayuda del programa y de las fotografías, redacta la noticia de forma clara, explicando y describiendo con detalle en qué consistió cada actividad. Debes incluir también un titular que sea divertido y contar, al menos, la siguiente información:</p> <ul style="list-style-type: none"> - ¿Cuándo se celebraron? - ¿Dónde se celebraron? - ¿Qué ocurrió? - ¿Cómo ocurrió? - ¿Quién o quiénes participaron? 	
Indicador de logro	<p>ÍTEM 6EEXXXX. Transmite las ideas con claridad, sentido y progresiva corrección.</p>
	<p>ÍTEM 6EEXXXX. Transforma la noticia en hechos cotidianos cercanos a su realidad ajustándose a la estructura, registro y lenguaje propios del género.</p>

A partir de la respuesta a un ítem debe poder inferirse, con cierto grado de fiabilidad y validez, un nivel de dominio de la competencia evaluada a través del indicador de logro al que se asocie (INTEF, 2014).

➤ Unidades de evaluación

Entendemos por unidad de evaluación el conjunto formado por un **estímulo** o centro de interés y varias **preguntas e ítems** asociados. En cada unidad de evaluación se presenta una situación o caso mediante un estímulo a partir del cual se plantea un conjunto de cuestiones.

➤ Guía de codificación

La guía de codificación debe incluir criterios de corrección y codificación para todos los ítems de los que se componga la prueba de evaluación a la que esté asociada. Es una labor fundamental ya que, de esta manera, se podrá indicar inequívocamente a quienes codifiquen las pruebas de evaluación cómo han de valorar las respuestas de las personas evaluadas.

Por cada una de las preguntas que compongan la prueba de evaluación debe completarse una plantilla o ficha de corrección donde se indique, entre otros datos, el ítem o ítems asociados a la pregunta. Además, para cada ítem se incluirá el indicador de logro que valora, el tipo de pregunta que se está formulando, la respuesta o respuestas correctas a la misma, los criterios de codificación, etc. A cada ítem debe corresponderle un único indicador de logro. Con el objetivo de orientar al codificador, si es posible se incorporarán asimismo ejemplos de respuestas correctas, parcialmente correctas, e incorrectas.



Elaboración de pruebas de evaluación

Indicadores de logro

¿Qué son?

Los indicadores de logro son descripciones concretas de lo que el alumno o alumna debe saber, debe comprender, es capaz de comprender y es capaz de hacer. Son, en definitiva, evidencias observables y medibles que permiten graduar el rendimiento o desempeño alcanzado en la competencia.

¿Cómo redactarlos?

Acción

¿Qué realiza el alumno o alumna?

- Ej. Identifica
- Ej. Maneja, analiza e interpreta

Contexto

¿En qué contexto?

- Ej. en situaciones familiares
- Ej. en aplicaciones informáticas

Conocimiento involucrado

¿Qué conocimientos, capacidades o actitudes?

- Ej. datos cualitativos y cuantitativos
- Ej. variables económicas

Finalidad

¿Para qué?

- Ej. para reconocerlos en la vida real.
- Ej. para presentar sus valoraciones de carácter.

Tipos de preguntas y utilidad

En una prueba de evaluación, cada unidad debe estar configurada por preguntas de diferentes formatos. Habitualmente la clasificación de los tipos de preguntas se realiza atendiendo a su forma de respuesta. Existen diferentes clasificaciones de los ítems. En esta guía se considerarán las preguntas clasificadas en alguno de los siguientes formatos: **respuesta cerrada**, **semiconstruida** y **abierta**. Dentro de cada uno de estos tipos, existen a su vez diferentes formatos de preguntas. Debe tenerse en cuenta que el tipo de pregunta utilizado y su formato pueden afectar a la dificultad y la validez de la misma (Benson y Crocker, 1979).

Respecto a la utilidad de cada tipo de pregunta, las de respuesta cerrada y las de respuesta semiconstruida se consideran adecuadas para medir:

- Conocimientos de terminología, hechos específicos, principios, métodos y procedimientos.
- La habilidad para identificar hechos y principios, interpretar relaciones de causa-efecto y justificar métodos y procedimientos.
- La habilidad para realizar operaciones aritméticas concretas (Stecher, *et al.*, 1997).
- La habilidad para identificar la adecuación de afirmaciones de hechos, definición de términos, frases de principios y similares (Cortés de las Heras, 2019).
- La habilidad para reconocer relaciones de causa-efecto o aspectos de lógica (Cortés de las Heras, 2019).
- La habilidad para reconocer y recordar datos básicos.
- La capacidad para la interpretación simple de datos.

En cuanto a las preguntas de respuesta abierta, son de utilidad en la evaluación de la ejecución de una actividad compleja, la evaluación de la resolución de un problema o la evaluación de la capacidad de desarrollo de una respuesta.

En los siguientes apartados se describen los diferentes tipos de preguntas y además se incluye el formato habitual de codificación de las respuestas, información que siempre debe ir incluida en la guía de codificación.

➤ Preguntas de respuesta cerrada

Las preguntas de respuesta cerrada son aquellas que no requieren de una elaboración o construcción por parte de la persona que realiza la prueba, sino que solamente precisan que esta seleccione una opción de entre varias ofrecidas, relacione diferentes posibilidades, ordene ciertas opciones, etc. En este tipo de cuestiones se identifica el término “pregunta” con el de “ítem”.

Características

- **Versatilidad:** pueden evaluar el resultado de un aprendizaje a distintos niveles: identificación, aplicación, análisis, etc. Son adecuadas para evaluar la habilidad de comprender conceptos y principios, plantear inferencias, elaborar juicios, interpretar datos, aplicar información, valorar razonamientos, etc.
- **Eficiencia:** son sencillas de administrar y pueden ser codificadas de forma objetiva mediante procedimientos mecánicos o automáticos.

- **Validez:** dada la rapidez con la que se pueden contestar los ítems de respuesta cerrada, las pruebas en las que predominan este tipo de ítems pueden abarcar una muestra relativamente más amplia de niveles de desarrollo de la competencia que se quiera evaluar, lo que aumenta la validez del resultado de la evaluación.
- **Limitación:** este tipo de ítems presenta algunas limitaciones en la evaluación de capacidades como las de organizar información, elaborar explicaciones o desarrollar ideas creativas. Por este motivo, en las pruebas de evaluación habitualmente se combinan este tipo de preguntas con otras de diferentes formatos.

Subtipos

Existen varios subtipos de preguntas de respuesta cerrada:

- **Preguntas o ítems de respuesta cerrada de opción (o elección) múltiple.** Se presenta un enunciado con varias alternativas de las cuales solo una es correcta. Los componentes que configuran este tipo de preguntas son:
 - **Tronco o vástago:** parte inicial de la pregunta en la que se especifica la tarea. Puede ser una pregunta, instrucciones, una frase incompleta, etc.
 - **Opciones:** todas las alternativas de respuesta al tronco.
 - **Clave:** respuesta correcta.
 - **Distractores:** opciones incorrectas: se proponen para “distrar” al alumnado con conocimiento parcial debido a su apariencia de opciones plausibles.
 - En ocasiones el ítem se complementa con alguna información adicional. En las pruebas de evaluación en las que el ítem forma parte de una unidad de evaluación, esta información a veces se localiza en el estímulo correspondiente.

Evidentemente, cuanto mayor sea el número de distractores mayor fiabilidad tienen estas preguntas, ya que son menos susceptibles de ser acertadas al azar. La Figura 2.4 muestra un ejemplo de ítem de opción múltiple. Más adelante se tratará con mayor detalle sobre la elaboración de este tipo de preguntas y sobre el diseño de distractores adecuados y el número recomendado de estos.

Figura 2.4. Ejemplo de ítem de opción múltiple

¿Cuántos meses tiene un año? Elige la respuesta correcta.

- A. 2
- B. 11
- C. 12
- D. 17

- **Preguntas o ítems de respuesta cerrada de opción múltiple compleja.** Estas preguntas pueden presentar varios planteamientos:

A) El más habitual es el de un conjunto de afirmaciones con respuesta dicotómica, es decir, a cada una de las cuales debe responderse eligiendo entre dos alternativas: verdadero/falso, sí/no, causa/consecuencia, etc.

Es importante que cada una de las afirmaciones planteadas necesite, para su resolución, un razonamiento independiente del requerido para las otras, ya que de otra manera las respuestas estarían relacionadas entre sí. Si esta condición no se verificara, quizá sería más adecuado reformular la pregunta en el formato de opción múltiple. Sin embargo, el conjunto de afirmaciones que se plantean agrupadas en un mismo ítem debe tener coherencia interna y dirigirse a la evaluación de un mismo indicador de logro; si esto no ocurriera, se puede desdoblar el ítem en dos o más en el momento de su elaboración. La Figura 2.5 ilustra este tipo de pregunta.

Figura 2.5. Ejemplo de ítem de respuesta cerrada con opciones dicotómicas

Señala qué razones de las siguientes pueden hacer que la introducción de una especie invasora provoque la extinción de una especie autóctona (que son las que viven de manera natural en un determinado sitio).

	Sí	No
La especie invasora compite con la autóctona por el alimento.		
La especie invasora es más pesada que la autóctona.		
La especie invasora puede introducir nuevas enfermedades en el ecosistema.		
La especie invasora ocupa todo el espacio donde habita la especie autóctona.		

- B) Las preguntas de este tipo pueden presentarse en otros formatos y no siempre como tabla con afirmaciones independientes. Por ejemplo, pueden presentar respuestas no dicotómicas. En la Figura 2.6 se muestran dos ítems, el primero con dos respuestas dicotómicas, pero presentado con un formato diferente a la tabla, y el segundo con opciones de respuesta no binarias.

Figura 2.6. Ejemplos de ítem de respuesta cerrada de opción múltiple compleja

Si la corona estuviese hecha de oro, ¿crees que podría conducir la electricidad? **Marca la palabra en cada recuadro que corresponda.**

Sí Aislante.

No , porque el oro es un material Conductor.

En las columnas 1 y 3 de la tabla siguiente se incluyen hechos que preceden a la toma de la Bastilla según el **Documento 1**. Ahora vas a analizar si, entre los hechos incluidos en cada fila, hay relación de causa-efecto. Para ello, lee cada fila de izquierda a derecha y marca con una X la casilla correspondiente de la columna central. **No olvides marcar SOLO una X en cada fila.**

(1785-89) Varios años de malas cosechas	<input type="checkbox"/> fue causa de <input type="checkbox"/> fue consecuencia de <input type="checkbox"/> sin relación	Hambre y protestas de las clases populares
Luis XVI reúne en Versalles a los Estados Generales (Cortes) para aprobar nuevos impuestos	<input type="checkbox"/> fue causa de <input type="checkbox"/> fue consecuencia de <input type="checkbox"/> sin relación	El gobierno de Luis XVI tenía grandes deudas financieras
Se extienden las ideas de la Ilustración contra el absolutismo y el Antiguo Régimen	<input type="checkbox"/> fue causa de <input type="checkbox"/> fue consecuencia de <input type="checkbox"/> sin relación	El tercer estado se reúne en Asamblea Nacional para acabar con el absolutismo
La nobleza y el clero dudan si apoyar al rey o unirse a la Asamblea Nacional	<input type="checkbox"/> fue causa de <input type="checkbox"/> fue consecuencia de <input type="checkbox"/> sin relación	El rey recurre al ejército para reprimir los tumultos y disolver la Asamblea Nacional

- C) Preguntas o ítems de opción múltiple con varias opciones correctas. Son ítems con la misma estructura que las preguntas de opción múltiple en los que varias de las opciones presentadas son correctas. La Figura 2.7 muestra un ejemplo de pregunta de este tipo, en la que en el enunciado se indica el número de opciones que deben marcarse.

Figura 2.7. Ejemplo de pregunta o ítems de opción múltiple con varias opciones correctas en la que se indica el número de opciones que se deben marcar

Las pilas de los auriculares de María se han gastado, y las tira en el punto limpio del hotel. Marca las DOS razones por las que lo que ha hecho María es beneficioso para el medio ambiente:

- Al tirar las pilas al contenedor adecuado ayudamos a mantener desordenada la basura.
- La mayoría de las pilas contienen sustancias muy contaminantes que ponen en riesgo a los ecosistemas de la tierra.
- Las pilas perjudican el aire que respiramos por los gases que desprenden.
- Al tirar las pilas al contenedor adecuado estamos beneficiando la salud de todos los seres vivos respecto a su alimentación.

La Figura 2.8 muestra una pregunta similar, pero en este caso el enunciado no proporciona el número de opciones correctas existentes. Es importante indicar que en este tipo de ítems es más conveniente indicar el número exacto de opciones que la persona que realiza la prueba debería marcar, ya que en caso contrario puede dar lugar a ítems con baja calidad o con poca capacidad de discriminación, esto es, la probabilidad de que el alumno obtenga una puntuación positiva no se relaciona con su nivel de competencia. Esto ocurre especialmente cuando, además, la pregunta ofrece codificación como “parcialmente correcta”.

Figura 2.8. Ejemplo de pregunta o ítems de opción múltiple con varias opciones correctas en la que no se indica el número de opciones que se deben marcar

Cuando vuelva del viaje, María piensa subir algunas fotos a su blog personal. Marca aquellas cosas que NO debería hacer, por ser peligroso desde el punto de vista de la seguridad y la protección en Internet.

- Añadir datos personales, como números de teléfono, dirección...
- Utilizar usuario y contraseña privada.
- Publicar algo que han recibido como privado sin pedir permiso.
- Navegar por Internet con el permiso y control de sus padres.
- Publicar fotos personales sin que lo sepan sus padres.
- Contar detalles personales de sus vidas.

- **Pregunta de respuesta cerrada de emparejamiento o relación:** en estas preguntas la persona evaluada debe emparejar los elementos presentados en dos o más conjuntos de información según la relación existente entre ellos. También se clasificarían dentro de este tipo de preguntas aquellas consistentes en rellenar huecos siempre que se proporcione un listado de términos o elementos entre los que haya que seleccionar. Para relacionar los distintos elementos se pueden usar flechas, números, letras, etc. Hay que tener en cuenta que las respuestas con flechas pueden dificultar la corrección, por lo que en el diseño del ítem debe valorarse la modificación del formato de pregunta para

que la respuesta se muestre con claridad. La Figura 2.9 muestra un ejemplo de este tipo de ítem, con el formato de unir con flechas, y su reformulación para que la respuesta tenga mayor claridad. La Figura 2.10 muestra una pregunta en la que hay que rellenar huecos seleccionándolos de un listado dado.

Figura 2.9. Ejemplos de ítem de respuesta cerrada de emparejamiento o relación y posibilidad de reformulación

Una de las actividades consistía en realizar un recorrido de descenso de una ladera no muy inclinada en bicicleta. **Relaciona, uniendo con flechas, los operadores mecánicos de una bicicleta con su utilidad.**

Operador	Utilidad
Manivelas	Facilitan el desplazamiento de la bicicleta.
Engranajes	Transmiten el movimiento de los pedales a la rueda trasera.
Palancas	Hacen girar los platos.
Ruedas	Accionan mecanismos como los frenos.

Reformulación del ítem

Una de las actividades consistía en realizar un recorrido de descenso de una ladera no muy inclinada en bicicleta. Relaciona los operadores mecánicos de una bicicleta con su utilidad escribiendo la letra correspondiente en la columna de en medio.

Operador	Escribe aquí la letra correspondiente	Utilidad
A. Manivelas		Facilitan el desplazamiento de la bicicleta.
B. Engranajes		Transmiten el movimiento de los pedales a la rueda trasera.
C. Palancas		Hacen girar los platos.
D. Rueda		Accionan mecanismos como los frenos.

Figura 2.10. Ejemplos de ítem de respuesta cerrada de rellenar huecos

Elige las palabras adecuadas de la nube para rellenar el texto (hay una por cada hueco):

encantada
imaginativa
comprensiva
juguetona



Amanda era una niña muy _____ a la que siempre se le ocurrían nuevos inventos. Un día los padres de Amanda le regalaron un perrito, y, aunque ella estaba _____ con el regalo, el cachorro no la dejaba estudiar mucho, porque siempre estaba corriendo y llamando la atención de su dueña. Y Amanda, que también era _____, dejó de hacer inventos. En el barrio comenzaron a echar de menos a su inventora favorita, así que la directora del colegio, que era muy _____, dejó a Amanda llevar al perro al colegio a cambio de que siguiera inventado.

Corrección y codificación de las preguntas de respuesta cerrada

En general, el proceso de corrección de las preguntas o ítems de respuesta cerrada se limita a la codificación, que consiste en registrar las respuestas del alumnado mediante códigos numéricos para su posterior tratamiento y análisis usando medios informáticos. Dependiendo de las instrucciones específicas que se proporcione a los codificadores y del tipo de análisis que vaya a realizarse con los datos, el registro de información puede ser más o menos exhaustivo. En la ficha de codificación de cada una de las preguntas o ítems de respuesta cerrada se debe especificar con claridad la **clave** o **respuesta correcta**. La Figura 2.11 muestra las respuestas correctas a las preguntas de las figuras anteriores.

