

# GUÍA DE EVALUACIÓN EDUCATIVA

*para el profesorado de educación primaria y secundaria*

J. Cortés De las Heras



Guía de evaluación educativa para el profesorado de educación primaria y secundaria

Autor: Javier Cortés De las Heras (email: cortes\_jav@gva.es)

Diseño y maquetación: Perro-Ballena Productions

Imagen de portada: Zsuzsanna Kilian

Esta obra está sujeta a la licencia Reconocimiento-CompartirIgual 4.0 Internacional de Creative Commons.

Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by-sa/4.0/>.



# Índice de contenidos

<b>0. La evaluación educativa a vista de pájaro.....</b>	<b>5</b>
<b>1. Los resultados de aprendizaje.....</b>	<b>7</b>
<b>2. Tipos de evaluación.....</b>	<b>11</b>
<b>3. Procedimientos, pruebas e instrumentos de evaluación y escalas de medida.....</b>	<b>17</b>
<b>4. Fases en la evaluación referida al criterio para clase.....</b>	<b>25</b>
<b>Glosario.....</b>	<b>27</b>
<b>Lecturas recomendadas.....</b>	<b>29</b>



# 0. La evaluación educativa a vista de pájaro

En esta guía no se ofrecerá una disquisición teórica sobre qué se entiende por evaluación educativa, actitud y acción muy frecuente en el ámbito académico. Todo lo contrario, se ofrece una única definición, de entre las múltiples y a menudo bastante coincidentes, salvo en pequeños detalles. Esta única definición y tantas otras que aparecerán a lo largo de esta guía, o en el *Glosario*, se basan generalmente en una traducción de los términos del tesoro de ERIC (<http://www.eric.ed.gov>), que se tomará, siempre que sea posible, como referencia.

Empecemos por definir el término evaluación en sentido amplio, para después concretarlo en el término que más nos concierne en esta guía, evaluación educativa<sup>1</sup>.

## Evaluación:

*Valorar o juzgar personas, organizaciones o cosas en relación con objetivos, estándares o criterios establecidos.*

## Evaluación educativa:

*Determinar e interpretar el logro de objetivos educativos (en el ámbito nacional, regional o local) para su uso en la planificación educativa y su desarrollo, la política de formación y la dotación de recursos.*

Las cuestiones que surgen en definitiva serán:

- ¿cómo se determinan e interpretan los logros de personas (alumnos) en relación con esos objetivos educativos? procedimientos y pruebas.
- ¿con qué se determinan e interpretan? instrumentos de medida.

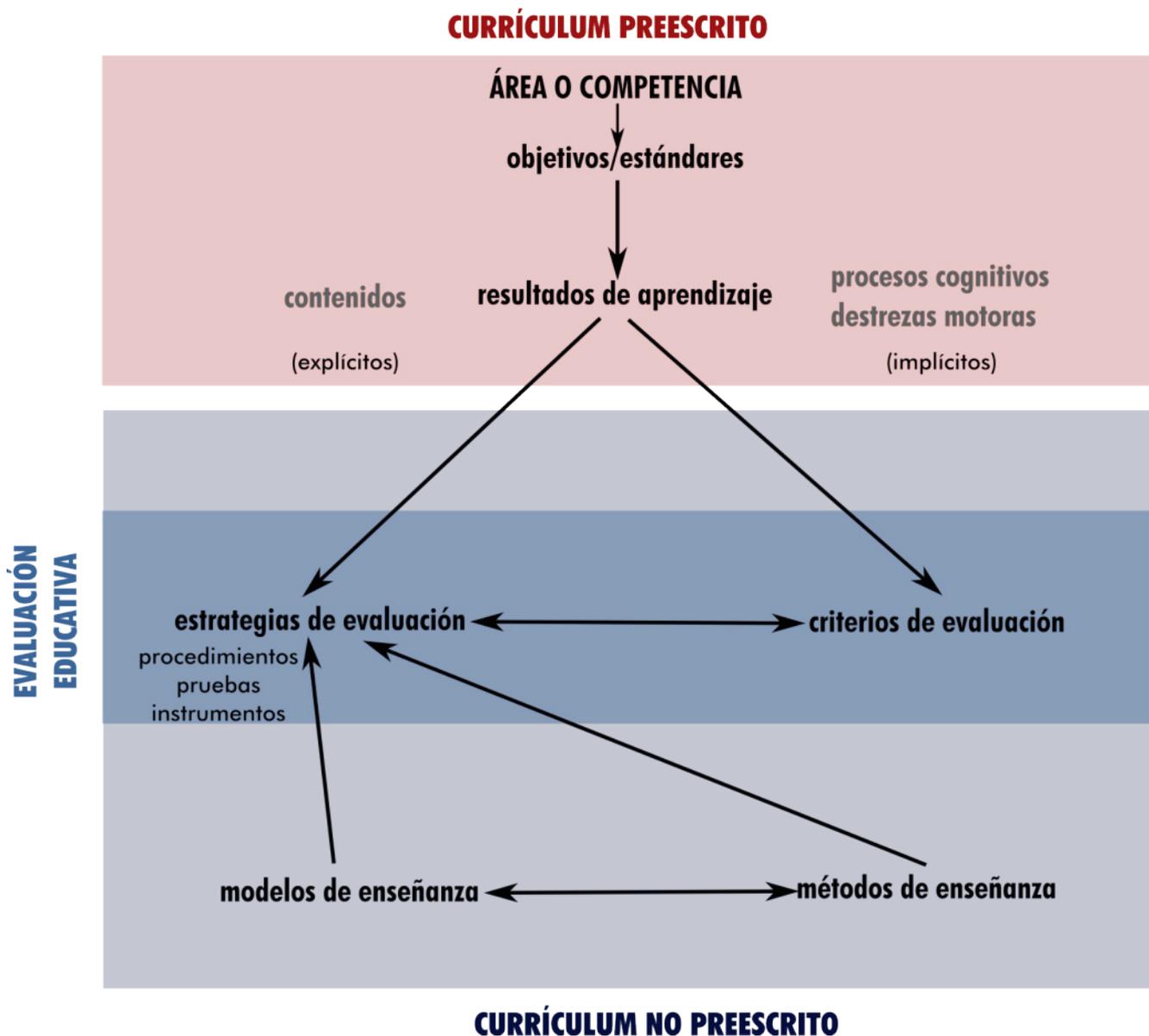
Una vez entrados en el campo de la evaluación educativa, uno de los grandes problemas de ésta, al igual que ocurre en otras disciplinas, es la dificultad que una comunidad de profesionales, en este caso docentes y administradores de la educación, tiene para entender a qué nos referimos con cada uno de los términos que la conforman.

La tradición en evaluación es larga y a lo largo del tiempo se han ido acuñando términos que en algunos casos han sido malinterpretados, especialmente aquellos que provienen de lenguas extranjeras (casi siempre) y que son traducidos a nuestra lengua.

En esta introducción van a aparecer en el esquema de la página siguiente muchos de estos términos. La mayor parte de ellos van a ser definidos en las siguientes secciones de este documento o, como se ha dicho, en el *Glosario*. Así que, por ahora, basta mirarlo y entender las relaciones que se establecen entre estos elementos, a modo de panorámica general e introductoria, pero también como resumen final, una vez se haya leído toda la guía.

---

(1) En el ámbito anglosajón se hace una distinción clara entre *evaluation*, aquí denotado como evaluación y *assessment (educational)*, denotado en esta guía como evaluación educativa. En educación, en el primer caso se suele aplicar a instituciones o programas, en el segundo a individuos.



Como se aprecia en el esquema hay algunos elementos que nuestro currículum no acuña de la misma forma o, utilizándolos, se entienden de otra manera. Tal es el caso de los *criterios de evaluación*, de los cuales se van dando pinceladas seguidamente y una definición en el *Glosario*. Lo interesante de este mapa es que pone de manifiesto que la evaluación no es una cuestión aislada del currículum, una acción independiente de los objetivos o resultados de aprendizaje, más aún, tampoco es independiente del modo y el método de enseñanza.

# 1. Los resultados de aprendizaje

Un resultado de aprendizaje es *una declaración de lo que el estudiante se espera que conozca, comprenda y sea capaz de hacer al finalizar un período de aprendizaje.* (trad. Moon, 2002).

Un resultado de aprendizaje bien escrito debe contener los siguientes componentes:

1. Un **verbo** que indique lo que el estudiante se espera que sea capaz de realizar al finalizar el período de aprendizaje.
2. Una palabra/s que indiquen sobre qué o con qué el estudiante actúa, **contenido**. Si el resultado es sobre destrezas entonces la palabra o palabras deberían describir el modo en el que la habilidad se ejecuta (p.ej.: saltar eficazmente arriba y abajo).
3. Una palabra/s que indique la **naturaleza** (en contexto o en términos de estándar) de la ejecución requerida como evidencia de que el aprendizaje se ha logrado.