Figura 2.11. Ejemplos de claves o respuestas correctas según el tipo de preguntas

Tipo de pregunta	Respuesta correcta
Ítem de opción múltiple (Figura 2.4)	C
Ítem de respuesta cerrada con opciones dicotómicas (Figura 2.5)	Sí – No – Sí – No
Ítem de respuesta cerrada de opción múltiple compleja (Figura 2.6)	Sí – Conductor
Ítem de respuesta cerrada de opción múltiple compleja (Figura 2.6)	Fue causa de – Fue consecuencia de – Fue causa de – Sin relación
Ítem de opción múltiple con varias opciones correctas en el que se indica el número de opciones que se deben marcar (Figura 2.7)	Opciones 2. ^a y 3. ^a
Ítem de opción múltiple con varias opciones correctas en el que no se indica el número de opciones que se deben marcar (Figura 2.8)	Opciones 1. ^a , 3. ^a , 5. ^a y 6. ^a
Ítem de respuesta cerrada de emparejamiento o relación (Figura 2.9)	Manivelas: Hacen girar los platos. Engranajes: Transmiten el movimiento de los pedales a la rueda trasera. Palancas: Accionan mecanismos como los frenos. Ruedas: Facilitan el desplazamiento de la bicicleta.
Ítem de respuesta cerrada de rellenar huecos (Figura 2.10)	imaginativa – encantada – juguetona – comprensiva

Codificación de las preguntas de opción múltiple

La codificación de este tipo de ítems consiste, en general, en registrar la respuesta sin valorar si es correcta o incorrecta, algo de lo que se encargará el procedimiento posterior de análisis, que será mucho más informativo, ya que permite analizar también los porcentajes de elección de los distractores, estudiar su calidad y analizar el tipo de errores que el alumnado comente con más frecuencia.

Además, la ficha de codificación de este tipo de ítem suele incluir un código para distinguir las respuestas nulas (Figura 2.12), por ejemplo, si en la respuesta se selecciona más de una opción; o en blanco, que facilita la identificación de la pregunta como no adecuada al nivel si un gran porcentaje de individuos la dejaran sin responder; o ítem no alcanzado, que sirve para probar si el tiempo para la realización de la prueba es adecuado si tras el análisis de datos se muestra que no se completaron las últimas preguntas.

Figura 2.12. Codificación de una pregunta de opción múltiple

Se codificará la respuesta que haya dado la persona evaluada: A, B, C o D.
Código 7: respuesta nula.
Código 9: respuesta en blanco.

Codificación de preguntas de respuesta cerrada distintas a las de opción múltiple

La ficha y las instrucciones de codificación de este tipo de preguntas pueden incluir la grabación de las respuestas proporcionadas por el alumnado, bien directamente o bien convertidas en códigos que se habrán definido *ad hoc* para cada pregunta.

Independientemente de que se realice o no el registro anterior, las respuestas a este tipo de preguntas suelen codificarse con los códigos 1, representando la respuesta que se valorará como correcta, y 0, representando la respuesta que se valorará como incorrecta. Dependiendo de la complejidad de la pregunta, también pueden establecerse los códigos 2, 1 y 0 para valorar las respuestas como correctas, parcialmente correctas e incorrectas, respectivamente. Adicionalmente, se suele incluir un código para registrar las respuestas en blanco (9). Las Figuras 2.13 y 2.14 muestran la codificación de este tipo de preguntas. En algunas guías, por ejemplo, en las pruebas internacionales de evaluación como TIMSS o PIRLS, los códigos de corrección tienen dos dígitos. El segundo de ellos sirve para agrupar los tipos de respuestas del alumnado, tanto correctas como incorrectas. Así, se podrían codificar las respuestas incorrectas con los códigos 01, 02, 03, etc., cada uno de ellos describiendo un posible error que el alumnado pueda cometer.

Figura 2.13. Codificación de respuesta cerrada dicotómica

Respuesta correcta: sí – sí – no – no – sí
Código 1: respuesta correcta o con un fallo en alguna de las afirmaciones.
Código 0: cualquier otra respuesta.
Código 9: respuesta en blanco.

Figura 2.14. Codificación de respuesta cerrada dicotómica incluyendo el código de parcialmente correcta

Respuesta correcta: verdadero – falso – verdadero – falso
Código 2: respuesta correcta.
Código 1: se admite un fallo.
Código 0: cualquier otra respuesta.
Código 9: respuesta en blanco.

Conviene señalar que el código 1 (en su caso, el código 2) se identifica con la valoración máxima positiva de la respuesta que no necesariamente requiere una respuesta completamente correcta, ya que puede admitirse algún error en la respuesta según se defina en la ficha de codificación.

➤ Preguntas de respuesta semiconstruida

Características

Estas preguntas exigen la elaboración de una respuesta por parte de la persona evaluada. En general solicitan que esta rellene una o varias casillas o espacios con una palabra o cifra, complete varias frases, o responda a alguna cuestión con una respuesta corta. Se clasificarán en esta categoría las preguntas siempre que no se proporcione un listado de posibles respuestas. Estas preguntas pueden exigir una breve construcción, por ejemplo, un número que dé respuesta a un problema matemático o una palabra que complete una frase, o bien que la persona evaluada responda a alguna cuestión recordando algún conocimiento adquirido.

Existen dos grandes subtipos de preguntas de respuesta semiconstruida, según la respuesta correcta sea única e inequívoca o que existan diferentes posibilidades de respuesta, todas ellas válidas.

- A. Pregunta de respuesta semiconstruida con respuesta única inequívoca (tipo hueco o GAP).** La persona evaluada debe rellenar uno o varios espacios en blanco (*gaps*, en inglés) con la información disponible en la propia pregunta o en algún estímulo al que se asocie. Debe señalarse que, aunque en principio al redactar la pregunta pueda pensarse que la respuesta es única, en los estudios piloto suele comprobarse que hay muchas formas de expresar esa respuesta. Por ello, en la ficha de codificación debe tenerse en cuenta el indicador de logro que el ítem persigue evaluar; en general no se penalizará por faltas de ortografía, letra poco clara, escritura en mayúsculas o minúsculas o con cifras o letras, por ejemplo. Si alguna de estas circunstancias fuera a valorarse de forma negativa deberá indicarse claramente en el enunciado lo que se pretende que la persona evaluada haga, con indicaciones del tipo: “escribe con cifras”. También, puede incluirse un ejemplo orientativo que guíe a la persona evaluada en el formato de respuesta que se persigue, como se muestra en la Figura 2.16. Las Figuras 2.15 y 2.16 muestran dos ejemplos de este tipo de preguntas.

Figura 2.15. Ejemplo de pregunta semiconstruida con respuesta única e inequívoca

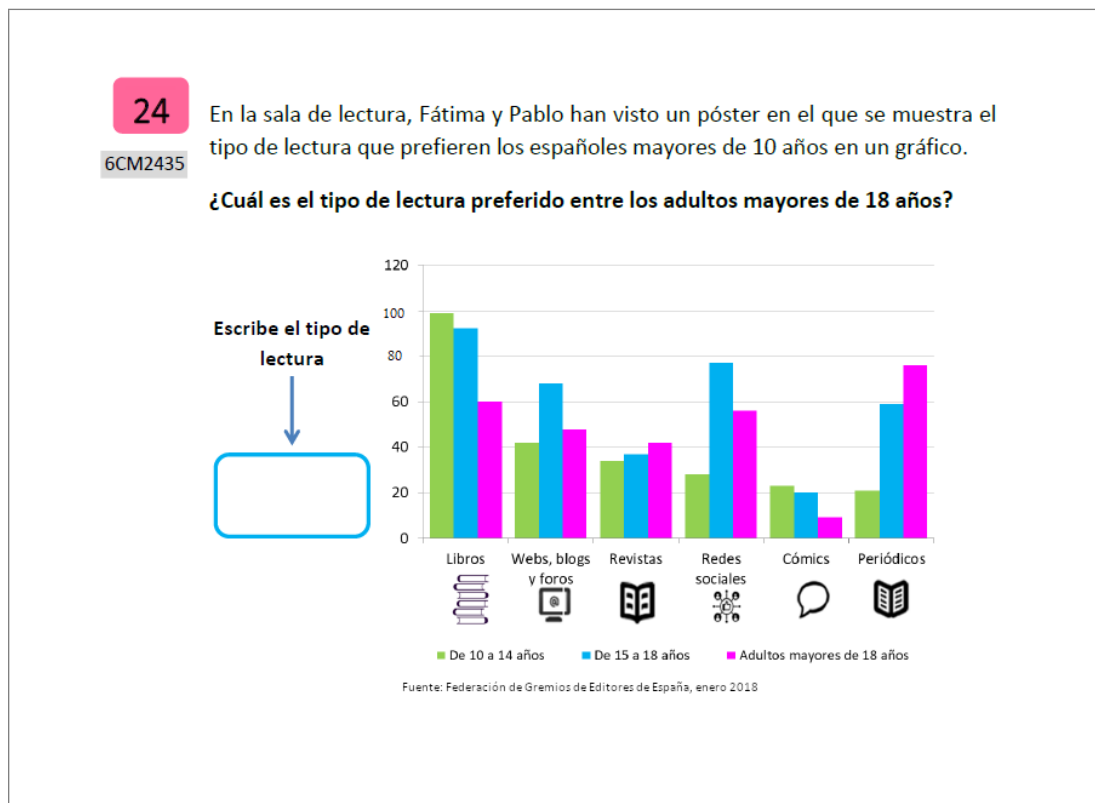


Figura 2.16. Ejemplo de pregunta semiconstruida con ejemplo orientativo de respuesta

Fíjate en el dibujo en el que se muestra el mapa del parque de atracciones.

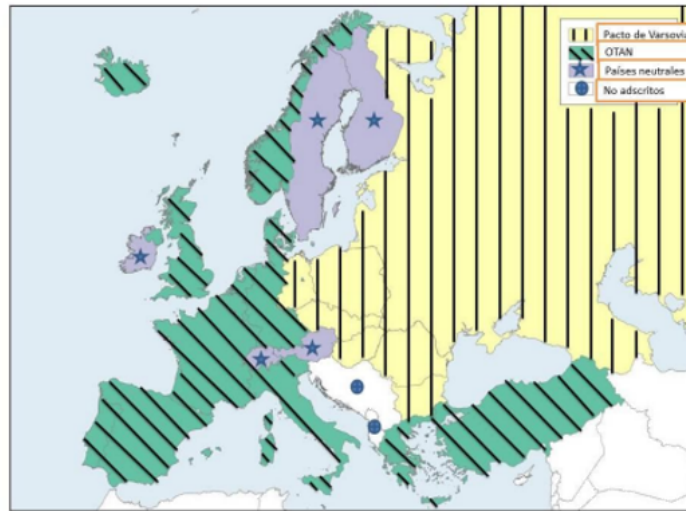


Ahora completa siguiendo el ejemplo:

El puesto de palomitas, los baños y la montaña rusa tienen forma de	Cuadrado
Las sillas voladoras tienen forma de	
La noria y el tiovivo tienen forma de	
La sala de espectáculos tiene forma de	
La casa del terror y el tren de la bruja tienen forma de	

- B. Pregunta de respuesta semiconstruida con varias respuestas posibles.** En este tipo de preguntas existen numerosas posibilidades de respuesta, todas ellas correctas, y que no solo se diferencian en el modo de registrarlas (mayúsculas o minúsculas, cifras o letras...). Por ejemplo, se incluirían en este grupo de preguntas las que piden que se indiquen dos adjetivos de los que aparecen en un texto (si en el texto figuran más de dos) o que se citen dos obras de un autor, por ejemplo.

Figura 2.17. Ejemplo de pregunta semiconstruida con varias respuestas posibles



En el anterior mapa de Europa aparecen los países de la OTAN y del Pacto de Varsovia y otros países neutrales o no adscritos. **Completa los espacios dejados en blanco en el siguiente texto.**

- A) Ambas organizaciones tienen o tenían como principal objetivo (1) _____.
- B) En la actualidad, una de ellas ya no existe, se trata de (2) _____.
- C) La OTAN, que significa (3) _____, es una organización liderada por (4) _____ y aglutina a sus países aliados de la Europa (5) _____. La economía de estos países es básicamente de tipo (6) _____.
- D) El Pacto de Varsovia fue liderado por (7) _____ y aglutinó a sus países aliados de Europa (8) _____. La economía de estos países era fundamentalmente de tipo (9) _____.

Corrección y codificación de las preguntas de respuesta semiconstruida

Las preguntas de respuesta semiconstruida, en general, no puede codificarlas un ordenador, ya que ello supondría que los algoritmos diseñados tuvieran todas las indicaciones necesarias para asegurarse de que se evalúa el indicador de logro para el que la pregunta fue diseñada y no se penaliza por no alcanzar otros diferentes. Como ya se ha señalado, en la corrección de aquellas preguntas que requieran rellenar casillas o espacios con palabras o números deberán tenerse en cuenta las indicaciones en cuanto al uso de letras, números, mayúsculas y minúsculas, etc. que aparezcan reflejadas en la guía de codificación.

La codificación habitual de estas preguntas es con 1, respuesta correcta, 0 respuesta incorrecta y 9, respuesta en blanco. En las preguntas de respuesta semiconstruida se debe especificar con claridad la **clave de respuesta correcta** (puntuación máxima o completa) y, si hay posibilidad de puntuación parcial, las **posibilidades asociadas**.

Como se ha dicho, según la complejidad del ítem, puede concederse puntuación parcial a algunas opciones de respuesta incompleta: se suelen codificar con código 2, puntuación positiva completa, o código 1, puntuación positiva parcial. La puntuación parcial suele darse a aquellas preguntas en las que se responda correctamente a todas las opciones salvo a una o alguna de ellas, o a las preguntas que presenten cierto grado de complejidad para alcanzar la puntuación completa.

La Figura 2.18 muestra la codificación de las preguntas que se recogen en las figuras 2.15, 2.16 y 2.17.

Figura 2.18. Ejemplo de codificación de las preguntas anteriores

Codificación de la pregunta de la Figura 2.15

Respuesta correcta: periódicos

Código 1: respuesta correcta.

Código 0: cualquier otra respuesta.

Código 9: respuesta en blanco.

Codificación de la pregunta de la Figura 2.16

Código 1: escribe correctamente el nombre de las figuras: pentágono, círculo (se admite circunferencia), triángulo y rectángulo (se admite cuadrilátero/paralelogramo). No se tiene en cuenta la ortografía.

Código 0: cualquier otra respuesta.

Código 9: respuesta en blanco.

Codificación de la pregunta de la Figura 2.17³

Respuesta correcta: (1) la cooperación en la defensa militar entre sus países miembros. (2) el Pacto de Varsovia. (3) Organización del Tratado del Atlántico Norte (4) Estados Unidos (5) occidental, capitalista o democrática (6) capitalista o de libre mercado (7) La Unión de Repúblicas Socialistas Soviéticas (URSS) (8) del Este / oriental / comunista (9) comunista o economía planificada.

Código 1: completa correctamente al menos seis espacios.

Código 0: cualquier otra respuesta.

Código 9: respuesta en blanco.

➤ Preguntas de respuesta abierta o construida

Características

Las preguntas de respuesta abierta o construida son aquellas que exigen construcción por parte de la persona evaluada y, por tanto, necesitan de un corrector o correctora que revise el procedimiento llevado a cabo para alcanzar la respuesta, compare este proceso con los ejemplos recogidos en la guía de codificación y decida qué puntuación le corresponde. Este tipo de preguntas resulta particularmente importante para valorar el proceso de reflexión del alumnado, por lo que suelen emplearse para evaluar la calidad del pensamiento más que el resultado final (OCDE, 2017).

En ocasiones se definen las preguntas de respuesta construida como aquellas en las que existe la necesidad de alcanzar un resultado único, aunque para llegar a este puedan describirse diferentes procedimientos o métodos. Este sería el caso, por ejemplo, de un problema de matemáticas en el que puedan llevarse a cabo distintos procedimientos para obtener un mismo resultado. Cuando se codifican este tipo de preguntas deben valorarse, por tanto, procedimiento y resultado, lo que posibilita el establecimiento de distintos niveles de corrección en la respuesta en función del grado de desarrollo competencial evidenciado.

La principal diferencia entre las preguntas de respuesta construida y las de respuesta abierta es que estas últimas no tienen una sola respuesta correcta inequívoca, sino que admiten respuestas diversas todas ellas igualmente válidas y correctas. Dentro de este tipo se engloban las producciones escritas o las respuestas que exigen el desarrollo de procedimientos y la obtención de resultados, entre otras. Una pregunta abierta normalmente suele estar asociada a más de un ítem, sobre todo en el caso de las producciones escritas. Cada ítem asociado evalúa la consecución de un único indicador de logro y aparece reflejado en la correspondiente ficha de codificación según se recoge en la tabla de especificaciones.

Corrección y codificación de las preguntas de respuesta construida

En las preguntas de respuesta construida **sin posibilidad de puntuación parcial** es necesario que se especifique con la máxima precisión posible la respuesta o respuestas que se consideran correctas. En el caso de tratarse de preguntas **con posibilidad de puntuación parcial** hay que especificar:

- La respuesta o respuestas correctas.
- La respuesta o respuestas parcialmente correctas.

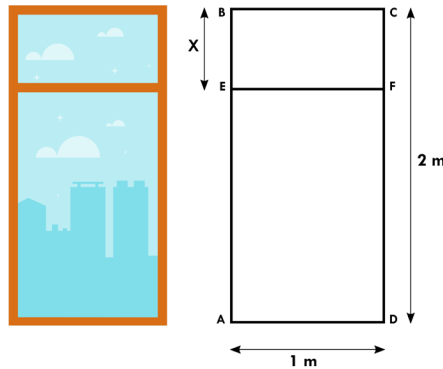
A cada posibilidad pueden asociarse una serie de ejemplos de respuesta con su correspondiente clasificación.

³ Debe observarse que es improbable que el alumnado redacte su respuesta con las mismas palabras que la respuesta correcta que se propone en esta pregunta. Sin embargo, en las respuestas a las preguntas 2.15 y 2.16 el término solicitado es único.

En las Figuras 2.19 y 2.20 se muestran dos ítems con respuesta construida y sus correspondientes guías de codificación. Debe resaltarse que en este tipo de ítems es necesario que el enunciado especifique si solo se tendrá en cuenta el resultado, caso en el que se clasificaría como pregunta semiconstruida, o si también se valorará el procedimiento, en cuyo caso se clasificará como de respuesta construida.

Figura 2.19. Ejemplo de ítem con respuesta construida

La siguiente figura representa un esquema de la ventana que queremos colocar en el salón.



Si los rectángulos ABCD y EBCF que forman la ventana son semejantes, **¿cuánto mide la longitud x indicada en el dibujo?** Razona tu respuesta.

Codificación:

Código 1: respuesta correcta.


Establece la proporción $\frac{2}{1} = \frac{1}{x} \Rightarrow x = 0,5m$ o utiliza cualquier otro razonamiento correcto.

Es necesario que indique correctamente la unidad de magnitud.

Código 0: cualquier otra respuesta.

Código 9: respuesta en blanco.

Figura 2.20. Ejemplo de ítem con respuesta construida

Los amigos de Ana se han puesto de acuerdo para regalarle un balón por su cumpleaños. Cada amigo ha puesto 2€. Si de 3.º A han participado 7 amigos y de 3.º B, 12. ¿Cuánto ha costado el balón? 

Contesta indicando las operaciones que hay que hacer.

✎ Operaciones:

Codificación

Código 2

Suma 7 y 12 y el resultado lo multiplica por dos y obtiene el resultado correcto (38),

$$(7 + 12) \times 2 = 38$$

o multiplica 7 por 2 y 12 por 2 y suma los resultados,

$$7 \times 2 + 12 \times 2 = 38$$

Código 1

Plantea bien las operaciones y falla en el cálculo.

Plantea mal las operaciones, o no las plantea, pero da el resultado correcto.

Código 0: cualquier otra respuesta.

Código 9: respuesta en blanco.

Corrección y codificación de las preguntas de respuesta abierta corta

Las preguntas de respuesta abierta se responden libremente, sin limitaciones, salvo las indicaciones o instrucciones incluidas en la cuestión formulada. En las pruebas de las que estamos tratando la respuesta suele ser una producción escrita (u oral) que puede consistir, por ejemplo, en realizar un resumen, proporcionar una explicación o una descripción más o menos larga o responder de forma detallada o profunda a la pregunta planteada.

Este tipo de preguntas presentan como ventaja la posibilidad de medir procesos cognitivos más complejos, como el razonamiento. Sin embargo, la codificación resulta una tarea mucho más compleja y requiere mayor tiempo de realización. Además, la corrección de estas cuestiones está muy sujeta a la subjetividad del corrector: para minimizarla, a la hora de codificar las preguntas de respuesta abierta la guía de codificación debe proporcionar pautas precisas y concretas, dejando el menor margen posible a la valoración personal del corrector. Además, la formación de los correctores y la unificación de los criterios de codificación es imprescindible para tratar de reducir los posibles sesgos debidos a la corrección y codificación de las respuestas.

Si la respuesta requerida en la pregunta abierta es corta, su codificación suele realizarse con una guía de valoración similar a las descritas en la sección dedicada a las preguntas semiconstruidas. Tres ejemplos de este tipo de preguntas se muestran en las Figuras 2.21, 2.22 y 2.23. En el momento de elaboración de las guías de codificación, los expertos encargados del trabajo proponen *a priori* diferentes niveles de elaboración de las respuestas, asignando un código a cada uno de ellos. Más adelante, los ítems suelen probarse con una muestra reducida de alumnos en lo que se denomina “prueba piloto”. Del análisis de las respuestas reales del alumnado a estas pruebas se obtiene información sobre el funcionamiento de las preguntas, lo que permite mejorar su guía de codificación y proporcionar ejemplos ilustrativos de las respuestas correctas y parcialmente correctas.

Las preguntas abiertas pueden relacionarse con la consecución de un único indicador de logro, y, por lo tanto, con un único ítem, o con varios indicadores. En ese caso, suelen identificarse con varios ítems. La pregunta que se recoge en la Figura 2.21 está midiendo, al menos, dos indicadores de logro diferentes, según estén definidos en el marco de la evaluación. En primer lugar, la capacidad de resumir el texto y, en segundo, la ortografía.

Figura 2.21. Ejemplo de pregunta abierta corta y su guía de codificación

Resume con tus propias palabras el contenido del texto en cinco líneas como máximo.
Codificación:
Código 2
Comete, como máximo, una falta de ortografía y dos errores en la colocación de tildes. En ningún caso se admitirán respuestas en las que la persona evaluada haya reproducido literalmente fragmentos del texto.
El resumen completo debe contener, al menos, las siguientes ideas:
<ul style="list-style-type: none"> • Manrique (el protagonista) es un joven con mucha imaginación, muy soñador. • Decide salir por la noche y cree ver una misteriosa mujer. • La persigue, pero no consigue alcanzarla. • Con el paso del tiempo, no consiguen hacerle entrar en razón y continúa buscando un amor ideal e imposible.
Ejemplo
Manrique es un joven con mucha fantasía. Una noche sale a pasear y cree ver a una mujer, a la que comienza a perseguir, aunque no consigue alcanzarla. Años más tarde, continúa pensando en encontrarla, aunque cree que es imposible.
Código 1
Incluye tres de las cuatro ideas indicadas.
Comete, como máximo, dos faltas de ortografía y tres errores en la colocación de tildes.
Código 0: cualquier otra respuesta.
Código 9: respuesta en blanco.

Figura 2.22. Ejemplo de ítem con respuesta abierta del estudio PIRLS 2016

Da una razón por la que Jack era el mejor para ir a pedir ayuda. Da una razón por la que Jack NO era el mejor para ir a pedir ayuda.

Codificación:

Código 2: comprensión total.

La respuesta proporciona una razón creíble basada en el texto por la que Jack era el mejor para buscar ayuda **Y** una razón por la que Jack no era el mejor para buscar ayuda de la lista que hay a continuación.

Por qué Jack era el mejor:

- Sabía a dónde ir.
- Sabía cómo llegar a la casa del granjero/al valle.

Por qué Jack NO era el mejor:

- Es demasiado viejo para llegar tan lejos.
- Estaba demasiado lejos para que él viajase hasta allí.
- No se las podía arreglar en esa distancia.
- Tenía que quedarse y vigilar a las ovejas.
- Es demasiado lento/débil.
- Últimamente apenas se las arregla con el rebaño.

Código 1: comprensión parcial.

La respuesta proporciona una razón creíble basada en el texto por la que Jack era el mejor para buscar ayuda **O** una razón por la que Jack no era el mejor para buscar ayuda.

Código 0: sin comprensión.

La respuesta no proporciona una razón creíble basada en el texto por la que Jack era el mejor para buscar ayuda o una razón por la que Jack no era el mejor para buscar ayuda.

Ejemplo:

- Era el mejor porque era el más rápido.

Código 9: respuesta en blanco.

Figura 2.23. Ejemplo de ítem con respuesta abierta del estudio PIRLS 2016

Indica dos razones por las que Charlotte era diferente del resto de las ovejas.

Codificación:

Código 2: comprensión total.

La respuesta proporciona dos aspectos, basados en el texto, de la lista de respuestas correctas que hay a continuación. Los aspectos de la lista que se encuentran en la misma línea son considerados como la misma idea y no se pueden puntuar por separado.

NOTA PARA LOS CORRECTORES: se consideran correctas las respuestas que parafrasean estas ideas siempre que el significado esté claro.

Aspectos en los que Charlotte era diferente de las demás ovejas:

- Era valiente. / No era tímida.
- Iba saltando por ahí (mientras otras ovejas se quedaban al lado de sus madres). / No se quedaba al lado de su madre.
- Estaba preparada para la aventura.
- Saltó la orilla del río y cruzó a nado.
- Por las noches deambulaba en secreto por el campo.
- Se escondió en su lugar secreto (y observó la luna) (mientras las demás ovejas dormían).
- Fue a salvar al pastor. (Las acciones concretas que formaron parte del salvamento del pastor, como conducir una camioneta, también son correctas).

Código 1: comprensión parcial.

La respuesta proporciona un aspecto, basado en el texto, de la lista de respuestas correctas que hay más arriba.

Código 0: sin comprensión.

La respuesta no proporciona ningún aspecto basado en el texto en el que Charlotte era diferente de las demás ovejas.

Ejemplos:

- *No tenía miedo de Jack.* (El texto no dice que las demás ovejas tengan miedo de Jack).
- Era tímida.
- Estaba asustada.

Código 9: respuesta en blanco.

Corrección y codificación de las preguntas de respuesta abierta larga

Cuando la tarea que hay que realizar para responder a la pregunta abierta es más compleja, la codificación requiere métodos más elaborados que permitan minimizar la subjetividad del corrector. La realización correcta de estas tareas se relaciona, en general, con la consecución de varios indicadores de logro y, por lo tanto, suelen relacionarse con varios ítems.

A lo largo de los años se han desarrollado diversos métodos de codificación y puntuación que incluyen aproximaciones tanto holísticas como analíticas (Stecher *et al.*, 1997). Los principales métodos de corrección de estas preguntas incluyen las escalas de valoración, las listas de cotejo y, en su versión más completa, las rúbricas. Otra aproximación a la elaboración de la guía de codificación es la corrección comparativa con patrones.

A) Listas de cotejo, escalas de valoración y rúbricas

Una **lista de cotejo** consiste en un listado de aspectos a evaluar (criterios de evaluación, indicadores de logro, metas de aprendizaje, etc.) que deben ser demostrados en la producción de la respuesta a la pregunta abierta. Se codificará con 0 si el aspecto no es alcanzado y con 1 en caso contrario.

Una **escala de valoración** admite, para cada uno de los aspectos evaluados, una gradación del nivel de logro alcanzado, codificándose desde el 0, representando el “logro no alcanzado”, a un máximo que sería “logro completamente alcanzado”. Las escalas de valoración deben representar los criterios de evaluación o indicadores de logro utilizados para evaluar procedimientos complejos. La descripción de cada uno de los niveles de logro debe ser lo más objetiva y consistente posible.

Para lograr su objetivo, en la construcción de las escalas se tendrán en cuenta los siguientes aspectos:

- Enfocar la escala a la medida de los indicadores de logro planteados.
- Utilizar un rango de rendimiento esperado y definir criterios específicos de valoración ordenados en niveles que indiquen en qué grado se cumple la tarea planteada.
- Evitar criterios ambiguos, subjetivos o que abarquen varios aspectos que deban ser evaluados por separado. Por ejemplo, deben evitarse términos como “bien”, “correctamente”, “comete algún fallo”, etc.
- Utilizar rangos claramente diferentes uno de otro, comprensibles y descriptivos. Cada uno de ellos puede ilustrarse con ejemplos.

En general, se prefiere:

- Establecer el menor número de niveles posible.
- Establecer un número par de niveles. Si se elige número impar (por ejemplo, niveles 0, 1 y 2) los correctores tenderán a codificar con el nivel intermedio en la mayoría de las correcciones, el 1, y el ítem no discriminará correctamente.

La Figura 2.24 muestra un ejemplo de escala de valoración mal construida para la codificación de la elaboración de un resumen. El problema de la escala es que abarca dos aspectos que deben ser evaluados por separado: ortografía y calidad del resumen. Además, no están bien delimitados los niveles. Por ejemplo, no sería objetiva la corrección de un resumen con las dos ideas solicitadas pero que presente dos faltas de ortografía.

Figura 2.24. Esquema de un ítem de respuesta abierta con agrupación y niveles de desempeño

Código del nivel	Descripción
1	Comete más de tres faltas de ortografía o el resumen no incluye ninguna de las siguientes ideas: idea A e idea B.
2	Comete, como máximo, tres faltas de ortografía y realiza un resumen en el que solo se incluye una de las siguientes ideas: idea A e idea B.
3	Comete, como máximo, una falta de ortografía y realiza un resumen en el que incluye esas ideas: idea A e idea B.

Una **rúbrica de evaluación** o **tabla de codificación** se presenta generalmente en forma de cuadrícula o tabla y es una herramienta muy útil para codificar de forma objetiva las respuestas del alumnado. Mediante la rúbrica se muestra de forma explícita la tabla de especificaciones establecida en el marco de la evaluación y el desglose según los indicadores de logro o criterios de evaluación establecidos, y se codifica el desempeño de los estudiantes teniendo a estos como referencia.

Los elementos principales de una rúbrica son:

- **Indicadores de logro:** metas de aprendizaje o criterios de desempeño. Constituidos por los elementos observables que deben estar presentes en la producción del alumnado.
- **Niveles de logro o desempeño:** para establecer estos niveles debe servir de guía el marco de la evaluación. Habitualmente se define una escala ordinal para valorarlos y se etiquetan de diversos modos dependiendo de la intención de la evaluación. En general, es suficiente con tres niveles, ya que la descripción de un número mayor suele resultar compleja. Por otro lado, la utilización de un número impar de niveles suele conducir a una mayor frecuencia en la elección del que ocupa el lugar central por lo que, si es posible, se optará por definir cuatro niveles. Algunos niveles y etiquetas posibles son:

Mal – Regular – Bien –Muy bien

Nivel 1 – Nivel 2 – Nivel 3 – Nivel 4 – Nivel 5

Inicial – Intermedio – Avanzado

Aunque en la evaluación externa los niveles no suelen etiquetarse y simplemente se les asocian códigos (0, 1, 2...).

- **Descriptorios:** para cada uno de los criterios de evaluación seleccionados, debe describirse cada uno de los diferentes niveles de consecución. Esta tarea es compleja ya que requiere una definición cualitativa de los diferentes niveles de desempeño y la descomposición de los indicadores de logro en elementos observables y medibles.

Es necesario señalar que la elaboración de una rúbrica es un trabajo complejo que a lo largo del ciclo de elaboración de una prueba de evaluación requiere una revisión continua. La elaboración suele partir de un análisis detallado por un grupo de expertos y su posterior validación conjunta. Lo ideal es probar el comportamiento de una rúbrica previamente a su utilización. Esto suele hacerse, por ejemplo, con las producciones del alumnado en las pruebas piloto o en una muestra seleccionada de alumnado en lo que se suele denominar “laboratorio cognitivo”. La Figura 2.25 muestra un ejemplo de rúbrica analítica (que evalúa varios indicadores de logro) para la valoración de un restaurante.