En evaluación son de suma importancia, ya que van a permitir:

1. Delimitar la calidad de la ejecución, es decir, establecer criterios de evaluación.
2. Determinar las estrategias de evaluación (procedimientos, pruebas e instrumentos).

Obviamente, también van a marcar claramente cómo vamos a enseñar, los modelos y métodos más adecuados para que el alumnado alcance los resultados esperados.

Asimismo, también van a tener suma importancia a la hora de determinar el nivel de competencia de un alumno concreto, precisamente por la exigencia de la tarea a demostrar, marcada especialmente por el verbo. El uso de taxonomías, como la de Bloom, permite establecer distintos niveles de exigencia en el aprendizaje, de forma que habrá resultados de aprendizaje más fácilmente alcanzables por parte del alumnado que otros. Ejemplos de esto también se reflejan en los niveles de competencia establecidos en las pruebas de PISA.

En nuestro currículum<sup>2</sup>, los resultados de aprendizaje podemos encontrarlos en forma de "criterios de evaluación". Un análisis de alguno de ellos pone de manifiesto que dichos criterios no son más que manifestaciones de lo que se espera que un alumno alcance al finalizar un ciclo o nivel en un área o materia determinada.

---

(2) En la Ley Orgánica 8/2013 de Mejora de la Calidad Educativa, art. 6, aparecen asimilados a estándares de aprendizaje (es decir se entiende que son lo mismo) aunque después la redacción de la ley se decanta por utilizar el término estándar, como una especificación de los criterios de evaluación, para que éstos sean evaluables. Sin embargo, ni el término criterio de evaluación ni el término estándar de aprendizaje está utilizado de forma adecuada, según las definiciones proporcionadas en este documento.

## Análisis de un resultado de aprendizaje (criterio de evaluación)

En este apartado, se muestra el análisis de un criterio de evaluación (resultado de aprendizaje) del currículo de Ed. Primaria de la Comunitat Valenciana para el área de Matemáticas en el 2º ciclo.

### 14. Situar y expresar correctamente la localización de un elemento en un croquis o en un plano sencillo.



Situar y expresar



la localización de un elemento



en un croquis o en un plano sencillo correctamente

Dificultades encontradas:

1. Situar y expresar son dos verbos, opción no demasiado recomendable, especialmente si una de ellas no se sabe muy bien a qué acción llama al estudiante a hacer en relación con el contenido (*expresar*).
2. "Correctamente" no aporta mucho. Naturalmente lo que pretendemos es que la acción siempre se haga correctamente, pero esto no ayuda a establecer umbrales estándar de ejecución a partir de los cuáles poder valorar el grado de ejecución. El docente tendrá que determinar qué elementos hacen que la tarea sea valorada como correcta.

No siempre aparece la naturaleza o contexto en el resultado de aprendizaje, pero si lo hace suele hacer referencia, respecto de la tarea demandada, a:

- **cómo** – modos, estrategias de realización.
- **dónde** – elementos físicos (espacios, tiempos) de realización.
- **con qué** – materiales para la realización.
- **para qué** – fin perseguido con la realización.
- **cuánto** – frecuencia de realización.

A partir del resultado de aprendizaje, podemos establecer el criterio de evaluación. Es decir, delimitar las bases sobre las que se va a realizar el juicio o valoración de adecuación del trabajo realizado. O dicho de otro modo, los elementos que especifican cómo va a ser evaluado, y en qué condiciones, un determinado trabajo (procedimientos, pruebas e instrumentos y otros condicionantes).

#### Tarea propuesta:

Analiza algún resultado de aprendizaje (criterios de evaluación en el currículo) de tu área, mejor si son varios.

- ¿se dan siempre los tres elementos de todo resultado de aprendizaje?
- ¿observas alguna incoherencia o mala especificación?
- ¿crees que los resultados de aprendizaje del currículo cubren todos los niveles de una taxonomía como la de Bloom (puedes consultar la página siguiente)?

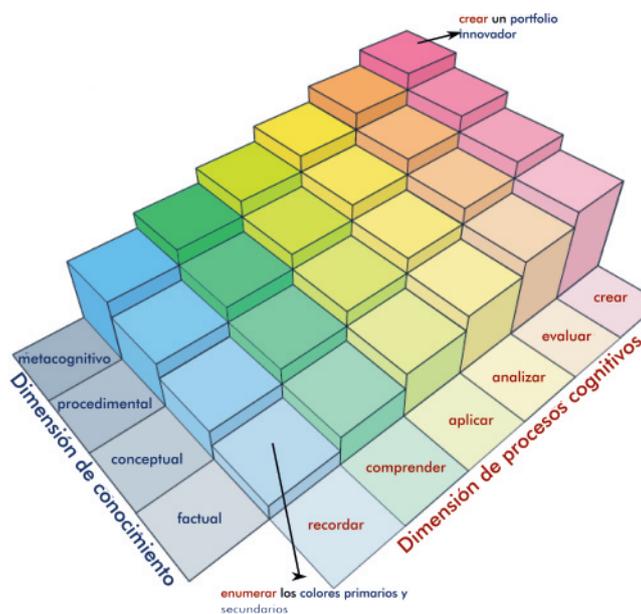
# Taxonomías

Es interesante hacer un repaso de alguna taxonomía, como la de **Bloom** (revisada), de clasificación de objetivos educativos. Como se sabe, se establecen tres dominios: cognitivo, afectivo y psicomotor. El más conocido, y por ejemplo usado en evaluaciones internacionales como PISA de la OCDE o TIMSS de la IEA, o en nacionales como la Evaluación General de Diagnóstico de 2009, es el dominio cognitivo.



Otro tipo de taxonomías como la de **Reigeluth** o la de **Merrill**, también permiten clasificar los objetivos de aprendizaje, y por consiguiente a ayudar al alumnado a alcanzarlos y al profesorado a evaluarlos.

En la siguiente imagen se utiliza una taxonomía basada en la obra de Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives* (Complete edition). New York: Longman, realizada por Rex Heer (<http://www.celt.iastate.edu/teaching/RevisedBlooms1.html>).



# Trazabilidad de los resultados de aprendizaje

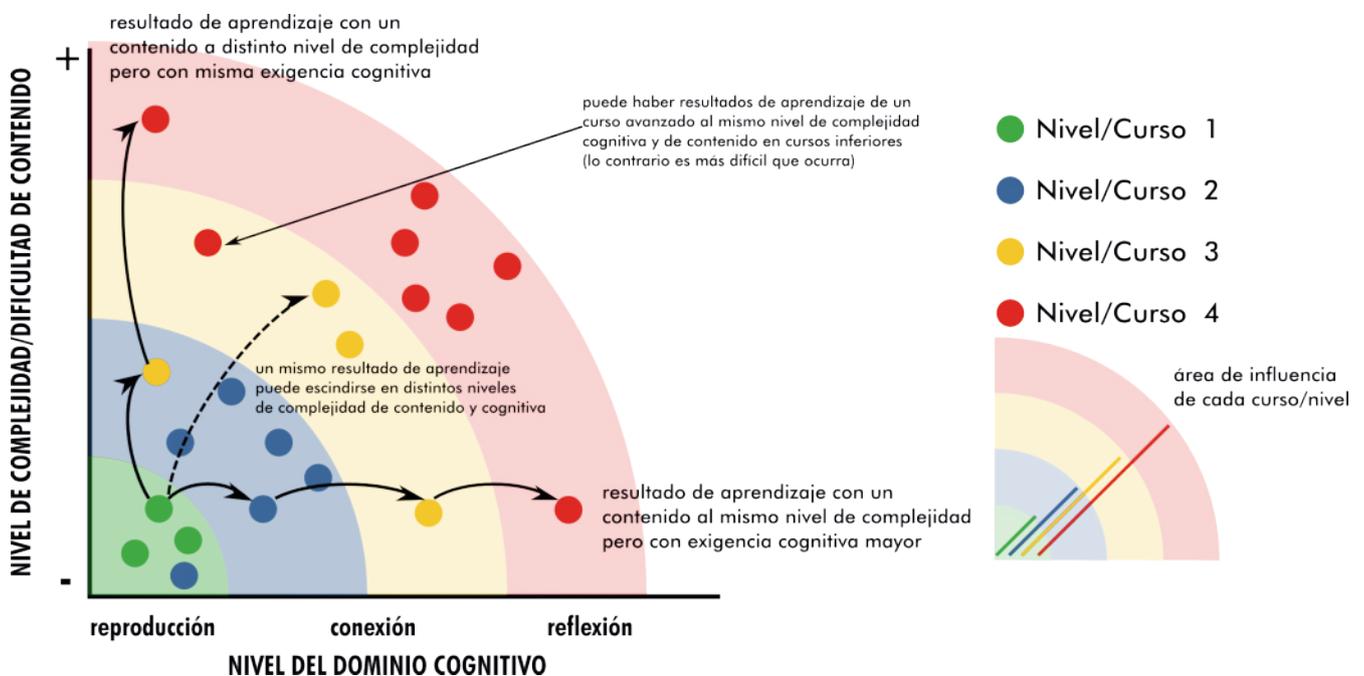
Bien los estándares u objetivos, bien su concreción, resultados de aprendizaje deberían haber estado concebidos en el currículum atendiendo a su "trazabilidad". Es decir, a aquella característica que permite establecer una ruta lógica de aprendizaje desde niveles educativos inferiores a los superiores:

- En función de su contenidos (de más simple a más complejo).
- En función de la exigencia en el dominio cognitivo, en el dominio psicomotor o en el dominio afectivo.