Figura 2.25. Ejemplo de rúbrica para la valoración de un restaurante

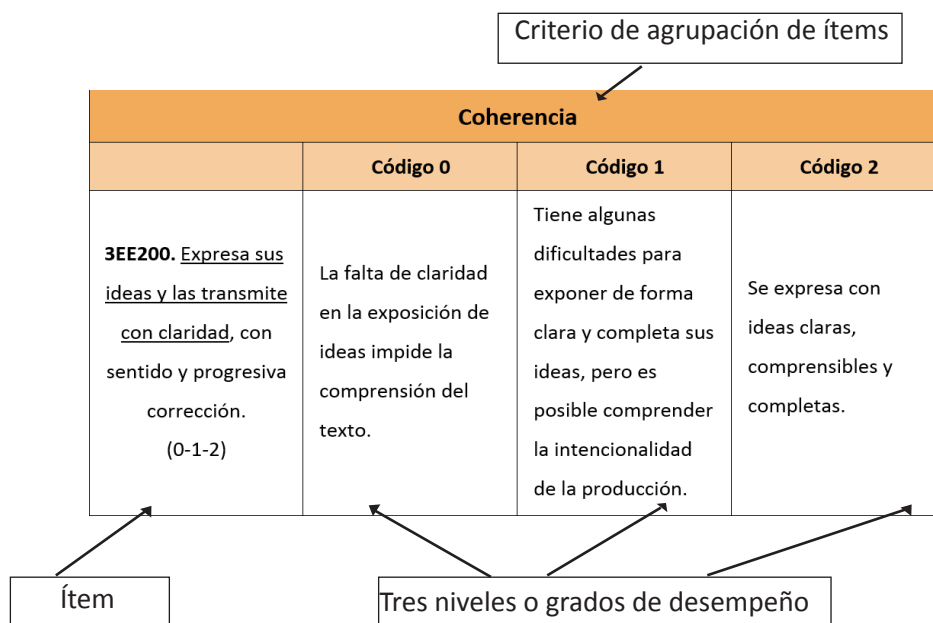
	Bajo	Medio	Alto	Muy alto
Indicador 1: calidad del producto	<i>El producto no es de calidad.</i>	<i>Calidad del producto media.</i>	<i>Calidad del producto alta.</i>	<i>Máxima calidad del producto.</i>
Indicador 2: calidad del cocinado	<i>Mala calidad del cocinado.</i>	<i>Cocinado aceptable y sabroso.</i>	<i>Buen cocinado y sabroso, pero no tiene el toque especial que se esperaba.</i>	<i>Muy bien cocinada, sabrosa y con un toque especial.</i>
Indicador 3: decoración y limpieza	<i>Decoración poco cuidada e instalaciones algo anticuadas. Baños sucios.</i>	<i>Decoración sencilla, pero adecuada. Algunos aspectos de limpieza deberían mejorarse.</i>	<i>Decoración cuidada y adecuada. Algunos aspectos de limpieza deberían mejorarse.</i>	<i>Decoración muy cuidada e instalaciones limpias.</i>
Indicador 4: servicio	<i>Servicio lento y mala atención a los clientes.</i>	<i>Servicio adecuado. Se deben mejorar algunos aspectos relacionados con la atención al cliente.</i>	<i>Servicio adecuado. Personal agradable.</i>	<i>Atención muy buena hacia el cliente.</i>
Indicador 5: relación calidad-precio	<i>El precio es muy alto en relación al servicio recibido.</i>	<i>El precio es alto en relación al servicio recibido.</i>	<i>Precio adecuado al servicio recibido.</i>	<i>Dinero muy bien invertido en relación al servicio y calidad de la comida.</i>

Las rúbricas de evaluación empleadas en las pruebas de evaluación de diagnóstico suelen evaluar cada uno de los indicadores de logro (identificados con ítems) mediante una escala de códigos enteros que varía de 0 hasta, como máximo, 3. El análisis de respuestas codificadas con escalas más amplias suele mostrar un mal funcionamiento si se introducen demasiadas preguntas con códigos diferentes al 0/1, por lo que se reducirá al máximo el número de ítems con codificación 0, 1 y 2, y se evitarán los ítems con código adicional 3. La construcción de la rúbrica requiere describir el desempeño correspondiente a cada código de nivel de forma pormenorizada, de modo que conduzca a la asignación del mismo código independientemente del codificador que realice la tarea de corrección. Cada una de las celdas de la rúbrica puede completarse con ejemplos, que habitualmente se obtienen de la prueba piloto. Una forma de establecer la gradación de los niveles de forma objetiva es la cuantificación de los errores.

Para que el alumnado sepa lo que se espera que realice en relación con la tarea que se le plantea, esta debe contener instrucciones precisas, como por ejemplo el número de palabras o de líneas que debe completar. Además, se indicará si se valorará o no la ortografía, el formato requerido de respuesta, etc. No debería valorarse en la rúbrica ningún aspecto opcional para el alumnado. La guía de codificación debe incorporar, además, instrucciones sobre el modo de proceder en caso de que no se alcancen los mínimos objetivos de la tarea. Además, como es habitual, se incluye un código para identificar las respuestas en blanco.

En la elaboración de las guías de codificación de las preguntas abiertas debemos considerar el **grado de desempeño evaluable y observable** en las respuestas de los alumnos en cada uno de los ítems que se plantean. Estos ítems deben identificarse claramente y ordenarse o agruparse con otros ítems afines (por ejemplo, en expresión escrita se pueden agrupar por cualidades textuales, como coherencia, cohesión, adecuación...). En la Figura 2.26 vemos un ejemplo de ello:

Figura 2.26. Esquema de un ítem de respuesta abierta con agrupación y niveles de desempeño



Los niveles de desempeño se pueden establecer atendiendo a distintos criterios: en el ejemplo se ha considerado la **intensidad**, pero podrían ser otros, como la **cantidad**, por ejemplo, como ocurre en la Figura 2.27, en la que se recogen los criterios para codificar en base a las faltas de ortografía cometidas:

Figura 2.27. Ejemplo de ítem de respuesta abierta con criterio de cantidad

Adecuación		
	Código 0	Código 1
3EE208. Completa y escribe textos sencillos aplicando las normas ortográficas adecuadas a su nivel (b/v, g/j, h). (0-1)	Comete más de dos fallos ortográficos en el uso de: <ul style="list-style-type: none"> - b/v, - g/j, - la h en palabras con hie- y hue-, - m antes de p y b, - r/rr, - palabras terminadas en -z y en -d - uso correcto de la c, z y el dígrafo qu, - palabras terminadas en -y, - uso correcto de las grafías ga, go, gu, gue, gui, güe, güi. 	Comete como máximo dos fallos ortográficos en el uso de: <ul style="list-style-type: none"> - b/v, - g/j, - la h en palabras con hie- y hue-, - m antes de p y b, - r/rr, - palabras terminadas en -z y en -d - uso correcto de la c, z y el dígrafo qu, - palabras terminadas en -y, - uso correcto de las grafías ga, go, gu, gue, gui, güe, güi.

También la **frecuencia** de aparición del indicador de logro (en el ejemplo que sigue, la riqueza de vocabulario) puede ser un criterio de codificación, como se ve en la Figura 2.28:

Figura 2.28. Ejemplo de ítem con respuesta abierta con criterio de frecuencia

Cohesión		
	Código 0	Código 1
<p>3EE106. Utiliza en sus expresiones un vocabulario adecuado a su edad. (0-1)</p>	<p>Repeticiones: la producción presenta un lenguaje repetitivo sin uso, en caso de necesidad, de sinónimos, pronombres o verbos de acción variados.</p>	<p>Repeticiones: la producción presenta un lenguaje variado, utilizando los recursos necesarios para evitar las repeticiones (sinónimos, pronombres...).</p>

Dependiendo del tipo de prueba y del objetivo final de la misma, se escogen los criterios más adecuados, o incluso se pueden combinar, en una propuesta más compleja, varios de ellos (por ejemplo, el número de faltas de ortografía -cantidad- y, en el mismo ítem, si las faltas de ortografía cometidas son más o menos importantes -cualidad): en todo caso, deben ser siempre criterios observables, que permitan una codificación fiable y que aporten un rasgo relevante del grado de adquisición de la competencia evaluada por parte del alumno.

La Figura 2.29 muestra un ejemplo de pregunta de respuesta abierta (producción escrita correspondiente a una evaluación de la destreza de expresión escrita en 3.º de Educación Primaria). La Figura 2.30 recoge un extracto de la tabla de codificación de esta tarea.

Figura 2.29. Ejemplo de pregunta de respuesta abierta asociada a más de un ítem


UN CUMPLEAÑOS ACCIDENTADO

Las fiestas de cumpleaños son un lugar perfecto para pasarlo genial y en ocasiones ocurren cosas inesperadas.

Te ofrecemos una historia en imágenes a la que le faltan las palabras y tú serás su escritor.

¡Debes escribir al menos 5 líneas en cada imagen!


1. Érase una vez...




¡Feliz cumpleaños!

2. _____

3. _____



¿Casa equivocada?



¡Fuegoooooo!

¡Gracias por tu trabajo!

Figura 2.30. Ejemplo de extracto de la tabla de codificación de la pregunta de la Figura 2.29

- Se codificará como 9 en todos los ítems cuando no haya ningún tipo de producción. Se codificará 0 en todos los apartados de la tabla cuando en las tres imágenes haya un texto inferior al número mínimo de líneas requerido o cuando en una de ellas se alcance el mínimo, pero en las otras dos, no.
- En caso de que en dos imágenes se alcance el número mínimo y en otra no, se procederá a codificar según la tabla.
- También se codificará 0 cuando la producción no tenga ningún tipo de relación con la tarea requerida en ninguno de los pasos.

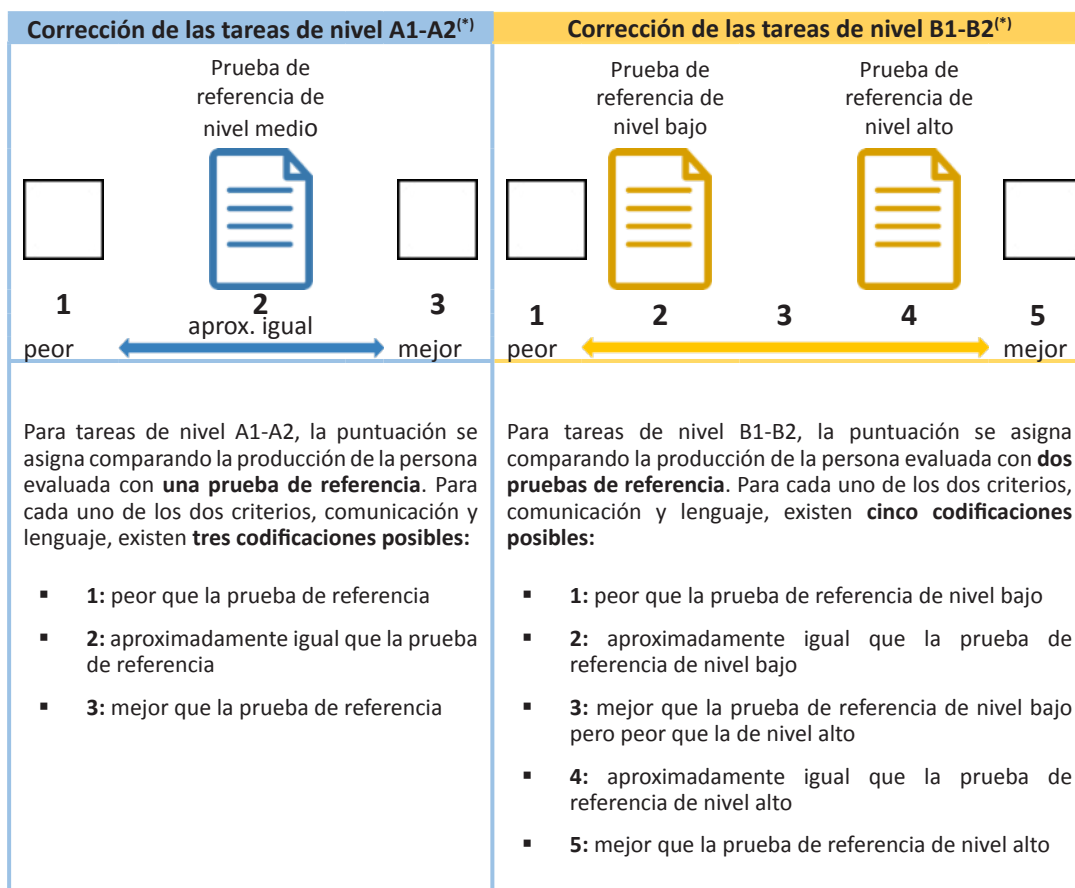
Coherencia			
Ítems asociados al ejercicio. En cada ítem se valora el aspecto subrayado del indicador de logro	Código 0	Código 1	Código 2
ÍTEM 3EEXX01. Expresa sus ideas coherentemente, respetando las opiniones de los demás. (0-1)	Relevancia de la información: aporta información que no está relacionada con la tarea, no aporta información o esta es irrelevante.	Relevancia de la información: la información se ajusta a los objetivos de la tarea.	Se expresa con ideas claras, comprensibles y completas.
ÍTEM 3EEXX02. Expresa sus ideas y las transmite con claridad, con sentido y progresiva corrección. (0-1-2)	La falta de claridad en la exposición de ideas impide la comprensión del texto.	Tiene algunas dificultades para exponer de forma clara y completa sus ideas, pero es posible comprender la intencionalidad de la producción.	
ÍTEM 3EEXX03. Expresa sus ideas y las transmite con claridad, con sentido y progresiva corrección. (0-1)	La producción no presenta progresión temática. Se observan saltos temporales y omisiones injustificadas entre la secuencia de imágenes. Ejemplo: <ul style="list-style-type: none"> • Se salta alguno de los pasos de la historia. • Repite información cuando no es pertinente. 	La producción presenta progresión temática y no se observan saltos temporales ni omisiones injustificadas entre las secuencias de imágenes.	

B) Corrección comparativa con patrones

Este método de corrección de preguntas de respuesta abierta consiste en **comparar las producciones** de las personas evaluadas **con pruebas de referencia (de distintos niveles)** y determinar si las producciones son aproximadamente iguales, mejores o peores que estas últimas en relación con ciertos criterios previamente establecidos. Este modelo de codificación fue utilizado en 2011 en el Estudio Europeo de Competencia Lingüística (EECL) (INEE, 2012) en la codificación de las pruebas de expresión escrita. En la corrección de estas pruebas se tuvieron en cuenta dos criterios: **comunicación** (que se refiere al grado de cumplimiento de la tarea con respecto a los requisitos sobre su contenido que se especifiquen en el enunciado de la misma y el estilo o registro empleados en la redacción) y **lenguaje** (que valora la adecuación a la prueba de aspectos como el vocabulario, la ortografía, la gramática y la organización lingüística). La principal ventaja que presenta este método es que neutraliza en gran medida la subjetividad de la persona correctora, ya que los juicios son comparativos y no absolutos.

La Figura 2.31 ilustra el procedimiento de corrección en el EECL. En las producciones de los niveles A1 y A2 del Marco Común Europeo de Referencia de las Lenguas (MCERL) (Consejo de Europa, 2001) se mostraba un único patrón de corrección, mientras que en las tareas de niveles B1 y B2 la comparación se hacía con dos producciones de referencia. La Figura 2.31 muestra un esquema del método de codificación, mientras que en la Figura 2.32 se desglosan los aspectos que se deben considerar en la corrección en cada uno de los criterios.

Figura 2.31. Diagrama ilustrativo del método de corrección de las producciones escritas en el EECL



^(*) Los niveles de referencia (A1, A2, B1 y B2) del EECL se corresponden con los niveles de competencia definidos en el MCERL.

Figura 2.32. Criterios del método de corrección comparativo de preguntas de respuesta abierta del EECL

Comunicación (cumplimiento de la tarea) (requisitos de contenido, estilo)	Lenguaje (organización, ortografía, gramática)
¿Hasta qué punto cumple la respuesta con los requisitos de la tarea?	¿Cómo de adecuado a la tarea es el vocabulario/organización lingüística/precisión?
<ul style="list-style-type: none"> ▪ ¿Cuántos de los puntos de contenido de la tarea se tratan con claridad? ▪ ¿Cómo de desarrollados están estos puntos de contenido? ▪ Estilo – registro 	Aspectos del lenguaje a tener en cuenta: <ul style="list-style-type: none"> ▪ Coherencia ▪ Vocabulario ▪ Cohesión ▪ Precisión

Se incluyen en las Figuras 2.33 y 2.34 dos ejemplos de tareas de respuesta abierta en inglés propuestas en el EECL, junto con sus correspondientes pruebas de referencia. Puede observarse que las pruebas de referencia contienen errores, puesto que, como se ha dicho, no se corresponden con ningún nivel concreto.

Figura 2.33. Ejemplo de tarea con respuesta abierta de nivel A1 y sus pruebas de referencia para la corrección (EECL 2011)

<p>Tarea de nivel A1: Holiday Photo</p> <p>You're on holiday. Send an email to an English friend with this photo of your holiday. Tell your friend about:</p> <ul style="list-style-type: none"> • the hotel • the weather • what the people are doing <p>Write 20-30 words.</p>
<p>Prueba de referencia para la corrección:</p> <p>Alcanza A1 (prueba de referencia de nivel medio):</p> <p>“Hi! I living in Hotel Bellevue and this is nice, We have swimming pool and a nice resturant. The weather is very good, its sunny and very hot. And the people play vollyball and they are nice. Good bye!”</p> <p>Pre-A1:</p> <p>They play voleyball. The namn of the hotel is Belleevue. Have a greates tree.</p>

Figura 2.34. Ejemplo de tarea con respuesta abierta de nivel B1 y sus pruebas de referencia para la corrección (EECL 2011)

Tarea de nivel B1: Favourite family member

This is part of an email you receive from an English pen friend:

In your next email, tell me about someone in your family that you like a lot. What sorts of things do you do together? Why do you get on well with each other?

Write an email to your friend, answering your friend's questions.

Write 80-100 words.

Pruebas de referencia para la corrección:

Alcanza B1 (prueba de referencia de nivel alto):

Hallo,

My family is great and I love it, but I love my mother the most. We always going shoping together, or do some funny different stuff. I love when we watching a scary movie. We making so much popcoen and laughing all the time. My mother is always with me, that is why I love her so much. She is the strongest person in the world. It is so funny with her. We love singing and my father goes crazy. In the winter we always go skiing and that is one of the best things in the year. I love my family, but my mother ls at the top, she is the best.

Aún en A2 (prueba de referencia de nivel bajo):

Dear John,

Thanks for your email.

In my family I like a lot Marie. It's my sister. I have 3 sisters but I'm going to talk you about Sophie. Sometimes we go shopping together and we kocht a lot of clothes. Marie is very friendly. We talk a lot together about our personnal life: about boy friends, school.. It's funny. Last week I wend in her flat in Brussel. She's a student in chemistry, The day we went shopping for find a dress for her. We finded it and she's very beautiful.