Y también como elemento que puede añadir complejidad:

- En función del contexto.

En el gráfico sólo están representadas dos dimensiones, el contenido y los procesos, en este caso del dominio cognitivo.



Naturalmente, este esquema es un modelo, y como tal no tiene porque tener un reflejo exacto en la realidad. Pero en la medida en la que nuestro currículum esté diseñado de esta forma, será más fácil observar el progreso en el proceso de enseñanza/aprendizaje y su evaluación.

## Tarea propuesta:

Intenta seguir el rastro de un resultado de aprendizaje de tu área (criterios de evaluación en el currículum) en distintos niveles/cursos de la etapa donde impartes clase.

- ¿se escinde en niveles superiores o se agrupa?
- ¿desaparece en un nivel determinado o puedes seguir su evolución hasta finalizar la etapa?

## 2. Tipos de evaluación

Cuando se pretende hablar de tipos de evaluación, algunos autores suelen introducir elementos que pueden causar confusión en el lector (p.ej. Chatterji (2003), introduce como tipos de evaluación toda una serie de instrumentos clasificados según distintos criterios). En otros manuales se hace referencia a tipos de evaluación más centrados en ámbitos laborales o de gestión de calidad, en otros se establecen relaciones de tareas que deben hacer los estudiantes evaluados, etc.

En definitiva, existe bastante confusión a la hora de delimitar los tipos de evaluación o al menos sobre qué tipo de evaluación se está hablando y a qué se está aplicando. Esta breve referencia que se expone en este apartado no pretende tampoco dar respuesta a todas las incógnitas que surgen a la hora de describir qué tipo de evaluación hacemos con nuestros estudiantes. La clasificación que se ofrece puede ser revisada con más profundidad de la que aquí se expone, es discutible y provisional, pero al menos puede servir como un primer intento de “encajar las cosas”.

Los criterios de clasificación de los tipos evaluación que se manejan en este documento son los siguientes. Según:

1. El agente evaluador.
2. El momento.
3. El propósito.
4. El grado de formalidad.
5. Los usos e interpretación de la puntuación.
6. Los métodos de puntaje.
7. La tradición evaluativa.
8. El nivel de impacto.

Un tipo de evaluación respecto a un criterio de clasificación no impide que se dé otro tipo de evaluación en otro criterio. De hecho, muchos tipos de evaluación están altamente asociados, otros son transversales. Por ejemplo, la evaluación sumativa la podemos utilizar en la evaluación de ejecuciones o en la evaluación mediante pruebas objetivas; y también la formativa, por qué no.

Asimismo, también se podría mencionar como criterio el posicionamiento teórico de la evaluación que puede incluir la evaluación tradicional y la evaluación auténtica, asimilándola con otros términos como el de evaluación alternativa, ambas frente a la primera citada. Se expone una definición de *evaluación alternativa* como descriptor del Tesoro de la base de Education Resources Information Center (ERIC) (<http://www.eric.ed.gov>).

*(trad.: Estimación del conocimiento o las habilidades por otros medios distintos a los empleados tradicionalmente, pruebas objetivas, especialmente las pruebas de elección múltiple estandarizadas). (ERIC Thesaurus, 15/09/2000).*

A su vez este término incluye como término más específico el de *evaluación auténtica*, asimilada en el tesoro a *evaluación basada en ejecuciones*. Sin embargo, algunos autores argumentan el sentido que se le da a este tipo de evaluación yendo más allá de una evaluación basada en ejecuciones (Chatterji, 2003; Shelley y Howell, 2008, Wiggins, 1998).

CRITERIO DE CLASIFICACIÓN	TIPOS DE EVALUACIÓN
1. Agente evaluador	- Autoevaluación - Coevaluación - Heteroevaluación
2. Momento	- Evaluación diagnóstica (o inicial) - Evaluación continua - Evaluación final
3. Propósito	- Evaluación sumativa - Evaluación formativa
4. Grado de formalidad	- Evaluación informal - Evaluación estandarizada
5. Usos e interpretación de la puntuación	- Evaluación referida a la norma - Evaluación referida al criterio - Evaluación ipsativa
6. Consideración del objeto evaluado	- Evaluación holística - Evaluación analítica
7. Tradición evaluativa	- Evaluación tradicional - Evaluación alternativa
8. Nivel de impacto	- Evaluación de bajo impacto - Evaluación de alto impacto

## 1. La evaluación según el agente evaluador:

Hace referencia a quién es la persona que evalúa a otra persona o a sí mismo respecto a un objeto de evaluación (acción).

Puede distinguirse claramente entre: a) **autoevaluación**: la que realiza un individuo sobre su propia acción; b) **coevaluación**: la que realizan distintos individuos que realizan la acción u objeto de evaluación entre sí (la evaluación que realiza la clase a los distintos grupos de la misma clase que exponen un tema, la evaluación por pares,...); c) **heteroevaluación**: la más practicada, la que realiza una persona (p.ej.: en el proceso e/a, el profesor) sobre otra que realiza una acción (prueba, exposición, ensayo,...).

## 2. La evaluación según el momento en el que se realiza:

Distingue entre una evaluación realizada a lo largo de todo el proceso de enseñanza aprendizaje (e/a), de forma frecuente, muchas veces asimilada a la evaluación formativa, ya que ésta última ha de tener como requisito la continuidad. Sin embargo, la **evaluación continua** muchas veces puede no ser formativa. La **evaluación final** propone un único momento de evaluación al finalizar un proceso de e/a (puede ser una prueba parcial, a final de curso, al cabo de tres años) y generalmente no supone retroalimentación de la información para la mejora del propio aprendizaje. Su característica principal es que etiqueta el proceso de aprendizaje de alguna manera (pasa/no pasa; apto/ no apto; suspenso / aprobado / notable /...;...). La **evaluación diagnóstica**, es un tipo de evaluación

que se suele realizar antes de comenzar una intervención (formativa, terapéutica, reeducativa,...) para detectar fortalezas y debilidades de los evaluados y realizar así las adaptaciones que sean necesarias para que el proceso de aprendizaje o de modificación de conductas se adapte a las características de los individuos. Generalmente tiene un carácter individualizado.

### 3. Según el propósito:

Se distingue entre: a) **evaluación formativa**: la que evalúa el proceso de e/a de forma frecuente e interactiva proporcionando información para la mejora tanto del aprendizaje como de la enseñanza, por tanto ha de ser necesariamente continua, y b) **evaluación sumativa**: mide mediante pruebas (de forma continua o final) con el propósito de obtener una puntuación del producto de aprendizaje.

### 4. Según el grado de formalidad:

La **evaluación informal** sería la valoración del aprendizaje por medios distintos a las pruebas estandarizadas. Así, las **pruebas estandarizadas** (más que evaluación) se refiere a pruebas para las que el contenido ha sido seleccionado y comprobado empíricamente, se han establecido normas y métodos uniformes de administración y pueden ser puntuadas con un relativo alto grado de objetividad.

### 5. Según usos e interpretación de la puntuación:

Se encuentran dos tipos evaluación, o más bien pruebas (tests): a) **referidas a la norma** (TRN), en los que la puntuación de un individuo se compara con relación a un grupo de individuos (grupo normativo) medidos en el atributo que mide el test. Y b) **referidos al criterio** (TRC), en los que la puntuación de un individuo se compara con relación a un dominio de aprendizaje y destrezas. Siguiendo la definición de Shrock y Coscarelli (2000), en el primer caso un individuo siempre se compara con otro en términos de puntuación, mientras que en el caso de los TRC, el individuo no se compara con otros sino con el nivel dado en referencia a unos estándares que especifican el dominio de aprendizaje o habilidad.