See you soon Isabelle

Elaboración de pruebas de evaluación

Una prueba de evaluación con enfoque competencial debe ser un instrumento en el que el estudiante demuestre con evidencias si puede realizar las tareas asociadas a la competencia exigida, entendida esta como combinación de conocimientos, capacidades y actitudes adecuadas al contexto. Las competencias clave son aquellas que todas las personas precisan para el desarrollo personal, así como para la ciudadanía activa, la inclusión social y el empleo, y que fomentan el desarrollo de habilidades que permitan resolver una situación real, compleja e inédita (Díaz Barriga y Hernández, 2010). Además, las tareas que se planteen deben de estar siempre contextualizadas, es decir, relacionadas con la realidad sin olvidar que los resultados del aprendizaje (criterios de evaluación, indicadores de logro, metas, etc.) deben de ser los que guíen la elaboración de las preguntas.

Para poder describir correctamente los diferentes niveles de desarrollo de la competencia, la prueba de evaluación debe contener preguntas con distintas dificultades, así como ítems de tipos variados, según se haya especificado en el marco de evaluación. En cuanto a la longitud, es importante planificar el tiempo de que dispondrá el alumnado para completar la prueba. Otro factor fundamental para elaborar las preguntas será tener en cuenta el material con el que contará el alumnado para completar la prueba, como son calculadoras, reglas, equipos de audio o vídeo, etc. La prueba debería planificarse de forma que se garantice que este material es exactamente igual para toda la población. Igualmente, se deberán contemplar en la planificación las adaptaciones de este material que fueran necesarias para los alumnos que así lo precisen.

Habitualmente las pruebas en papel se le presentan al alumnado en cuadernillos, que pueden ser diferentes según el diseño de la evaluación. En el diseño de los cuadernillos, la disposición y el orden de las preguntas es también importante: si las preguntas de menor nivel de dificultad se agrupan al principio, el alumnado con mayor competencia puede desmotivarse y no completar la prueba, igual que situar las preguntas de mayor dificultad al principio puede desmotivar al alumnado con nivel de competencia más bajo. En la actualidad, los medios digitales están abriendo múltiples posibilidades en evaluación educativa, como son la configuración de “cuadernillos digitales” distintos, la posibilidad de flexibilizar la prueba e ir adaptándose según las respuestas de cada alumno o alumna (“pruebas adaptativas”) o la posibilidad de ampliar el tipo de preguntas o el diseño interactivo de los ítems.

Ciclo de elaboración de pruebas

La elaboración de una prueba de evaluación de calidad es un trabajo complejo que lleva varias etapas. Muchas pruebas de evaluación externa de carácter competencial están constituidas por unidades de evaluación, aunque en ocasiones los ítems o preguntas aparecen de forma independiente. En todo caso, aquí asumiremos que los ítems se asocian a un estímulo formando una unidad de evaluación.

➤ Construcción de las unidades de evaluación

En esta primera fase de elaboración de las pruebas deben escogerse los estímulos a partir de los cuales van a desarrollarse las unidades de evaluación. Una vez escogidos, debe redactarse un número elevado de ítems de diferente tipología a ellos asociados que aseguren la cobertura de los diferentes niveles de desarrollo de la competencia que se quiere evaluar. Deberán establecerse asimismo los criterios de codificación para cada uno de los ítems indicando, cuando sea necesario, la graduación de sus posibles respuestas. Este trabajo debe ser llevado a cabo por personal debidamente capacitado.

➤ Revisión de calidad de los estímulos e ítems asociados

Las tareas de revisión deben recaer sobre personal cualificado diferente de quienes redactaron los ítems. La corrección de los ítems no se limita a detectar y corregir posibles defectos en la

redacción, sino que requiere comprobar que las respuestas y sus niveles de graduación sean correctos, que los ítems estén adaptados al nivel de la población a la que se destina la prueba, que sean independientes unos de otros, etc. En resumen, se trata de asegurar que la prueba cumple todos los parámetros incluidos en los apartados que se van a referir a lo largo de este apartado de la guía.

➤ **Configuración de las pruebas de evaluación**

Finalizada la revisión de ítems y estímulos debe llevarse a cabo la selección de unidades de evaluación e ítems que van a incluirse en las pruebas. Aunque, en cualquier caso, la configuración final de las pruebas debe obedecer a las matrices de especificaciones establecidas para cada una de las competencias que vayan a evaluarse, conviene recordar que deben considerarse variables como las características de la población a quien vaya dirigida la prueba o el tiempo del que se disponga para su aplicación, entre otros, a la hora de establecer el número mínimo y/o máximo de unidades e ítems que conforman una prueba de evaluación.

Una vez configuradas las pruebas, deben diseñarse y maquetarse en el formato en el que esté prevista su aplicación, siendo los dos más comunes los cuadernillos en papel y el formato digital.

➤ **Pruebas piloto**

Es aconsejable aplicar una prueba piloto, o incluso una prueba pre-piloto y una piloto, antes de llevar a cabo la aplicación del estudio principal. La finalidad de estos estudios es doble: por un lado, se analiza el funcionamiento general de los ítems, es decir, que estos cumplen su función, en una muestra de estudiantes con características similares a la población de interés. Por otro lado, el pilotaje sirve para probar cuestiones logísticas. Habitualmente en las pruebas piloto se suele probar un número de ítems superior al que finalmente configurará la prueba definitiva ya que algunos ítems son rechazados debido a su mal funcionamiento, bien debido a sus propiedades psicométricas o por otras cuestiones como el lenguaje empleado o el contexto al que se refieren.

A partir de los datos arrojados por una prueba piloto los ítems pueden revisarse, tanto mediante un análisis cualitativo como cuantitativo. El análisis cualitativo comprende la revisión de los errores de redacción, léxicos o gramaticales, la adaptación al contexto, la comprensibilidad, el formato de la maqueta, etc. Además, se revisan cuestiones como la redacción de las instrucciones generales de la prueba. En cuanto a la revisión cuantitativa, se realiza una calibración de los ítems: se valida su capacidad de discriminación, se mide su grado de dificultad, etc. Teniendo estos resultados en cuenta se seleccionan finalmente los ítems que configurarán la prueba definitiva eligiendo aquellos con las mejores propiedades como instrumentos de medida y descartando o modificando otros. También se obtiene información que permite completar y mejorar las guías de codificación de las respuestas abiertas.

Los resultados obtenidos en las pruebas pre-piloto y piloto también pueden servir para ajustar y poner a punto las dimensiones de la competencia y sus relaciones con los elementos del currículo que se esté evaluando. Puede realizarse, en su caso, una primera elaboración de escalas de rendimiento (aplicación de la teoría de respuesta al ítem) para asociar cada ítem a un nivel de dificultad.

Por regla general, el proceso de pilotaje se lleva a cabo con participantes para los que no debería ser difícil responder a las distintas preguntas, de modo que lo que se analiza es la validez y viabilidad de los ítems, así como su capacidad discriminatoria. Por ejemplo, si diseñamos una evaluación para el alumnado de 3.º EP, podríamos pilotarlo con alumnos de 4.º EP en los primeros meses del curso escolar. En los estudios internacionales de evaluación, la prueba piloto suele realizarse en el curso escolar anterior a la aplicación principal.

Por último, se configura una nueva versión de la prueba de evaluación que, una vez revisada por los expertos, será la que finalmente se aplique en el estudio principal.

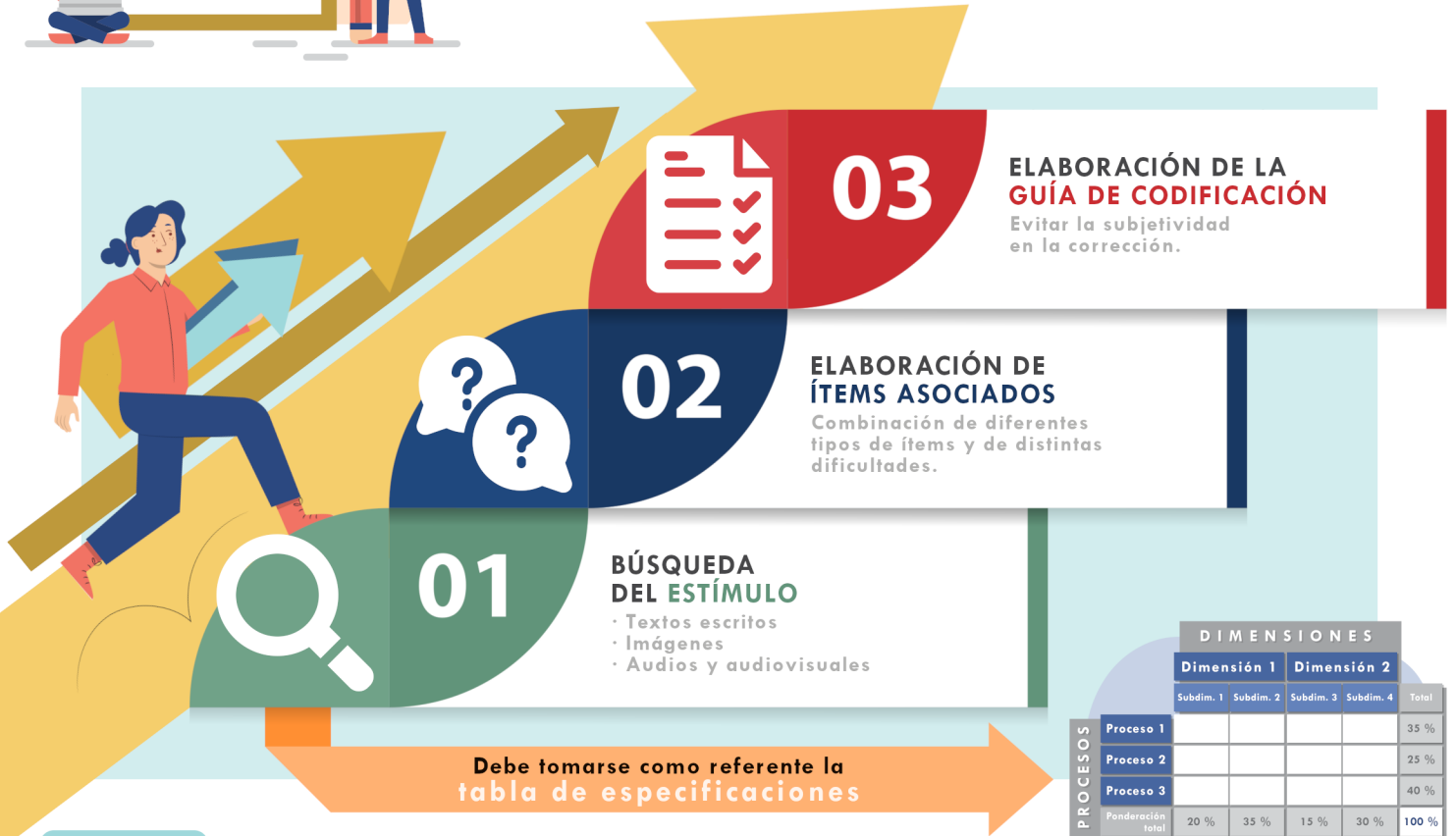
➤ Estudio principal

El proceso de aplicación debería constar, en primer lugar, de una fase preparatoria. Durante esta primera fase debería sensibilizarse al grupo al que se dirige la prueba (grupo evaluado, responsables de dicho grupo, aplicadores de las pruebas, etc.) sobre el sentido y la naturaleza de la evaluación que se va a llevar a cabo y, si fuera necesario, hacer hincapié en la formación de los aplicadores en lo que respecta a las instrucciones generales de aplicación e instrucciones específicas de las pruebas.

La segunda fase de aplicación sería la de ejecución de las pruebas, durante la que debería insistirse al grupo evaluado sobre la importancia que tiene realizar la prueba con la máxima concentración posible. Para garantizar la objetividad de la prueba, lo ideal sería que su aplicación y corrección corriesen a cargo de aplicadores externos al contexto en el que se desarrolle.



Elaboración de pruebas de evaluación



¿Cómo plantear preguntas para la evaluación por competencias?

RETOS

Se deben plantear cuestiones que planteen problemas nuevos, que supongan un desafío para el alumnado, de forma que su respuesta requiera la resolución de un problema real, más o menos complejo y relevante.

CONTEXTUALIZADAS

Se plantean en situaciones que tienen que ver con la realidad y requieren analizarla e indicar cómo actuar de forma razonada.

EVALUAR SABER, SABER HACER Y SABER SER

Se deben evitar únicamente cuestiones memorísticas o que se reduzcan a reproducir el aprendizaje adquirido. Estos deben tratarse como componentes de la adquisición de la competencia: conocimientos (saber), destrezas (saber hacer) y actitudes (saber ser).

COHERENTES CON LOS APRENDIZAJES DE CLASE

Se deben adecuar al nivel educativo al que se dirigen. Además, debe tenerse en cuenta que para realizar una evaluación por competencias la metodología de enseñanza-aprendizaje debe basarse también en competencias.

Recomendaciones generales para la elaboración de pruebas

➤ Formato de la prueba

Es importante señalar que, desde un punto de vista formal, las pruebas deben facilitar su uso por parte de sus destinatarios y destinatarias y resultar atractivas y motivadoras, sean en papel o en formato digital.

Independientemente del formato, la maquetación de las pruebas debe siempre tener en cuenta una serie de aspectos formales que han de contribuir a cubrir los dos criterios mencionados:

- Tipo y tamaño de letra
- Color empleado en dibujos y/o gráficos

En el caso de las pruebas en papel, además, debe cuidarse la amplitud de márgenes y los espacios para la respuesta. En las pruebas digitales es conveniente posibilitar la visualización simultánea de estímulo e ítem asociado y diseñar una plataforma con la máxima usabilidad posible y adaptada a la edad de los destinatarios.

➤ Características generales

En general, las unidades que conforman una prueba de evaluación deben tener una serie de características en común:

- **Fiabilidad.** La fiabilidad se define como el grado de consistencia con el que una prueba evalúa el resultado de un aprendizaje. Una prueba debe ser capaz de asegurar resultados similares si se repite en circunstancias similares.
- **Validez.** La prueba debe ser precisa en la medición de aquello que pretenda evaluar.
- **Objetividad.** La prueba debe ser independiente del grupo a evaluar y del corrector o correctora.
- **Valor discriminatorio de los ítems.** Se refiere a la capacidad de cada ítem para diferenciar entre sujetos que presentan diferentes niveles de competencia, esto es, los individuos con nivel de competencia alta deben tener mayor probabilidad de responder correctamente los ítems de mayor dificultad. Asimismo, los individuos con menor nivel de competencia deben tener menor probabilidad de responder acertadamente los ítems de dificultad alta.
- **Dificultad de los ítems.** Los ítems no tienen que ser ni muy fáciles, ya que los contestarían bien todas las personas evaluadas, ni muy difíciles, ya que nadie sabría responder. En definitiva, la dificultad de los ítems debe de estar adaptada al grupo evaluado, aunque para poder describir el nivel de competencia de forma adecuada deben diseñarse ítems de todos los niveles de dificultad.
- **Eficiencia.** La prueba debe ser fácil de llevar a la práctica.
- **Adecuación a la competencia.** Las unidades deben estar adecuadas tanto a la competencia objeto de evaluación como al nivel de las personas a las que se dirija.
- **Contextualización.** En la medida de lo posible, y como ya se ha señalado anteriormente, los estímulos en los que se basan las unidades deben remitir a situaciones cotidianas o cercanas a quien haga la prueba. En las etapas obligatorias es habitual que los estímulos se refieran a situaciones personales, escolares, sociales y científicas o humanísticas.

- **Lenguaje apropiado.** El lenguaje empleado en las pruebas debe adaptarse a la edad y/o nivel educativo de las personas que van a ser evaluadas, tanto más cuanto más joven sea el grupo al que se dirige la prueba. Una sintaxis compleja o un vocabulario inusual podrían introducir una dificultad añadida en las pruebas.
- **Ambiente cultural apropiado.** Debe tenerse en cuenta el entorno cultural del receptor de la prueba, para evitar que se produzcan situaciones de extrañeza, rechazo o desconocimiento que perjudiquen a la realización de la misma. Es importante considerar los principios de no discriminación y educación entre iguales cuando se lleva a cabo la elección del estímulo o estímulos que configuran una prueba.
- **Diversidad de estímulos.** Pueden utilizarse las redes sociales, páginas web, correos electrónicos, etc., como recurso para escoger o elaborar estímulos ricos y creadores de sentido. Otras posibilidades a las que puede recurrirse para asegurar la diversidad de estímulos son: folletos informativos, textos o anuncios de periódicos, fotografías de situaciones reales, esquemas, gráficos, planos de edificios, cómics, cartelería, instrucciones de juegos, etiquetas de productos, etc.
- **Diversidad de preguntas.** Al igual que los estímulos, las preguntas que conforman una unidad de evaluación también deben ser diversas. La tipología y el número de preguntas que conforman una prueba de evaluación debería variar dependiendo de cuál sea la finalidad de la evaluación que se esté llevando a cabo y del tiempo que vaya a dedicarse a la realización de la prueba.