### 6. Según la consideración del objeto evaluado:

Se suelen distinguir dos tipos de evaluación, principalmente usados en la evaluación de ejecuciones: a) **evaluación analítica**: basada en las valoraciones o juicios emitidos sobre las partes de un proceso o producto de aprendizaje de forma individual, que en suma darían una valoración del conjunto de ese proceso o producto. Suelen utilizarse para la valoración escalas analíticas. b) **evaluación holística**: basada en el juicio de un proceso o producto de aprendizaje en su conjunto, suelen utilizarse escalas holísticas (*scoring rubrics*) o ejemplos típicos de comparación que describen niveles de ejecución. Aunque no pertenezca a este criterio de clasificación, la evaluación holística además tiene una connotación más amplia asimilándose a la evaluación de procesos, continua y formativa.

### 7. Según la tradición evaluativa:

La **evaluación alternativa** es el uso de otras estrategias y pruebas de evaluación diferentes a la utilizadas habitualmente, las pruebas objetivas o pruebas escritas (especialmente aquellas de elección múltiple), que se situarían en el polo más tradicional de la evaluación.

### 8. Según el nivel de aplicación

La **evaluación de bajo impacto** (*low stakes*) es aquella cuyo uso se limita al ámbito de la instrucción en una clase o centro educativo y suele ser informativa. No suelen tener consecuencias más allá de la guía de los aprendizajes. Por el contrario, la evaluación cuyos resultados son los determinantes principales de consecuencias significativas en los individuos o instituciones, es la **evaluación de gran impacto** o consecuencias (*high stakes*) (*graduación, promoción o seguimiento a gran escala*).

## Pistas para una evaluación auténtica...

Se había comenzado el apartado haciendo mención a *evaluación alternativa* y *evaluación auténtica*, sin dejar claro a qué nos referimos con este último término. Es interesante conocer qué aspectos la definen, ya que la evaluación auténtica es algo más que hacer pruebas diferentes a las pruebas objetivas y no siempre tiene porque tener un componente de evaluación basada en ejecuciones, aunque esta característica es prácticamente inherente a su conceptualización.

En un artículo de revisión de literatura sobre el concepto (Frey, Schmitt y Allen, 2012) se delimitan nueve características que se han encontrado en definiciones realizadas por cuarenta y siete autores sobre la evaluación auténtica. De ellas, las más citadas para alumnado de educación primaria, y que reproducimos aquí, se refieren a:

### 1. Contexto

- Plantea una actividad o contexto realista (60% de las definiciones).
- Plantea tareas complejas desde el punto de vista cognitivo (30%).

### 2. Papel del estudiante

- Es una evaluación de carácter formativo (31%).

### 3. Puntuación

- Los estudiantes conocen o desarrollan los criterios de puntuación (47%).
- Se utilizan múltiples indicadores o se usa portfolio (como instrumento que aglutina diversas pruebas (evidencias de aprendizaje) e indicadores (53%).

En educación secundaria o formación profesional destacan, más o menos las mismas características.

La cuestión de que la evaluación auténtica ha de estar basada en ejecuciones cobra especialmente importancia en educación primaria (23%) y en mayor medida en áreas como la educación física o educación artística (plástica y música).

## ¿Qué tipo de evaluación está presente en nuestro currículo?

En el cuadro de la página siguiente se recoge un resumen del carácter de la evaluación educativa en la normativa de la Comunitat Valenciana. No se valorará críticamente ningún aspecto, se trata de señalar cómo se conceptualiza la evaluación normativamente para que sirva de reflexión sobre la práctica y sobre la propia política educativa. Así, se ha señalado en el cuadro si se dan en los distintos niveles de la enseñanza básica características de la evaluación: global/diferenciada, formativa/sumativa, inicial/continua/final y algún que otro aspecto más.

Parece que está bastante asumido, pero es necesario recordar que la normativa deja suficiente autonomía a los docentes para decidir especialmente sobre **cómo** se va a realizar la evaluación del alumnado. Por ejemplo, en las distintas normativas sobre evaluación no se hace mención, salvo algún apunte en Educación Infantil, al tipo de técnicas, instrumentos, pruebas o criterios de calificación a utilizar.

	Ed. Infantil	Ed. Primaria	Ed. Secundaria Obligatoria
<b>Normativa de referencia</b>	ORDEN de 24 de junio 2008, de la Conselleria de Educación, sobre la evaluación en la etapa de Educación Infantil (DOCV núm 5814)	ORDEN de 13 de diciembre de 2007, de la Conselleria de Educación, sobre evaluación en educación primaria.	ORDEN de 14 de diciembre de 2007, de la Conselleria de Educación, sobre evaluación en Educación Secundaria Obligatoria.
<b>Global vs. Diferenciada</b>	Global	Global En la práctica es diferenciada: cada maestro decide sobre los resultados de evaluación del área que imparta (art. 3.1) además los documentos oficiales de evaluación hacen diferenciación por áreas.	Diferenciada (materias, ámbitos y módulos)
<b>Continua</b>	Sí	Sí	Sí
<b>Formativa vs Sumativa</b>	Formativa	Formativa	Formativa
<b>Inicial</b>	Sí	Sí	Sí
<b>Dominio</b>	Capacidades (expresadas en objetivos de ciclo y criterios de evaluación; Decretos 37/2008 y 38/2008).	Sin perjuicio del carácter global se tomará como referencia los criterios de evaluación por áreas para valorar el grado de adquisición de: - competencias básicas - objetivos - conocimientos (Decreto 111/2007)  - Anexo V, de la orden de evaluación se establece un modelo para calificar la adquisición de competencias. No se establece el método o procedimiento para valorarlas a través de las áreas.	- Diferenciada por áreas. Se tomará como referencia los criterios de evaluación por materias, ámbitos y módulos para valorar el grado de adquisición de: - competencias básicas - objetivos (Decreto 112/2007)
<b>Técnicas e instrumentos</b>	Observación directa e indirecta Registro	Se hace mención a que se concretará en la programación didáctica anual las estrategias e instrumentos de evaluación.	Se hace mención a que se concretará en la programación didáctica anual las estrategias e instrumentos de evaluación, así como mecanismos para que el alumnado participe en el propio proceso de evaluación.
<b>Resultados / Escala de medida</b>	Cualitativa	Ordinal (SB, NT, BI, SU, IN)	Ordinal (SB, NT, BI, SU, IN) Cuantitativa (escala de 1 a 10)

En la Orden 45/2011 sobre la estructura de las programaciones didácticas en la enseñanza básica se hace referencia en el apartado 7 a la especificación de la evaluación. Así, se deben definir:

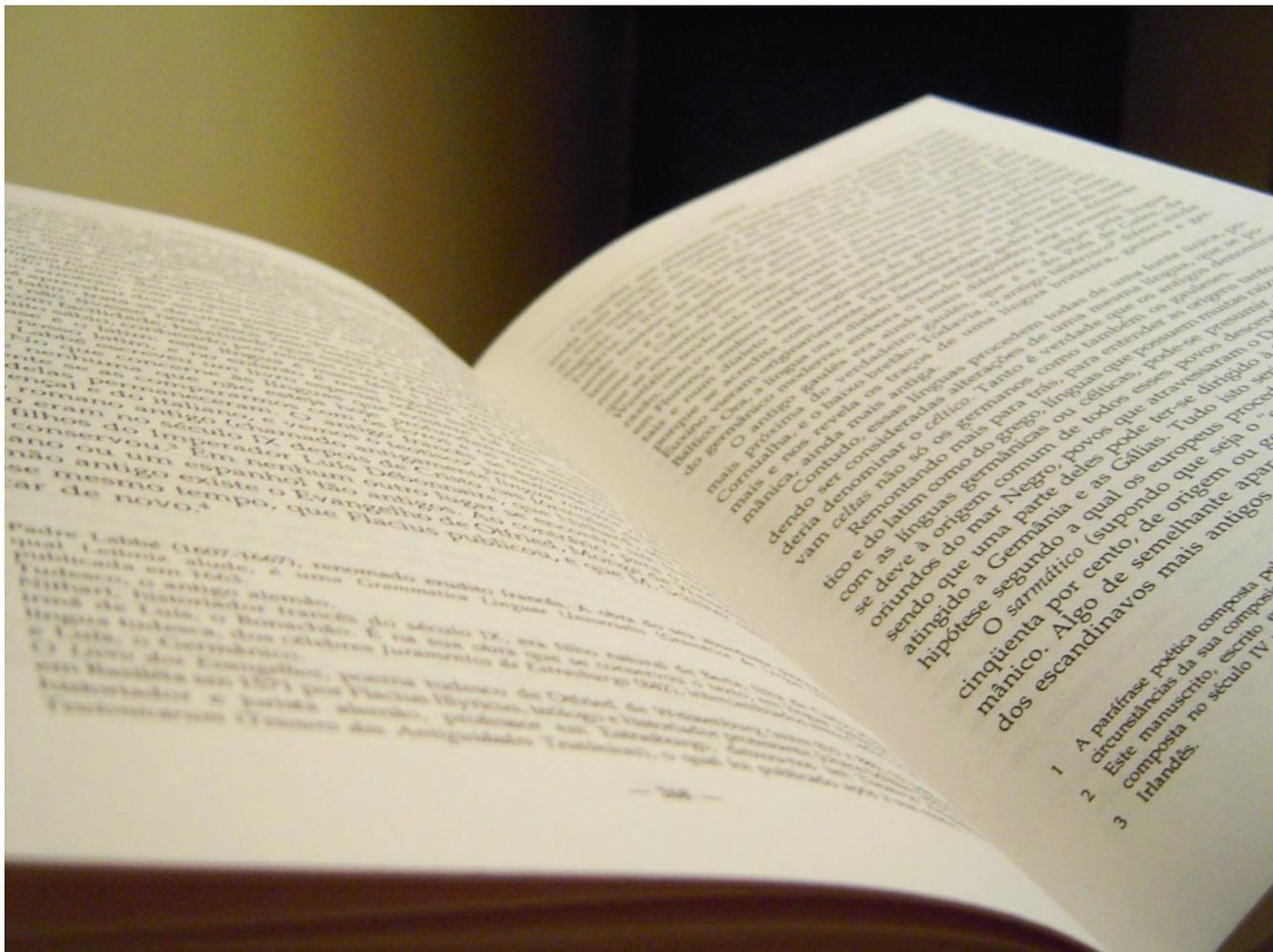
- a) criterios de evaluación
- b) instrumentos de evaluación
- c) tipos de evaluación
- d) criterios de calificación

y otra serie de puntos que no se refieren específicamente a la evaluación en general o a la del alumnado específicamente:

- e) actividades de refuerzo y ampliación,
- f) evaluación del proceso de enseñanza y aprendizaje.

Assí, la mesma normativa indica que desde la administração se proporcionarán criterios e indicaciones orientativas para elaborar modelos abiertos de programaciones didácticas. Sin embargo, todavía no se disponen de ese tipo de indicaciones ni en la normativa se indica qué se entiende por criterio de evaluación (si se refiere a algo distinto a los criterios de evaluación establecidos en el currículo), instrumentos de evaluación, tipos de evaluación y criterios de calificación.

También es interesante tener como referencia la Orden 32/2011 por la que se regula el derecho del alumnado a la objetividad en la evaluación, y se establece el procedimiento de reclamación de calificaciones obtenidas y de las decisiones de promoción, de certificación o de obtención del título académico que corresponda.



# 3. Procedimientos, pruebas e instrumentos de evaluación y escalas de medida

## 1. Procedimiento de evaluación

Los procedimientos para la evaluación son el conjunto de estrategias y métodos centrados en las acciones que ha de realizar el evaluador para someter a los estudiantes a una prueba y así valorar mediante un instrumento el grado de consecución de sus aprendizajes.

- Método observacional (tareas, ejecuciones, productos).
- Pruebas escritas.
- Métodos de encuesta.
- Métodos de entrevista.

En educación, habitualmente usamos los dos primeros.

## 2. Pruebas de evaluación

Las pruebas de evaluación son retos o situaciones en las que se demanda a un sujeto (alumno, en educación) que realice una acción o serie de acciones con el objeto de evaluar sus capacidades, destrezas, conceptos, habilidades, actitudes, etc. Así, las pruebas pueden ser de realización escrita (contestar a preguntas en formato lápiz y papel), pruebas con ordenador, oral, física, de destreza manual, de elaboración de un producto o material o de ejecución de una tarea o procedimiento donde se combinen cualquiera de las anteriores.

Las condiciones en las que se realiza la prueba adecuadas a las características de los probandos, los procedimientos a utilizar, los criterios de evaluación, los resultados de aprendizaje, así como los instrumentos que van a permitir la valoración de estos aprendizajes, son elementos que han de estar alineados con la misma prueba y a su vez con los modelos y metodología docentes.

El foco entonces está en la tarea que debe hacer un sujeto que va a ser evaluado y que, desde el punto de vista de una evaluación auténtica puede que no esté alejada de las tareas de aprendizaje que realiza el alumno habitualmente (piénsese en el propio cuaderno de ejercicios del alumno, como prueba de un aprendizaje determinado).

## 3. Instrumentos de evaluación (calificación o valoración)

Son herramientas, entendidas como objetos, aparatos, que permiten establecer una valoración sobre el grado de consecución de los aprendizajes. Esa valoración, puede tener un carácter cualitativo o cuantitativo. La tradición en educación hace que lo cuantitativo tenga mucha peso en la evaluación educativa, de forma que la mayor parte de aprendizajes son cuantificados y posteriormente calificados cualitativamente en un nivel de ejecución (p.ej.: el 6 de una escala de 1 a 10 se califica generalmente con un "bien"). Es decir que en general, se establece una medida, aunque no siempre.

Como instrumentos de evaluación que permiten valorar el grado de consecución de los aprendizajes podemos tener:

- **Pruebas objetivas**, el término prueba aquí es ambiguo ya que remite a prueba escrita, aunque no todas las pruebas escritas son instrumentos. En este caso las pruebas objetivas (estandarizadas o no) constituyen un instrumento a partir del cual se obtiene una puntuación que valora el grado de consecución de los aprendizajes,

siempre y cuando estén bien diseñadas. Estos instrumentos están conformados por preguntas vinculadas a procesos cognitivos y contenido, y en general al dominio cognitivo (los ítems son las pruebas en sí, entendidas como se está definiendo en este documento).

- **Escala de actitudes**, con su gran variedad de formatos (Thurstone, Likert, Gutman, diferencial semántico,...), que generalmente es escrita y está conformada también por preguntas o reactivos (p.ej.: adjetivos a valorar por el alumno en función de cómo se autopercibe). Son apropiados para medir actitudes.
- **Cuestionarios**. Los cuestionarios están conformados por preguntas (que se presentan en forma oral o escrita). Generalmente versan sobre algún tema del dominio cognitivo, pero más que recoger aspectos sobre rendimiento se utilizan para contrastar estados de opinión.
- **Listas de cotejo**.
- **Escalas de observación**.
- **Rúbricas** o escalas de valoración. Suelen clasificarse en holísticas o analíticas, aunque estas últimas se pueden confundir con las escalas de cotejo (*check list*).
- **Instrumentos** para la medición de variables físicas (cronómetro, pulsómetro, báscula, cinta métrica,...)

#### 4. Escala de medida

Muchos de los instrumentos de evaluación, especialmente aquellos que cuantifican el nivel de consecución de los aprendizajes, conllevan una escala de medida implícita. Piénsese en una regla (instrumento), que tiene una graduación que va de 0 a 1000 cm (escala de medida con unidades de medida en mm). Las escalas de medida pueden ser distintas para instrumentos iguales, (p. ej.: una regla que mida en pulgadas). De los anteriores instrumentos los susceptibles de tener escalas de medida asociada son las pruebas objetivas, las escalas de actitudes y las rúbricas, y por supuesto los instrumentos para la medición de variables físicas. En algunos casos, como las pruebas objetivas o escalas de actitudes, la unidad mínima de medida es el ítem, preguntas o cuestiones.

#### Ejemplo de definición de un procedimiento, prueba, instrumento y escala

Unidad didáctica: La Prehistoria - 1º ESO

Indicadores/resultados de aprendizaje (vinculados a criterios de evaluación del Decreto 112/2007)

- Elabora en grupo exposiciones orales.
- Utiliza el vocabulario pertinente.
- Utiliza la corrección formal adecuada.
- Elaborar, individualmente o en grupo, trabajos y exposiciones orales sobre temas de la materia, utilizando el vocabulario pertinente y corrección formal adecuada.