➤ **Elaboración de ítems**

El proceso de redacción de ítems debe apoyarse en el análisis de las competencias objeto de evaluación y tener en cuenta los criterios de gradación en niveles de la competencia y sus indicadores de logro, así como lo recogido en las correspondientes tablas de especificaciones. Los principios básicos que deben regir la construcción de cualquier banco de ítems son: representatividad, relevancia, diversidad, claridad, sencillez y comprensibilidad (Muñiz *et al.*, 2005):

- Cada ítem debe referirse a una única subdimensión de las descritas en el marco de la evaluación (tema, proceso cognitivo, contexto, etc.). Salvo en el caso de las preguntas de respuesta abierta que, como se ha indicado anteriormente, pueden llevar asociadas varios ítems, por regla general cada pregunta se asocia a un ítem y, por tanto, mide un solo resultado del aprendizaje (indicador de logro, criterio de evaluación, meta...).
- Las personas encargadas de construir los ítems de una prueba de evaluación deben, por tanto, conocer en profundidad los contenidos de la competencia a evaluar y los aspectos concretos de los procesos cognitivos que se evalúan, así como el marco de evaluación de la prueba.
- Los ítems deben indagar sobre cuestiones no triviales.
- En pruebas con ítems asociados a un estímulo, debe tratarse de ligar los ítems al estímulo para lograr un contexto tan real como sea posible.
- Los ítems deben de ser independientes entre sí de forma que la información contenida en uno de ellos no conteste, oriente o contradiga la contestación de otro ítem, lo que reduciría la validez de la prueba.
- La dificultad del ítem debe radicar en el contenido de la pregunta no en otros elementos accesorios o en la redacción de la pregunta.

- Se debe tratar de ser creativo e innovador para medir los procesos cognitivos superiores.
- No deben preguntarse opiniones, puntos de vista personales, etc., excepto en el caso de ítems de respuesta abierta que así lo demanden.

El tronco o vástago

- Debe tener sentido por sí mismo y presentar un problema definido, ya que esto permite que la evaluación se centre en el resultado del aprendizaje en lugar de medir la capacidad de las personas evaluadas para hacer inferencias o sacar conclusiones a partir de un enunciado poco preciso. El enunciado debe, por tanto, estar redactado de tal forma que conduzca a una reflexión de la que se deduzca la opción de respuesta correcta.
- Debe proporcionar solo información precisa para responder. Si el tronco contiene información irrelevante puede reducir su fiabilidad y la validez de su resultado (Haladyna y Downing, 1989).
- Deben emplearse formas positivas en la redacción de los textos siempre que sea posible. Los enunciados negativos suelen plantear problemas de comprensión a las personas evaluadas (Rodríguez, 1997). En general, se deben evitar las negaciones y el uso de palabras como “no” o “excepto”. En caso de incluirlas, deben resaltarse con mayúscula o negrita (Figura 2.35) y es preferible enunciarlas de forma afirmativa.

Figura 2.35. Ejemplo de ítem con negación en el tronco y de alternativa mejorada

Ejemplo de negación en el tronco:	Alternativa mejorada:
<p>¿Cuál de las siguientes variables NO es cualitativa?</p> <p>A. La edad</p> <p>B. El color de pelo</p> <p>C. La ciudad donde vives</p> <p>D. El nombre del colegio al que vas</p>	<p>Las siguientes variables son cualitativas EXCEPTO...</p> <p>A. La edad</p> <p>B. El color de pelo</p> <p>C. La ciudad donde vives</p> <p>D. El nombre del colegio al que vas</p>

- Debe formularse en forma de pregunta o, en todo caso, de oración incompleta. Los troncos en forma de pregunta permiten a la persona evaluada centrarse en contestar a la pregunta en lugar de retener la oración incompleta en la memoria operativa para completarla con cada alternativa (Statman, 1988). Deben evitarse enunciados que empiecen con o contengan un espacio en blanco, ya que estas construcciones incrementan la carga cognitiva de la persona que responde.

Las opciones (clave y distractores). Ítems de opción múltiple

En los ítems de opción múltiple se plantean varias opciones de las cuales solo una de ellas, que llamamos “clave”, es correcta, denominándose el resto de las opciones “distractores”. La redacción de la opción correcta no suele llevar ninguna dificultad asociada, sin embargo, existen múltiples referencias sobre el modo de redactar correctamente los distractores (Brame, 2013; García González, 2017; Gierl *et al.*, 2017). Algunas recomendaciones para redactar las opciones son las siguientes:

- La clave debe ser evidentemente correcta.
- Se debe distribuir la posición de la clave al azar, evitando usar siempre la misma posición.
- Las opciones deben concordar sintácticamente con el enunciado de la pregunta evitando que el conjunto resulte ambiguo o que pueda provocar una interpretación errónea o dudosa del ítem.
- Las opciones deben ser mutuamente excluyentes.
- La redacción de las opciones debe ser clara y concisa. Si su redacción es demasiado extensa, el ítem evaluará en mayor medida la capacidad lectora que el cumplimiento del indicador de logro asignado al ítem en cuestión. Deben evitarse tecnicismos en los distractores. Como propuesta, se pueden emplear frases familiares incoherentes o bien afirmaciones correctas que no responden al enunciado.
- Debe evitarse la redacción de los distractores con frases negativas, especialmente si el enunciado está también formulado negativamente.
- No es aconsejable emplear el humor.
- Si una palabra o frase se repite en todas las opciones, se debe valorar su eliminación o llevarla al tronco (Figura 2.36).

Figura 2.36. Ejemplo de ítem con palabras repetidas en las opciones de respuesta y su alternativa mejorada

Ejemplo de palabras repetidas en las opciones:	Alternativa mejorada:
¿Qué longitud recorrió Juan? A. 20 metros B. 30 metros C. 40 metros D. 50 metros	¿Qué longitud, en metros, recorrió Juan? A. 20 B. 30 C. 40 D. 50

- Todos los distractores deben ser plausibles. Si no lo son, la pregunta tendrá baja discriminación y resultará demasiado sencilla. Por ejemplo, la alternativa B de la Figura 2.37 es claramente no plausible. Para diseñar distractores adecuados estos pueden basarse en errores comunes e ideas equivocadas del alumnado en relación con la tarea que se propone.

Figura 2.37. Ejemplo de ítem con distractores no plausibles

¿Para qué utilizará Paula el bote de crema solar?

- A. Para hidratarse.
- B. Para lavarse el pelo.
- C. Para lavarse los dientes.
- D. Para protegerse de los rayos del sol.

- No se deben utilizar opciones del tipo “todas las anteriores” y “ninguna de las anteriores”. Estas opciones conducen a una falta de cohesión interna del ítem y hacen que sea más fácil elegir la respuesta correcta teniendo un conocimiento parcial: cuando se introduce “todas las anteriores” como alternativa, aquellas personas que identifiquen con certeza más de una alternativa como correcta pueden responder bien a la pregunta sin estar seguros de que lo sean el resto de las alternativas.
- Todas las opciones deben ser homogéneas, esto es, de similar longitud, complejidad y estructura sintáctica. La respuesta correcta no debe ser la más larga, ni la más corta, ni la única distinta. A este respecto, resulta adecuado seguir siempre el mismo criterio de ordenación de las opciones, por ejemplo, de menor a mayor longitud o, en el caso de números, de menor a mayor.

El primer ítem de la Figura 2.38 muestra tres opciones con literatos españoles y otro francés, lo que puede inducir a elegir incorrectamente la respuesta correspondiente al francés. Por su parte, el segundo ítem muestra una opción con una redacción mucho más larga.

Figura 2.38. Ejemplo de ítems con opciones no homogéneas

¿Cuál de los siguientes escritores **NO** pertenece al Barroco?

- A. Molière
- B. Pérez Galdós
- C. Lope de Vega
- D. Luis de Góngora

En el panel que ha leído Lucía se hace referencia al plancton. El plancton está compuesto, entre otros, por protozoos y algas, formados por una sola célula eucariota. ¿En qué categoría los clasificarías?

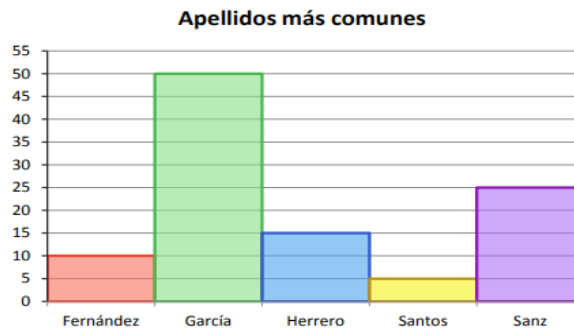
- A. Virus
- B. Hongos
- C. Bacterias
- D. Organismos unicelulares complejos

- Las opciones de respuesta no deben suponer diferentes reflexiones o razonamientos. Es el enunciado del ítem el único que debe provocar la reflexión por parte de la persona evaluada. Cuando existan preguntas en las que cada opción requiere una reflexión distinta conviene convertirlas en el formato de pregunta cerrada con opciones dicotómicas. La Figura 2.39 muestra un ejemplo.

Figura 2.39. Ejemplo de ítem reformulado como ítem de respuesta cerrada con opciones dicotómicas

Formulación como ítem de opción múltiple. Cada alternativa requiere de una reflexión

El siguiente gráfico muestra los cinco apellidos más comunes de la ciudad de Ana y Jorge, junto con el número de personas que tienen ese apellido por cada mil personas.



¿Cuál de las siguientes afirmaciones es correcta?

- A. El apellido menos común es García.
- B. Herrero es el cuarto apellido más común.
- C. Si sumamos los que se apellidan Fernández con los que se apellidan Herrero igualan a los que se apellidan Sanz.

Alternativa mejorada como pregunta de respuesta cerrada con opciones dicotómicas:

Marca con una X si las siguientes afirmaciones son verdaderas o falsas:

	Sí	No
El apellido menos común es García .		
Herrero es el cuarto apellido más común.		
Si sumamos los que se apellidan Fernández con los que se apellidan Herrero igualan a los que se apellidan Sanz .		

Aunque no hay consenso en cuanto al número óptimo de distractores, es evidente que cuantos más aparezcan más se ven afectadas sus propiedades psicométricas (dificultad, discriminación, fiabilidad, etc.) y más se reduce la probabilidad de acierto al azar. Sin embargo, un aumento de los mismos puede hacerlos menos plausibles y fáciles de descartar por el alumnado. Se ha probado que en general los ítems tienen mejores propiedades psicométricas cuando presentan dos o tres distractores (tres o cuatro alternativas de respuesta). Estas mejores propiedades psicométricas, unidas a la dificultad de redacción de más de tres alternativas plausibles incorrectas, hace que en la actualidad lo habitual sea redactar dos o tres distractores. En todo caso, todos los ítems de opción múltiple contenidos en una prueba de evaluación suelen contener todos ellos el mismo número de alternativas.

La redacción de distractores de calidad es una tarea compleja. Lo más conveniente es redactar tantos como sea posible atendiendo a los errores y aciertos previsibles que el alumnado pudiera cometer para descartar aquellos que resulten menos plausibles en un cribado posterior. Es de gran importancia considerar, además, el orden en el que se presentan las distintas alternativas, ya que se ha mostrado que la colocación de la respuesta correcta en último lugar suele aumentar la dificultad del ítem; sin embargo, si la clave se sitúa en las primeras posiciones, el alumnado tenderá a no leer todas las opciones. Por ello, se preferirá la distribución aleatoria de las claves en las distintas preguntas que configuran la prueba, lo que puede lograrse mediante la ordenación de las alternativas con algún criterio lógico: orden alfabético de clave y distractores en cada pregunta, orden numérico, longitud, etc.

➤ **Ítems de respuesta cerrada de opción múltiple compleja**

Como se describió anteriormente, en las preguntas de opción múltiple compleja que presentan varias opciones dicotómicas en las que la persona evaluada debe responder eligiendo entre dos alternativas (verdadero/falso, sí/no, causa/consecuencia, etc.), el ítem en su conjunto se orienta a la medida de un único resultado de aprendizaje. Por lo tanto, en la redacción de este tipo de preguntas se deben tener en cuenta las siguientes pautas:

- Las afirmaciones deben de ser independientes entre sí, de forma que la respuesta a una de ellas no dé pistas sobre las otras. Si esto no fuera posible, debe plantearse la posibilidad de redactar estas preguntas como opción múltiple.
- No conviene presentar más de cuatro o cinco afirmaciones. Si se plantean más, se tiende a la respuesta al azar, a la falta de discriminación de la pregunta o al aumento de dificultad de la pregunta, lo que conduce a su descarte tras la realización del estudio piloto.
- Puede concederse acierto parcial si la persona evaluada responde correctamente a todas las opciones salvo a una o alguna de ellas. Debe, no obstante, valorarse la posibilidad de acierto al azar: con cualquier número de afirmaciones la probabilidad de acertar por azar al menos a la mitad de ellas es, como mínimo, del 50 %.
- Como se indicó en el apartado anterior dedicado a los distractores, es aconsejable reformular con este formato las preguntas de opción múltiple cuyas alternativas exigen una reflexión individual.

➤ **Ítems de respuesta cerrada de emparejamiento o relación**

Anteriormente se indicó que en estas preguntas la persona evaluada debe emparejar mediante flechas, números, letras, etc., los elementos presentados en dos o más conjuntos de información según la relación existente entre ellos, o rellenar huecos siempre que se proporcione un listado de términos o elementos entre los que haya que seleccionar. Por lo tanto, en la redacción de este tipo de preguntas, deben tenerse en cuenta las siguientes pautas:

- Las directrices para contestar a la pregunta deben ser específicas, indicando claramente qué elementos deben relacionarse (los elementos de una columna con los de otra, palabras e imágenes, nombres y adjetivos...), cómo hacerlo (por ejemplo, mediante flechas) y si cada elemento puede o no ser emparejado más de una vez (Kelly, 2019).
- Debe procurarse que el formato de respuesta sea lo más claro posible, de forma que el codificador no tenga ninguna duda sobre lo que la persona evaluada ha querido indicar. Por ejemplo, el formato de “unir con flechas” puede resultar dudoso en la codificación.
- En caso de que la pregunta plantee la relación entre una lista de premisas y otra de respuestas, las respuestas deben ser más cortas que las premisas y estar organizadas alfabética o numéricamente (Kelly, 2019).

- En caso de utilizar plataformas de evaluación digital, deben tenerse en cuenta las funcionalidades que estas ofrecen, ya que este tipo de preguntas puede necesitar una modificación de formato al pasar de soporte físico (papel) a digital. En ocasiones puede hacerse necesario cambiar la redacción original e incluso la guía de codificación.
- En la guía de codificación de este tipo de ítems debe tenerse en cuenta que si el número de elementos que deben relacionarse es el mismo en los dos listados y la correspondencia es biunívoca, el error en una de las respuestas llevará necesariamente asociado otro. Esto se puede evitar, por ejemplo, incorporando un número diferente de elementos en uno de los conjuntos (por ejemplo, en los ejercicios consistentes en rellenar huecos con palabras se puede proporcionar un listado de palabras superior al número de huecos).

Figura 2.40. Ejemplo de ítem con respuesta cerrada de relación

De camino al pueblo, desde el autobús, se han encontrado con varias señales de tráfico con formas geométricas. **Relaciónalas con su imagen mediante flechas:**







	•	•	Círculo
	•	•	Triángulo
	•	•	Rectángulo

Figura 2.41. Posible adaptación del ítem de la Figura 2.40 a formato digital si la plataforma no permite la posibilidad de unir con flechas

De camino al pueblo, desde el autobús, se han encontrado con varias señales de tráfico con formas geométricas. **Marca con una X la forma geométrica que se corresponde con cada señal:**


	<input type="checkbox"/> Círculo	<input type="checkbox"/> Triángulo	<input type="checkbox"/> Rectángulo
	<input type="checkbox"/> Círculo	<input type="checkbox"/> Triángulo	<input type="checkbox"/> Rectángulo
	<input type="checkbox"/> Círculo	<input type="checkbox"/> Triángulo	<input type="checkbox"/> Rectángulo

➤ Ítems de respuesta cerrada de rellenar huecos

- Si se trata de una pregunta de completar espacios, debe evitarse introducir demasiados en un mismo elemento (por ejemplo, en una oración). Es preferible tener solo uno o dos espacios por elemento u oración (Centre for Teaching Excellence, University of Waterloo, s.f.). En todo caso, hay que asegurarse de que cada palabra se asigne a un hueco de forma inequívoca. La Figura 2.42 muestra una pregunta en la que “Después” y “A continuación” son intercambiables, por lo que debería evitarse este tipo de propuestas.

Figura 2.42. Ejemplo de ítem con respuesta cerrada de rellenar huecos

24. Ahora te daremos parte de la receta desordenada. Complétala con las siguientes palabras para que quede ordenada y con sentido:

Finalmente – Después – Primero - A continuación	
	
_____	unta el molde con la mantequilla...
_____	añade la harina junto con la levadura...
_____	mételo en el horno 30-40 minutos...
_____	añade los huevos enteros, el azúcar, la leche, el aceite y la esencia de vainilla ...

➤ Ítems de respuesta semiconstruida

Este tipo de preguntas exigen la elaboración de una respuesta por parte de la persona evaluada, bien rellenando una o varias casillas o espacios con una palabra o cifra, o bien completando frases, o respondiendo a alguna cuestión con una respuesta corta. En su elaboración se recomiendan las siguientes pautas:

- Debe indicarse claramente la respuesta que se espera. Si se trata, por ejemplo, de una pregunta de matemáticas cuya respuesta es un número decimal, debe indicarse el número de cifras decimales que se debe incluir en su respuesta (Kelly, 2018). La guía de codificación no debe penalizar por respuestas correctas que no se expresen en un determinado formato si esto no ha sido indicado en el enunciado. Por otro lado, no debe proponerse que se proporcione información adicional u optativa a la respuesta que va a ser considerada como correcta: es el caso de preguntas del tipo “cita al menos tres”, que deben evitarse; si el ítem se va a codificar como correcto con tres respuestas válidas no es necesario solicitar más. La pregunta de la Figura 2.43 solo valora citar dos elementos en cada casilla; sin embargo, el enunciado propone este número como mínimo. Puesto que la guía de codificación muestra que no tiene ninguna repercusión citar más de dos elementos, el enunciado debe modificarse.

Figura 2.43. Ejemplo de ítem con respuesta semiconstruida que solicita información “extra” y alternativa mejorada

Con la información del Documento 1 y tus conocimientos, completa la siguiente tabla incluyendo un mínimo de **dos elementos** en las casillas vacías:

Fases de la Revolución Industrial	Cronología	Motores y fuentes de energía descubiertos en el periodo	Medios de transporte (tierra, mar y aire)
Medios de transporte (tierra, mar y aire)	Hasta 1800 aproximadamente	- Muscular (animal y humana) - Molinos de viento y de agua	- Carro tirado por animales - Barco de vela
Primera Revolución Industrial	1800 – 1860		
Segunda Revolución Industrial	1860 – 1945		- Tren eléctrico y de gasoil - Automóviles - Barco de gasoil - Aviones
Tercera Revolución Industrial	1945 - Hoy		

Guía de codificación

Código 2: completa las 5 casillas con un mínimo de 7 elementos correctos.

Código 1: completa 4 casillas con al menos 4 elementos correctos.

Código 0: cualquier otra respuesta.

Código 9: respuesta en blanco.

Alternativa de enunciado mejorado

Con la información del Documento 1 y tus conocimientos, completa la siguiente tabla incluyendo **dos elementos** en cada casilla vacía:

Fases de la Revolución Industrial	Cronología	Motores y fuentes de energía descubiertos en el periodo	Medios de transporte (tierra, mar y aire)

Guía de codificación modificada

Código 2: completa todas las casillas escribiendo al menos 7 elementos correctos en total.

Código 1: escribe al menos 5 elementos correctos en total.

Código 0: cualquier otra respuesta.

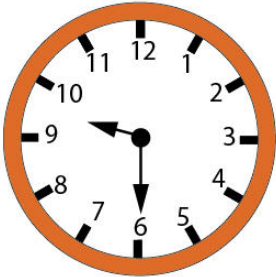
Código 9: respuesta en blanco.

- Debe evitarse ajustar el tamaño del espacio en blanco a la longitud aproximada de la que sería la respuesta correcta, ya que esto puede servir como pista para la persona evaluada (Kelly, 2018).
- Se puede orientar la respuesta con un ejemplo.
- Si se deben responder dos cuestiones es preferible indicarlo en el formato de respuesta, sobre todo cuando la evaluación se dirige a grupos de corta edad, como se muestra en la Figura 2.44.

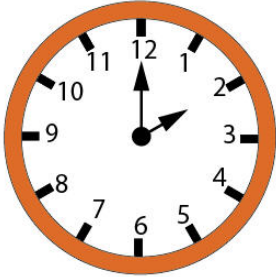
Figura 2.44. Ejemplo de ítem con respuesta a varias cuestiones y alternativa mejorada

Ejemplo de ítem con respuesta a varias cuestiones:

Cuando Hugo entró al parque por la mañana el reloj marcaba la siguiente hora:



Cuando salió a mediodía el reloj marcaba:



¿Ha estado Hugo más de 5 horas en la visita?, ¿por qué?

✎

Alternativa de planteamiento de la pregunta mejorada:

✎ ¿Ha estado Hugo más de 5 horas en la visita? *Responde sí o no.*

¿Por qué? _____

➤ Ítems de respuesta abierta o construida

Como se indicó con anterioridad las preguntas de respuesta abierta o construida son aquellas que exigen construcción por parte de la persona evaluada y, por tanto, necesitan de un corrector o correctora que revise el procedimiento llevado a cabo para alcanzar la respuesta, compare este proceso con los ejemplos recogidos en la guía de codificación y decida qué puntuación le corresponde. A la hora de elaborar este tipo de ítem se deben seguir las siguientes recomendaciones:

- El espacio que se deja para la respuesta debe ser acorde a la respuesta que se espera, que debe figurar en la guía de codificación. La Figura 2.45 muestra una pregunta en la respuesta correcta que propone la guía de codificación no se adecúa al espacio reservado para contestar en la pregunta.

Figura 2.45. Ejemplo de maquetación de un ítem de respuesta abierta para formato papel

La imagen anterior, **documento 9**, presenta a Lenin pronunciando un discurso a las masas. A su derecha aparecen los líderes Kamenev y Trotsky. Comenta brevemente el papel que jugó Lenin en la Revolución rusa y el problema sucesorio surgido tras la muerte de éste en 1924.

A) Su papel en la revolución:

B) Problema sucesorio:

En esta pregunta los alumnos/as deben contestar a las siguientes cuestiones:

A) El papel de Lenin durante la Revolución Rusa:

El papel de Vladimir Ilich Uliánov, Lenin (1870-1924) en la Revolución rusa es crucial. Fue **líder de los bolcheviques**. Vivió varias temporadas en el exilio donde aprovechó para escribir y dar forma a su pensamiento político inspirado en el marxismo. En **Las tesis de abril** recoge parte de su programa político: entre otras, salida de Rusia de la Primera Guerra Mundial, distribución de tierras entre el campesinado, control de las fábricas por parte de comités obreros, dar poder a los soviets, autonomía de las nacionalidades que formaban Rusia y nacionalización de la banca. Participa en la **Revolución de Octubre de 1917**. Si bien en 1918 la Revolución ha triunfado, existen importantes sectores contrarios a la Revolución (zaristas, terratenientes, mandos militares de alta graduación, Iglesia, etc.) que se organizan y crean el ejército blanco. Entre 1918 y 1921 tiene lugar una guerra civil ganada finalmente por los bolcheviques. En ese periodo, Lenin decreta un sistema de "Comunismo de Guerra"; nacionalizando la industria y colectivizando tierras. En 1921 establece la Nueva Política Económica cuyo objetivo será reconstruir la maltrecha economía.

En definitiva, en esta etapa se sientan las **bases del nuevo estado**. Lenin rompe con la estructura de estado anterior e instaura una dictadura del proletariado con un sistema político de partido único, el Partido Comunista o PCUS.

B) El problema sucesorio:

Al fallecer Lenin en 1924 queda la dirección del partido y de la URSS en manos de un reducido grupo de dirigentes entre los que estaban Stalin, Trotsky, Kamenev, Zinoviev, etc. A pesar de la existencia de un partido único, las discrepancias no se hicieron esperar. En este punto el alumno/a debería centrarse en el paulatino ascenso de Stalin al poder y en cómo lo va consiguiendo. Desde 1922 Stalin era el Secretario general del PCUS, defendía el planteamiento de "socialismo en un solo país" (contrario por tanto a la idea de revolución mundial). En 1929 Stalin ya controla los poderes del estado. Actuaciones como el culto a su personalidad, reforzar el poder del PCUS e instaurar una política de terror fueron algunos de los instrumentos empleados por él. En ese sentido, el alumno/a debería mencionar las purgas y depuraciones a cargos de PCUS (juicios o procesos de Moscú), el envío a campos de concentración siberianos (gulags) de disidentes, el control absoluto de la cultura, etc. Entre los líderes purgados están los dos que aparece en la fotografía, Trotsky (partidario de exportar la revolución a otros países), asesinado en su exilio de México en 1940, y Kamenev, acusado de complot contra Stalin y fusilado en el año 1935.

Como aclaración, se debe observar que el alumno/a debe contestar a las dos preguntas en un espacio reducido de unas tres líneas por cada apartado por lo que hay que valorar la claridad, coherencia y capacidad de síntesis de la respuesta.

Para lograr una valoración positiva bastará con que mencione los aspectos señalados en negrita.

Código 2: contesta de manera clara y coherente a los apartados A y B de la pregunta.

Código 1: contesta parcialmente a las dos preguntas o hace una bien y otra mal o de manera inexacta.

Código 0: cualquier otra respuesta.

Código 9: respuesta en blanco.

- En el enunciado de la pregunta deben indicarse claramente los aspectos de la respuesta que se van a valorar en la corrección. Además, puede especificarse el número mínimo o máximo de oraciones, líneas o palabras de los que debe constar la respuesta para que se pueda considerar válida (Figura 2.46 y Figura 2.47.)
- Deben evitarse términos que den lugar a interpretaciones subjetivas en la definición de la respuesta correcta. Es preferible evitar expresiones del tipo “responde correctamente”, “responde con coherencia”, “cualquier otro razonamiento correcto”, etc. Si la guía no especificara cómo deben interpretarse estas expresiones, estas deben quedar completamente clarificadas en las reuniones de coordinación de los equipos de correctores buscando siempre la objetividad en la corrección.

Figura 2.46. Ejemplo de instrucciones para una prueba de expresión escrita de 3.º de Educación Primaria

Instrucciones que debes seguir para realizar la prueba

En esta actividad vas a escribir dos textos relacionados con el mundo prehistórico. A continuación, te daremos algunas indicaciones sobre cómo debes hacerlo. Presta mucha atención:

- ✓ Evita las repeticiones.
- ✓ Usa adecuadamente los signos de puntuación: punto y seguido para separar frases, punto y aparte para separar párrafos y comas cuando sea necesario.
- ✓ Utiliza un vocabulario apropiado a las tareas.
- ✓ Evita las faltas de ortografía: pon mucha atención y repasa el texto cuando hayas terminado.
- ✓ Escribe con letra clara y cuida la presentación: no realices borrones y respeta los márgenes.
- ✓ Cuando acabes, revisa bien lo que has escrito.

¡A escribir!

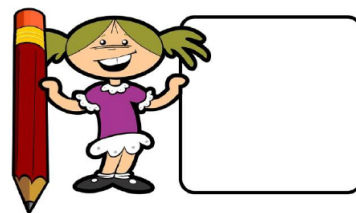


Figura 2.47. Ejemplo de pregunta de respuesta abierta con indicaciones sobre la corrección

A TEACHER'S DAILY LIFE

This is Tom. He is a teacher in a primary school.
What do you think he does every day? At what time? Where? ...

Write, at least, 5 sentences describing Tom and his daily life.

**Word limit:
Between 40 and
50 words**

6E111	0-1	
6E112	0-1	
6E112	0-1-2	
6E113	0-1-2	
6E114	0-1	
6E115	0-1	
6E116	0-1	
6E117	0-1-2	
6E118	0-1-2	
6E119	0-1	

REMEMBER

- Use the **present simple tense**.
- Be careful with **punctuation, capital letters and spelling**.
- You can use these words: *always, sometimes, never, and, but, also...*

This is Tom, and he is _____

- Si en el enunciado se indica que se deben escribir al menos cinco frases, una producción de esta longitud que contenga entre 40 y 50 palabras debe poder obtener la máxima puntuación.
- También es importante tener en cuenta que en la rúbrica de evaluación no se puede codificar de forma negativa si, por ejemplo, la persona evaluada no emplea ciertas formas verbales, tipos de expresiones o cualquier cuestión en la respuesta si esta no se le ha solicitado. Por ejemplo, en la pregunta de la Figura 2.47 se le da al alumnado una propuesta de términos (*always, sometimes, never, and, but, also...*), pero su uso no es obligatorio. Por tanto, la guía de codificación no puede penalizar por no emplearlos.



Elaboración de pruebas de evaluación

Tipos de ítems

ÍTEMS DE RESPUESTA CERRADA

Precisan que se seleccione una opción de entre varias ofrecidas, se relacione diferentes posibilidades, se ordenen ciertas opciones, etc.

Si se introduce un objeto en un recipiente con agua, la altura del líquido en el recipiente aumenta. Podemos medir cuánto ha aumentado la altura, pero, en realidad, ¿qué magnitud del objeto estaríamos midiendo al llevar a cabo este experimento?

A. La masa
B. El volumen
C. La altura
D. La densidad

Opción múltiple

Operador	Escribe aquí la letra que corresponda	Utilidad
A. Manivelas	...	Facilitan el desplazamiento de la bicicleta.
B. Engranajes	...	Transmiten el movimiento de los pedales a la rueda trasera.
C. Palancas	...	Hacen girar los platos.
D. Ruedas	...	Accionan mecanismos como los frenos.

Florence pertenece a una familia con mucho dinero por lo que de niña viajaba mucho. A Florence le pusieron ese nombre porque nació durante un viaje en la ciudad italiana de Florencia, una de las ciudades más bellas del mundo.

A Zaid le llama la atención la forma de los elementos de la fachada de una de sus iglesias, la Basílica de Santa María.

Observa la imagen y señala con X qué figura geométrica tiene cada una de sus partes:

- Triángulo
- Cuadrado
- Pentágono
- Circunferencia
- Círculo

Opción múltiple compleja

Emparejamiento o relación

ÍTEMS DE RESPUESTA SEMICONSTRUIDA

Implican que se rellenen una o varias casillas o espacios con una palabra o cifra, se completen varias frases, o responda a alguna cuestión con una respuesta corta.

Varias respuestas posibles

No solo se diferencian en el modo de registrarlas (mayúsculas o minúsculas, cifras o letras...)

Nombra a dos autores de cada movimiento literario:

1. Generación del 98: y

2. Generación del 27: y

Respuesta única tipo GAP

En la sala de lectura, Fátima y Pablo han visto un póster en el que se muestra el tipo de lectura que prefieren los españoles mayores de 10 años en un gráfico.

¿Cuál es el tipo de lectura preferido entre los adultos mayores de 18 años?

Tipo de lectura	De 10 a 14 años	De 15 a 18 años	Adultos mayores de 18 años
Libros	~95	~85	~80
Webs, blogs y foros	~45	~65	~45
Revistas	~25	~35	~35
Redes sociales	~15	~65	~45
Cómic	~10	~10	~10
Periódicos	~10	~10	~60

Escribe el tipo de lectura

ÍTEMS DE RESPUESTA ABIERTA O CONSTRUIDA

Exigen construcción por parte de la persona evaluada.

Si los rectángulos ABCD y EBCF que forman la ventana son semejantes, ¿cuánto mide la longitud x indicada en el dibujo? Razona tu respuesta.



Cuestionarios de contexto

Elaboración de cuestionarios de contexto

Dado que el contexto tiene una influencia decisiva en el rendimiento del alumnado, no es posible evaluar un sistema educativo con la pretensión de utilizar los resultados de esta evaluación para mejorarlo sin considerar el contexto en el que se encuentran sus estudiantes. Estos factores contextuales son dinámicos y complejos, y su estudio exige una preparación minuciosa y una idea clara de qué información se quiere obtener, de quién, y para qué. Para realizar un correcto diagnóstico del sistema, es preciso conocer estas variables que describen el contexto, los procesos que se den tanto en el aula como en el centro, así como los recursos educativos y de aprendizaje disponibles.

Los cuestionarios de contexto son instrumentos de recogida de información en los que se realizan una serie de preguntas que suelen ser numerosas y que abarcan una gran cantidad de aspectos relacionados con los estudiantes, el profesorado, las familias y los equipos directivos, aunque no necesariamente de todos ellos: dependerá de la amplitud o de las pretensiones del estudio que se quiera realizar y también del nivel educativo al que se dirija la evaluación. En los cursos de Educación Primaria, parece claro que debe complementarse la información que dé el alumno en su cuestionario con la del cuestionario de familia. En todo caso, debe tenerse en cuenta que en las respuestas hay siempre, en todos los niveles, un componente subjetivo que puede suponer un margen de error. Se trata, pues, un “sistema de cuestionarios” (Orden Hoz y Jornet Meliá, 2012; Jornet *et al.*, 2012) interrelacionados orientados a la recogida de una información capaz de explicar la evaluación del rendimiento educativo.

Diseño de los cuestionarios de contexto

El diseño de los cuestionarios de contexto requiere reflexionar sobre la información que desea recabarse y, por lo tanto, establecer unos indicadores que sean medibles. Este trabajo debe realizarse juntamente con las pruebas de evaluación cognitivas y confeccionarse sobre una base que comparta los mismos conceptos teóricos (por ejemplo, qué se entiende por “calidad” o por “equidad” en el ámbito educativo) y los mismos objetivos (la mejora del sistema) que rigen la elaboración e interpretación de las pruebas cognitivas. Por ello, es fundamental elegir a los agentes o informantes óptimos (tanto cualitativamente, ya que aportarán la mejor y la más apropiada información, como cuantitativamente, pues facilitarán la mayor cantidad de información relevante) a quienes se van a dirigir los cuestionarios. El éxito de todo el proceso comienza con la correcta selección de alumnado, familias, profesores, inspectores, equipos directivos, etc. En esta primera fase se deben plantear cuestiones fundamentales, por ejemplo, si dos agentes pueden informar sobre el mismo aspecto del estudio (para enriquecer o para contrastar la información) o si un agente, a pesar de ser teóricamente el idóneo, no es adecuado por no ser accesible o por prever que la información que facilite no va a ser veraz o completa.