Análisis de los indicadores

Nivel de ejecución (verbo de acción)

elaborar, utilizar, elaborar

### Contenido

- exposición oral (modo o contenido?)
- vocabulario
- corrección formal (modo o contenido?)
- temas de la materia

### Naturaleza de la acción

- pertinente (vocabulario)
- adecuado (corrección formal)

### Criterio de evaluación

Definición del tema: El criterio debe especificar el tema sobre el que debe centrarse el alumno.

Definición de los modos: exposición oral, con la delimitación de tiempo, si tiene que utilizar o no material de apoyo (p.ej.: audiovisual), definir los elementos que se refieren a la corrección formal de una presentación (introducción del tema, desarrollo cronológico, conclusiones, uso de conectores, tono de la voz,...), vocabulario adecuado (definir qué se entiende por vocabulario adecuado, palabras clave, conceptos,...)

### Procedimiento/Técnica

Observación

### Prueba

Exposición oral

### Instrumento de evaluación

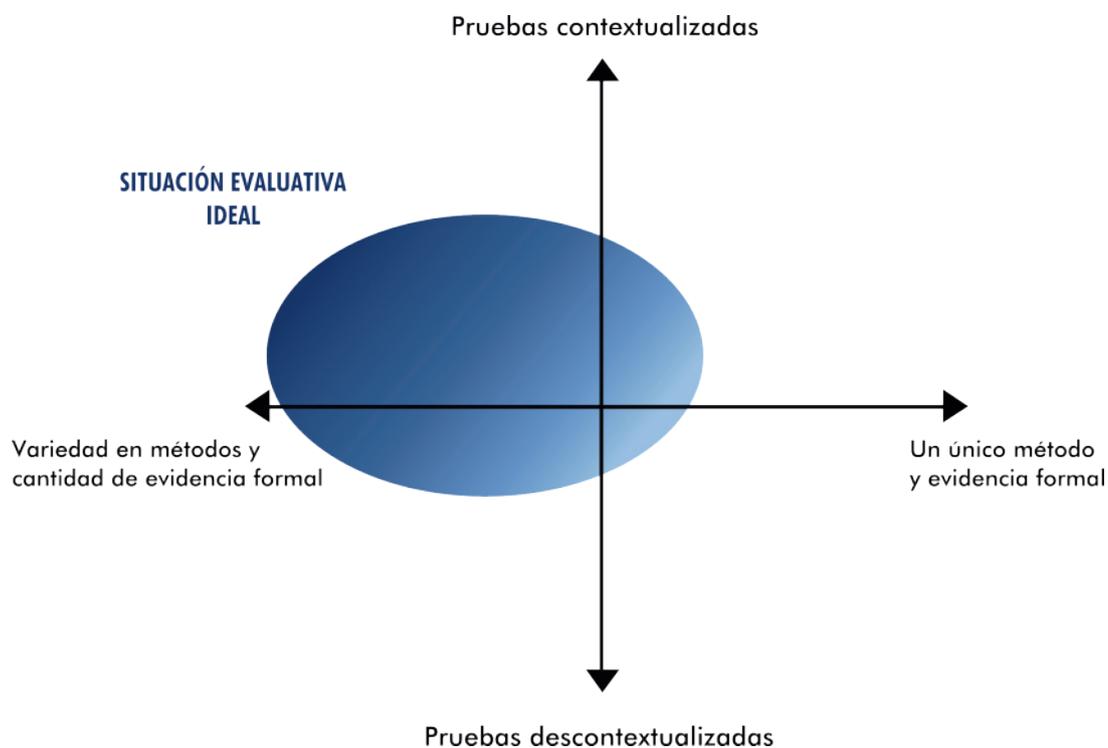
Rúbrica (escala de valoración holística).

### Escala de medida

De 1 a 4. 1: Inadecuado 2: Necesita mejorar 3: Adecuado 4: Excelente.

## Mestizaje en instrumentos y estrategias

A considerar, para Wiggins (1998) se trataría de buscar, como mejor alternativa para una evaluación auténtica, una evaluación equilibrada, en la que se den cita la máxima variedad de métodos y cantidad de evidencia formal de los aprendizajes.



## ¿Qué es medir?. La idea del muestreo y estimación.

S.S. Stevens, definió la medición en psicología, como la asignación de numerales a objetos o eventos de acuerdo a una regla. Y propuso cuatro escalas de medida: nominal, ordinal, intervalo y razón.

Hasta aquí parece clara, pero cuando entramos en el campo de la medida en psicología o educación, por ejemplo en nuestro caso y especialmente en educación secundaria, ¿tenemos clara la regla? El problema no parece residir en el hecho de asignar un numeral al aprendizaje del alumnado en una determinada materia o área, sino en qué significa ese numeral y si un mismo numeral en distintos alumnos significa exactamente lo mismo (dos alumnos que obtienen un 7 en una prueba de matemáticas, ¿han alcanzado el mismo nivel de aprendizaje?). En definitiva, cuando se desconocen las propiedades de las puntuaciones (lo que significan), lo apropiadas o no que sean las conclusiones también es un hecho desconocido.

Esto supone de inicio un serio problema a la hora de tomar las puntuaciones como algo taxativo para "cuantificar" el aprendizaje, más aún si se compara entre distintos profesores y más aún entre distintos centros. Existen en psicometría métodos para salvar este problema, como el modelo de Rasch o en general los modelos de la Teoría de Respuesta al Ítem, no sólo aplicables a ítems de pruebas objetivas, si no también a escalas presentadas en rúbricas. Sin embargo, estas teorías y técnicas psicométricas asociadas escapan al campo de la práctica docente.

¿Qué alternativa queda para la práctica de aula? Dotar de significado a esas puntuaciones confirmando validez a nuestras pruebas, en definitiva alinear la escala con el aprendizaje. De manera que un 7 de calificación signifique, aproximadamente, que todo el alumnado que obtenga esa calificación tenga adquiridas más o menos las mismas competencias del objeto de aprendizaje y que además sean más que un 5 y menos que un 10, por poner un ejemplo.

### Validez y fiabilidad de las pruebas

Además de las cuestiones inherentes a las escalas de medición también existen otras inherentes a las pruebas: su validez y fiabilidad.

Por **validez** entendemos aquella característica del instrumento de medición que hace que midamos lo que pretendemos medir.

Se consigue con algunas estrategias, y no es una característica absoluta, se pueden obtener evidencias sobre muchos tipos de validez. Las más comunes: criterial, de constructo o de contenido, pero también hay otras como la aparente, convergente, divergente, predictiva, ecológica...

De especial interés en la práctica de clase es la de contenido, sobre todo por la accesibilidad y la no necesidad de manejar datos. Para ello se puede utilizar una tabla de especificaciones (véase glosario). Otro tipo de estrategias pasan por la revisión de las preguntas realizadas de acuerdo a una serie de criterios (congruencia de la pregunta con el indicador/resultado de aprendizaje, sesgos de lenguaje, análisis de alternativas (en las preguntas de elección múltiple),...).

La **fiabilidad** tiene que ver con la estabilidad de la medida y su precisión. Es decir, un instrumento fiable es el que mide con precisión un constructo o una característica. Piénsese en una báscula; si por la mañana me peso tres veces seguidas, las tres veces habría de obtener más o menos el mismo peso.

La fiabilidad de una prueba es importante, pero para nuestro propósito docente, lo es más la validez. Además si un instrumento es válido podemos tener la seguridad de que es fiable, no a la inversa. Podemos medir la altura de una persona con una báscula muy fiable, pero desde luego no es un instrumento válido para ello.

Y la validez es una cuestión que podemos conseguir, como se ha dicho, con técnicas que están al alcance de cualquier docente. Fives y DiDonato-Barnes (2013) presentan las tablas de especificaciones como una estrategia que confiere validez a una prueba y permite

diseñarla de acuerdo a unos parámetros sencillos, además de servir también para diseñar la instrucción. Es una estrategia que se utiliza en toda evaluación a gran escala como PISA, TIMSS o PIRLS y también en las evaluaciones de diagnóstico de la Comunitat Valenciana. Es de uso habitual en el diseño de pruebas objetivas, pero también puede servir para otro tipo de instrumentos como las rúbricas.

Véase el siguiente ejemplo de tabla de especificaciones de la Evaluación General de Diagnóstico 2009 del Ministerio de Educación para la Competencia Matemática de Educación Primaria.