Composición de los cuestionarios de contexto

En una segunda fase del diseño de los cuestionarios, además de las instrucciones para cumplimentarlos que suelen aparecer al principio, y que deben ser claras, concisas y adaptadas a su destinatario, se deben diseñar los ítems propios de estos cuestionarios, que se componen de enunciados (normalmente preguntas o afirmaciones) que algunos autores llaman “reactivos”, y sus opciones correspondientes de respuesta, que se realiza normalmente con una “x” o un círculo que se coloca en una casilla (a veces las respuestas son abiertas, pero hay que valorar muy bien su inclusión en el cuestionario, pues no son cómodas de responder para el encuestado y suponen una dificultad añadida en la codificación y en el manejo de datos). Del diseño y la redacción de las preguntas que componen el cuestionario depende la calidad de la información que posteriormente se obtiene. Por ello es necesario incluir una presentación o justificación del estudio al principio y un agradecimiento por la colaboración al final. El orden de las preguntas es también importante, por lo que se suele estructurar en secciones con preguntas relacionadas entre sí, por ejemplo cuestiones sociodemográficas al inicio, preguntas relacionadas con el contexto educativo a continuación, primero las más generales y luego las más específicas (por ejemplo, relacionadas con el aprendizaje de las ciencias en una evaluación de esta competencia) y preguntas referentes a emociones o sentimientos en una parte avanzada del cuestionario, cuando la persona que responda esté más familiarizada con este. Debe indicarse que los cuestionarios que se realizan en evaluaciones educativas son siempre anónimos por lo que también ha de planificarse el procedimiento que garantizará el anonimato de las respuestas.

En cuanto a la redacción de las preguntas, el lenguaje debe adaptarse al colectivo al que se dirigen procurando que sea sencillo, entendible y con frases breves. La redacción debe de ser neutral no induciendo a una determinada respuesta. Asimismo, se evitará la ambigüedad con preguntas formuladas de forma directa y personal. Las preguntas más personales (por ejemplo, aquellas que indagan sobre el acoso o la violencia) se formularán de forma indirecta.

Los enunciados de un cuestionario de contexto se refieren a variables cuya variabilidad está vinculada de manera sistemática al conjunto de lo que se quiere evaluar, y que recogen indicadores relativos a los recursos de los centros, el contexto socioeconómico de alumnado, los procesos de enseñanza, etc., planteados a los agentes a través de los reactivos. Estos indicadores pueden ser “simples”, si es un único reactivo el que se plantea al encuestado, o “complejos”, cuando se combinan varias informaciones.

En este sentido los indicadores resultantes del análisis de las respuestas del cuestionario pueden clasificarse como “simples”, y se pueden expresar a través de un porcentaje; un indicador “complejo” se calcula mediante procedimientos estadísticos y se expresa, generalmente, mediante variables continuas, con una cierta media y desviación típica, de forma que se pueden realizar comparaciones o valoraciones comparativas entre diferentes conjuntos de datos (por ejemplo, diferentes grupos sociodemográficos).

La Figura 3.1 muestra los indicadores definidos para la evaluación final de Educación Primaria realizada en los cursos 2015-2016 a 2018-2019⁴.

Figura 3.1. Ejemplo de indicadores para el diseño de cuestionarios de contexto

		Categorías
Indicadores expresados en porcentajes	Género	Chico; Chica
	Trimestre de nacimiento	T1; T2; T3; T4
	Escolarización temprana	0 o 1 años; 2 años; 3 años; 4 o más años
	Antecedentes de inmigración	Ninguno de los progenitores nacido en España. Uno ha nacido en España y el otro no. Los progenitores han nacido en España.
	Expectativa de los progenitores respecto a los estudios de sus hijos	ESO, Bachillerato o FPGM, Grado o FPGS, Máster, Doctorado
Indicadores medidos de forma continua (media 10 y desviación típica 2)	Utilización de medios tecnológicos para tareas escolares	
	Satisfacción de los alumnos y las familias con el centro	
	Satisfacción de los alumnos con el profesorado y sus resultados	
	Índice de clima escolar y resultados	
	Metodología del profesorado y resultados	
	Ayuda que reciben los estudiantes y resultados	
Indicador medido de forma continua (con media 0 y desviación típica 1)	Absentismo escolar y resultados	
	Índice social, económico y cultural (ISEC)	

4 Debido a la pandemia por coronavirus esta evaluación no se realizó en el curso 2019-2020.

Los tipos de preguntas que se plantean en el cuestionario son de diferentes tipos (González Nagel, 2012). Principalmente se diferencian los reactivos de respuesta directa (ej. *¿A qué edad comenzaste a ir al colegio o guardería?*), reactivos de respuesta dicotómica o de opciones múltiples y las baterías de afirmaciones que se responden mediante escalas de tipo *Likert*, que es una herramienta graduada que permite conocer el grado de acuerdo o desacuerdo que el informante tiene en relación con la pregunta o afirmación planteada. Esta escala puede seguir también un criterio de frecuencia, nivel de importancia, valoración, etc. La Figura 3.2 recoge un ejemplo de este tipo de pregunta que procede del cuestionario del director o directora utilizado (junto con los cuestionarios del alumno y de las familias) en las pruebas de diagnóstico de evaluación final de Educación Primaria diseñadas por el INEE:

Figura 3.2. Ejemplo de enunciado de un cuestionario de contexto y sus reactivos

REACTIVO

1. Indique su grado de acuerdo con las siguientes afirmaciones referidas a su centro.

	Nada de acuerdo	Poco de acuerdo	Bastante de acuerdo	Muy de acuerdo
a) El personal del centro comparte una línea metodológica común	1	2	3	4
b) Existe un elevado nivel de cooperación entre el centro y la comunidad local	1	2	3	4
c) El personal del centro habla abiertamente sobre las dificultades que encuentra en el desempeño de su labor docente	1	2	3	4
d) Existe un respeto mutuo hacia las ideas de los compañeros	1	2	3	4
e) Existe una cultura de compartir éxitos y fracasos	1	2	3	4
f) Las relaciones entre el profesorado y el alumnado son buenas	1	2	3	4

REACTIVOS QUE SE COMBINAN DE MANERA COMPLEJA PARA RESPONDER A UN INDICADOR

➤ El Índice social, económico y cultural en los cuestionarios de contexto

Uno de los indicadores que ha mostrado mayor capacidad explicativa del rendimiento del alumnado es el que mide el nivel socioeconómico y cultural de las familias: el índice social, económico y cultural (ISEC). La construcción de este indicador y el análisis de su relación con los resultados de las pruebas cognitivas es fundamental en los estudios de evaluación educativa, puesto que la relación entre ambas variables es una medida de la equidad de los sistemas. Este se suele expresar mediante una variable continua que, para todo el conjunto del alumnado participante, tiene media 0 y desviación típica 1. A continuación, para cada país, región o comunidad autónoma se obtiene el valor del ISEC realizando la media de los índices de los alumnos, ponderándolo de forma que cada alumno o alumna tiene un peso equivalente al número de individuos de la población a los que representa. El ISEC se elabora a partir de datos como:

- El nivel de estudio de los progenitores del alumno.
- El nivel de desempeño profesional de los mismos.
- Los recursos culturales o educativos, o relacionados con sus estudios, de que dispone el alumno en su hogar (si dispone de una habitación o de una mesa de trabajo propias o debe compartirla, ordenador personal, conexión a internet, etc.).
- En general, la cantidad de libros que hay en casa del alumno se considera adicionalmente y de forma separada al resto de recursos, puesto que hasta el momento se ha mostrado como variable muy influyente en el rendimiento educativo ya que, hasta el momento actual y a pesar de la cada vez mayor digitalización de este recurso, funciona como un “proxy” muy adecuado del nivel cultural familiar.
- Otros recursos disponibles en el hogar.

Características de los cuestionarios de contexto

Como ocurre con las pruebas cognitivas, el resultado final de la elaboración de un cuestionario de contexto debe tener las siguientes características:

1. **Validez:** que la información que se recoja sea válida quiere decir que está en relación con los objetivos que se persiguen.
2. **Consistencia:** la consistencia interna de un estudio tiene que ver con la correlación de los reactivos empleados. Así, en el ejemplo de la Figura 3.2 la aparición de grandes discrepancias en las respuestas a los reactivos d) y e) serían indicio de inconsistencia.
3. **Objetividad:** que carezca de sesgo (es decir, que no favorezca o perjudique a una parte de la población a la que se dirige el estudio).
4. **Actualización:** dado que, como se señaló antes, los contextos, como los propios encuestados, son dinámicos y cambiantes, se debe ajustar la realidad de los mismos al cuestionario.

Recordemos además que un cuestionario se aplica junto con otros dirigidos a otros colectivos, y todos ellos junto con las pruebas cognitivas: entre todos estos elementos debe existir una armonía en el diseño y un calendario de aplicación coherente. Por ello, es conveniente que las pruebas piloto o pre-piloto que se realicen de las pruebas cognitivas incluyan también los cuestionarios de contexto para poder analizar su correcto funcionamiento y el cumplimiento de sus objetivos.

Bibliografía

- Barbero, M. I., Vila, E. y Suárez, J. C. (2006). *Psicometría*. Madrid: UNED.
- Benson, J. y Crocker, L. (1979). The effects of item format and reading ability on objective test performance: a question of validity. *Educational and Psychological Measurement*, 39 (2), 381-387.
- Bloom, B. S. (Ed.) (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. David McKay Company. p. 201-207.
- Brame, C. (2013). Writing good multiple choice test questions. Recuperado de <https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/>.
- Centre for Teaching Excellence, University of Waterloo (s.f.). Exam questions: types, characteristics and suggestions. Recuperado de <https://uwaterloo.ca/centre-for-teaching-excellence/teaching-resources/teaching-tips/developing-assignments/exams/questions-types-characteristics-suggestions>.
- Consejo de Europa (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. (CEFR). Cambridge University Press.
- INTEF (2014). Cervera, D., González, P., Sánchez, J. M. (INEE) Evaluaciones Externas Internacionales del Sistema Educativo. Bloque 2. Formación en red. Recuperado de http://formacion.intef.es/pluginfile.php/120708/mod_resource/content/1/EEISE_Modulo02_v11_parte1.pdf.
- Cortés de las Heras, J. (2019). Cuestionarios cognitivos (IV): Ítems de respuesta cerrada y abierta. Recuperado de <http://revolucion.blogspot.com/2019/06/cuestionarios-cognitivos-iv-items-de.html>.
- Díaz Barriga, F. y Hernández, G. (2010). *Estrategias docentes para un aprendizaje significativo (3.ª Ed)*. México: McGraw-Hill.
- García González, D. (2017). *Diez Comodines: Cómo redactar mejores exámenes tipo test*. Bilbao: Triple Eñe.
- García Ramos, J. M. (1989). *Bases pedagógicas de la evaluación: guía práctica para educadores*. Madrid: Síntesis.
- Gierl, M. J., Bulut, O., Guo, Q. y Zhang, X. (2017). Developing, Analyzing, and Using Distractors for Multiple-Choice Tests in Education: A Comprehensive Review. *Review of Educational*.
- González Nagel, E. J. (2012). Cuestionarios de contexto en los estudios de la IEA. *Bordón. Revista de pedagogía*, 63 (2), 29-39.
- Haladyna, T. M. y Downing, S. M. (1989). A Taxonomy of Multiple-Choice Item-Writing Rules. *Applied Measurement in Education*, 2 (1), 37-50.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Psychology Press.
- Jorba, J. y Sanmartí, N. (1993). La función pedagógica de la evaluación. *Aula de innovación educativa* (20), pp. 20-30: https://ddd.uab.cat/pub/artpub/1993/187512/aulinnedu_a1993n20aJorba.pdf.
- Jornet, J. M., González-Such, J. y Perales, M. J. (2012). Diseño de cuestionarios de contexto para la evaluación de sistemas educativos: Optimización de la medida de constructos complejos. *Bordón. Revista de pedagogía*, 64 (2), 89-110.

- Kelly, M. (2018). Creating effective fill-in-the-blank questions. Recuperado de <https://www.thoughtco.com/creating-effective-fill-in-the-blank-questions-8438>.
- Kelly, M. (2019). Tips to create effective matching questions for assessments. Recuperado de <https://www.thoughtco.com/effective-matching-questions-for-assessments-8443>.
- INEE (2012). Estudio europeo de competencia lingüística. EECL. Volumen I. Informe español. Recuperado de <https://www.educacionyfp.gob.es/inee/dam/jcr:4b869d66-f743-4394-9b9d-39007557a4ca/eeclvolumeni.pdf>.
- INEE (2015a). Marco General de la evaluación de 3^{er} curso de Educación Primaria. Recuperado de <https://www.educacionyfp.gob.es/inee/dam/jcr:40df2f8e-0ee4-4a17-9ce6-1f0fd1437240/marcoev3ep08012015.pdf>.
- INEE (2015b). Marco General de la evaluación final de Educación Primaria. MECD. Recuperado de <https://www.educacionyfp.gob.es/inee/dam/jcr:55b35040-4d36-4c38-a420-eeb440dedf41/marco-teorico-evaluacion-final-6ep.pdf>.
- Mullis, I. V. S. y Martin, M. O. (Eds.) (2015). PIRLS 2016. Marco de Evaluación. Boston College, TIMSS & PIRLS. Recuperado de <https://www.educacionyfp.gob.es/inee/dam/jcr:d79b8f8b-d4a8-42b7-b63e-42aed382a8e7/pirls2016webokk.pdf>.
- Mullis, I. V. S. y Martin, M. O. (Eds.) (2017). TIMSS 2019. Marco de Evaluación. Boston College, TIMSS & PIRLS. Recuperado de <https://www.educacionyfp.gob.es/dam/jcr:8bac8bae-0621-4bff-a6ce-75e4cc144410/timss%202019%20marcos%20de%20la%20evaluacion.pdf>.
- Muñiz, J., Fidalgo, A. M., García-Cueto, E., Martínez, R. y Moreno, R. (2005). Análisis de los ítems. Madrid: La Muralla.
- Organización para la Cooperación y el Desarrollo Económicos, OCDE (2017). *PISA 2018, Assessment and Analytical Framework*. París: OECD Publishing.
- Orden Hoz, A. de la (1989). Investigación cuantitativa y medida en educación. *Bordón: Revista de Orientación Pedagógica*, 41 (2), 217-236.
- Orden Hoz, A. de la, y Jornet Meliá, J. M. (2012). La utilidad de las evaluaciones sistemas educativos: el valor de la consideración del contexto. *Bordón: Revista de Orientación Pedagógica*, 64 (2), 89-110.
- Rodríguez, M. C. (1997). The art and science of item-writing: A meta-analysis of multiple choice item formal effects. Recuperado de <https://pdfs.semanticscholar.org/f39c/c381e1c3e8a3db6f5e9c884abbfbecf4b057.pdf>.
- Shepard, L. A. (2006). La evaluación en el aula. En R. L. Brennan (Ed.), *Educational Measurement* (pp. 623-646). México, D. F.: Instituto Nacional para la Evaluación de la Educación.
- Statman, S. (1988). Ask a clear question and get a clear answer: An enquiry into the question/answer and the sentence completion formats of multiple-choice items. *System*, 16 (3), 367-376.
- Stecher, B. M., Rahn, M. L., Ruby, A., Alt, M. N. y Robyn, A. (1997). *Using Alternative Assessments in Vocational Education*. Washington, D. C.: RAND.
- Villa A. y Poblete, M. (2007). *Aprendizaje basado en Competencias. Una propuesta para la evaluación de las competencias genéricas*. Estudios e investigaciones del I.C.E. de la Universidad de Deusto. Bilbao.



La evaluación educativa es un proceso esencial para realizar con garantías el diagnóstico y mejora de cualquier sistema educativo. Su diseño, aplicación y análisis supone la concurrencia de diversos campos del conocimiento, además de los directamente vinculados a los currículos escolares, cada uno con sus propios objetivos y peculiaridades pedagógicas.

La necesaria participación de estos conocimientos técnicos y conceptuales (estadística, sociología, epistemología, pedagogía...) en la evaluación educativa ha contribuido a que las investigaciones que los tienen como objeto se hayan acrecentado en los últimos años, propiciando una amplia bibliografía referida a la teoría y práctica de la evaluación educativa, entre la que ocupa un papel central la elaboración de las pruebas cognitivas, punto de intersección entre los equipos técnicos que las elaboran para las Administraciones u organismos y su genuino destinatario, que no es otro que el alumnado.

Con el fin de aportar nuestra experiencia en el diseño de estas pruebas cognitivas, desde el área de evaluación nacional del INEE hemos elaborado la presente *Guía para la elaboración de pruebas de evaluación educativa*, en la que han colaborado el resto de las áreas del Instituto. Si con ello podemos aportar algo útil a la comunidad educativa, daremos por bien empleado este trabajo.