Matriz de especificaciones de Educación Primaria

4º Educación Primaria							
Bloques de contenidos	Procesos						Total
	Reproducción		Conexión		Reflexión		
	Acceso e identificación	Comprensión	Aplicación	Análisis y valoración	Síntesis y creación	Juicio y regulación	
Números y operaciones							35%
La medida							20%
Geometría							25%
Tratamiento de la información, azar y probabilidad							20%
<b>Total</b>	<b>10%</b>	<b>15%</b>	<b>25%</b>	<b>20%</b>	<b>20%</b>	<b>10%</b>	<b>100%</b>

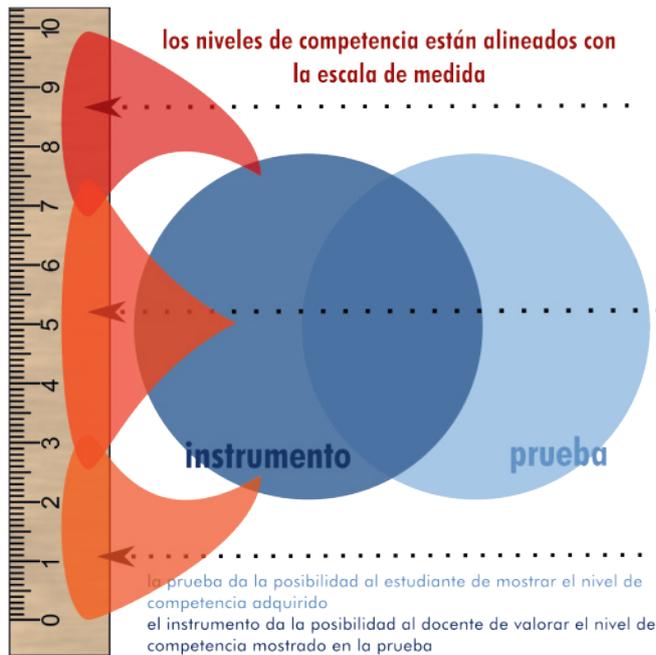
Con una **tabla de especificaciones** nos aseguramos equilibrar la prueba según un criterio que me permite establecer pesos (horas de instrucción de determinados bloques de contenido, o la dificultad que quiera conferir según procesos, generalmente).

Si además los ítems, preguntas o niveles de la rúbrica están bien diseñados y alineados con los indicadores/descriptores o resultados de aprendizaje, podemos prácticamente tener la seguridad de que dos 7 obtenidos por distintos alumnos, significan más o menos lo mismo, y además nos da una idea de qué aspectos de la competencia han sido alcanzados y cuáles no.

Así un alumno que tiene un 7, no habrá contestado o no habrá ejecutado o la habrá hecho incorrectamente, ítems o habilidades de procesos superiores; y mucho más será así para alumnos que hayan obtenido un 5.

Es decir, se ha alineado la escala de medida con el nivel de competencia alcanzado por el alumno. Este es el sentido de la evaluación referida al criterio. La puntuación obtenida significa *algo* en el dominio de aprendizaje.

## ESCALA DE MEDIDA



## NIVELES DE COMPETENCIA

### Nivel III

resultados de aprendizaje o indicadores que tienen que ver con procesos cognitivos de exigencia alta: evaluar y crear

### Nivel II

resultados de aprendizaje o indicadores que tienen que ver con procesos cognitivos de exigencia media: aplicar y analizar

### Nivel I

resultados de aprendizaje o indicadores que tienen que ver con procesos cognitivos de baja exigencia: recordar y comprender

## La idea del muestreo y la estimación

Se ha de tener en cuenta, que nunca se puede conocer con exactitud lo que un alumno ha aprendido. Únicamente tenemos instrumentos de evaluación que nos permiten realizar mediciones de ese aprendizaje y generalizarlas, con un margen de error, para posteriormente tomar decisiones.

Es justo la idea del muestreo y la estimación en el campo de la estadística, y sobre esto se va a hacer el símil en el campo de la evaluación de los aprendizajes.

### Muestreo

**Universo de medida**

**Experimentos**

**Tamaño de la muestra**

**Error muestral**

**Varianza**

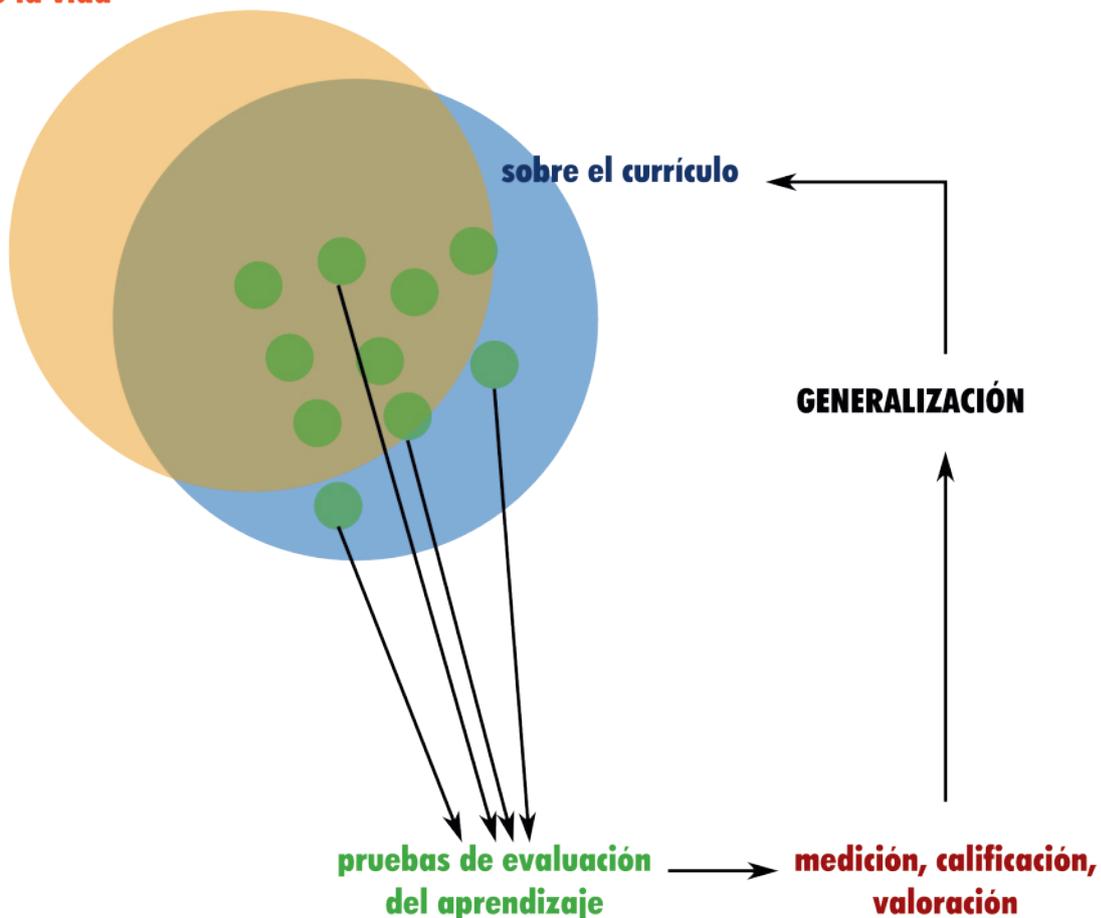
**Representatividad**

### Evaluación Educativa

- Saberes del alumno (formales, no formales, informales). Dominio de aprendizaje.
- Número de pruebas de evaluación realizadas; a mayor número de experimentos mejor vamos a poder estimar el verdadero valor del aprendizaje.
- Cantidad de evidencia de aprendizaje que abarcamos en una prueba (preguntas de todo el dominio de aprendizaje en una prueba escrita; condiciones suficientes para exhibir las habilidades orales en una exposición oral,...)
- Error que se asume al generalizar una calificación en una prueba al conjunto de saberes del alumno. Tiene relación inversa con la cantidad de evidencias de aprendizaje evaluadas. Pocas preguntas o condiciones para la exhibición (tiempo por ejemplo), mayor error, o viceversa.
- Característica de los aprendizajes evaluados. Existen aprendizajes más homogéneos (con dominios muy concretos o específicos) y otros más heterogéneos (con dominios muy variopintos y complejos). A complejidades mayores del dominio o dominios amplios necesitaremos más evidencias, más muestra.
- Se suele confundir con el tamaño de la muestra, sobre todo por los miembros de muchos tribunales en los Trabajos Fin de Máster. Un tamaño de la muestra pequeño no tiene por qué no ser representativo. La representatividad tiene que ver con el método de selección de la muestra, de las preguntas, en el dominio de aprendizaje, ¿hemos incluido preguntas de todo tipo de contenido y nivel cognitivo? A eso se refiere la representatividad, aunque haya pocas preguntas...

# saberes del alumno

sobre la vida



En la gráfica se resume el proceso de muestreo. ¿Qué cuestiones esenciales se deben entonces asumir en sucesivas evaluaciones del aprendizaje de los alumnos?:

1. Siempre hay un error de muestreo con el que vamos a contar a la hora de generalizar lo que un alumno sabe sobre el dominio de aprendizaje, aunque puede minimizarse.
2. El error de muestreo se minimiza aumentando el número de evidencias (preguntas, tiempo de ejecución,...). Lógicamente tiene que haber un equilibrio, no se pueden hacer pruebas muy largas.
3. En dominios donde los aprendizajes a evaluar sean complejos o amplios (varianza grande) se necesitarán más pruebas, de lo contrario aumentamos el error de muestreo.
4. Las pruebas han de ser representativas del dominio de aprendizaje (universo), se tiene que preguntar o someter a prueba los distintos tipos de contenido y procesos cognitivos por los que están conformados los aprendizajes del dominio. Con ello hacemos que la evaluación sea más válida.
5. Finalmente, el dominio de saberes sobre la vida (otro tipo de aprendizajes no dados en situaciones formales, como en la escuela) va a estar más evaluado conforme los saberes del currículo se solapan más con situaciones de la vida de los estudiantes. Es decir, conforme el currículo se centre más en competencias que en saberes descontextualizados.

# 4. Fases en la evaluación referida al criterio para clase.

## 1. Análisis del dominio de aprendizaje a evaluar.

Se trata de establecer estrategias para determinar a partir de los resultados de aprendizaje (criterios de evaluación o indicadores en el contexto de la legislación española), qué contenidos y qué procesos cognitivos (o destrezas psicomotoras, o niveles afectivos) forman parte de la evaluación. Es decir, analizar el dominio de aprendizaje que va a ser evaluado.

Una estrategia útil son las **tablas de especificaciones**, que además ayudarán a determinar el peso de determinados contenidos y procesos.

## 2. Especificación de criterios de evaluación, prueba, instrumento/s y escala/s asociadas a estos.

Los resultados de aprendizaje pueden guiar a determinar el contexto de la evaluación, los criterios para evaluar, los instrumentos y sus escalas de medida asociadas (si las hay). Véase el ejemplo de la p. 18 de este manual.

## 3. Elaboración de los instrumentos y determinación de escalas

Es hora de pasar manos a la obra y de elaborar los instrumentos de evaluación y sus escalas. En cada uno de estos instrumentos se requiere de una serie de técnicas que se irán desarrollando en una sucesivos cuadernos prácticos asociados a este manual.

- Técnicas para la elaboración de rúbricas y sus escalas de medida.
- Técnicas para la elaboración de pruebas objetivas (estímulos e ítems).
- Técnicas para la elaboración de cuestiones en cuestionarios.
- Técnicas para elaboración de categorías de observación para las escalas de observación.
- Técnicas para la elaboración de escalas de actitudes y sus ítems.

## 4. Revisión de los instrumentos.

Antes de comenzar cualquier prueba de evaluación y del uso de los instrumentos es conveniente probarlos. Se puede hacer algún ensayo o pedir a un compañero que los pruebe. Además, también es conveniente releerlos tras un período, seguro que se encuentran formas de redactar preguntas o de expresar rúbricas que parecen menos claras de los que en un principio parecían...Es hora de depurar el instrumento.

## 5. Realización de la prueba y evaluación del aprendizaje.

Es recomendable ser escrupuloso en el establecimiento de las condiciones de prueba: todos los alumnos han de cumplir las mismas condiciones, salvo las adaptaciones pertinentes establecidas para determinados alumnos. Las instrucciones que se proporcionan a cada alumno en una prueba determinada han de ser siempre las mismas, si es necesario se redactatan para ayudar a mantener el "guión" establecido.

Los criterios por los cuáles un alumno va a ser evaluado deben ser claros y públicos.



## 6. Información y toma de decisiones.

La última parte, pero no la menos importante. Los resultados de cualquier evaluación han de permitir que podamos tomar decisiones sobre cómo facilitar el proceso de aprendizaje del alumnado, y cómo modificar nuestra práctica docente, además de monitorizar la evolución del proceso (carácter formativo). En otros casos nos llevará a tomar decisiones que afectan a la promoción o certificación (carácter sumativo). Y por supuesto de todo ello se ha de informar a las partes interesadas (además de que por normativa así está establecido).



# Glosario

## Criterios de evaluación

*Cuestiones objetivas, especificaciones, requerimientos o estándares que actúan como referencia para valorar o juzgar individuos, organizaciones o cosas.*

## Estándares académicos

*Reglas, principios o criterios establecidos por una institución educativa mediante los que se determinan niveles o grados de logro de los estudiantes, generalmente en términos de la medición de la adecuación, aceptabilidad, cantidad, calidad o valor.*

## Evaluación:

*Valorar o juzgar personas, organizaciones o cosas en relación con objetivos, estándares o criterios establecidos.*

## Evaluación alternativa

*Estimación del conocimiento o las habilidades por otros medios distintos a los empleados tradicionalmente, pruebas objetivas, especialmente las pruebas de elección múltiple estandarizadas.*

## Evaluación educativa:

*Determinar e interpretar el logro de objetivos educativos (en el ámbito nacional, regional o local) para su uso en la planificación educativa y su desarrollo, la política de formación y la dotación de recursos.*

## Objetivos (educativos)

*Propósitos o fines hacia los que se dirige un esfuerzo, relacionados con los resultados o la organización del proceso educativo – se incluyen los establecidos por autoridades educativas en todos los niveles.*

## Prueba

*Las pruebas de evaluación son retos o situaciones en las que se demanda a un sujeto (alumno, en educación) que realice una acción o serie de acciones con el objeto de evaluar sus capacidades, destrezas, conceptos, habilidades, actitudes, etc.*

## Resultados de aprendizaje

*Declaraciones que describen aprendizajes significativos y esenciales que los estudiantes han alcanzado y pueden demostrar al finalizar un curso o programa.*

## Rúbrica

*Esquemas descriptivos utilizados para puntuar (calificar) que guían el análisis de la ejecución de procesos o productos de los estudiantes.*

## Tabla de especificaciones

*Estrategia para el diseño de pruebas de evaluación que consiste en una tabla de doble entrada, generalmente, donde se especifica el contenido y los procesos, así como el peso que van a tener en la prueba (usualmente en tanto por cien). Los resultados de aprendizaje, descriptores o indicadores se clasifican en la tabla de acuerdo a estas dos dimensiones.*

# Lecturas recomendadas

**Brookhart, S.M.** (1999). *The Art and Science of Classroom Assessment: The Missing Part of Pedagogy*. ASHE-ERIC Higher Education Report (Vol. 27, No. 1). Washington, DC: The George Washington University. Graduate School of Education and Human Development.

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

**De Miguel Díaz y otros** (2006). *Metodologías de enseñanza y aprendizaje para el desarrollo de competencias*. Madrid: Alianza Editorial.

**ERIC Thesaurus On-line.**

**Fives, H. & DiDonato-Barnes, N.** (2013). Classroom Test Construction: The Power of a Table of Specifications. *Practical Assessment, Research and Evaluation*, 18(3). Disponible online: <http://pareonline.net/getvn.asp?v=18&n=3>

**Frey, B.B., Schmitt, V.L. and Justin, P.A.** (2012). Defining Authentic Classroom Assessment. *Practical Assessment, Research and Evaluation*, 17(2). Disponible online: <http://pareonline.net/getvn.asp?v=17&n=2>

**Martín E. y Martínez F.** *Evaluación de los aprendizajes en el aula: una conceptualización renovada*. Ed. Santillana

**Morales Vallejo, P., Urosa Sanz, B. y Blanco Blanco, A.** (2003). *Construcción de Escalas de Actitudes tipo Likert*. Madrid: La Muralla.

**Padilla, M.T.** (2002). *Técnicas e instrumentos para el diagnóstico y la evaluación educativa*. Madrid: Editorial CCS.

**Rudner, L.M. y Schafer, W.D.** (Eds.) (2002). *What Teachers Need to Know about Assessment*. Washington, DC: National Education Association.

**Shelley, K. y Howell, S.L.** (2008). *The State of Authentic Assessment*. ERIC Report #ED503679.

**Shrock, S. y Coscarelli, W.** (2000). *Criterion-Referenced test Development: technical and legal guidelines for corporate training and certification*. (2nd Ed.). Silver Spring (MD): International Society for Performance Improvement.

**Wiggins, G.** (1998). *Educative Assessment. Designing Assessments To Inform and Improve Student Performance*. San Francisco (CA): Jossey-Bass Publishers.

