

# DISEÑO DE RÚBRICAS

*Instrumentos de evaluación educativa #02*

Javier Cortés De las Heras



Diseño de rúbricas. Cuaderno de instrumentos de evaluación #2

Autor: Javier Cortés De las Heras (e-mail: [cortes\\_jav@qva.es](mailto:cortes_jav@qva.es))

Diseño y maquetación: Perro-Ballena Productions

2014

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. Introducción .....</b>	<b>4</b>
<b>1. La evaluación de ejecuciones .....</b>	<b>5</b>
<b>2. Diseño de rúbricas.....</b>	<b>11</b>
<b>3. Revisión de la rúbrica.....</b>	<b>20</b>
<b>Bibliografía .....</b>	<b>22</b>

# 0. Introducción

En este segundo cuaderno de instrumentos de evaluación educativa se abordarán los pasos necesarios para diseñar rúbricas como instrumento privilegiado en la evaluación de ejecuciones o del desempeño. Así, como para el caso del diseño de *cuestionarios de rendimiento* (también llamadas pruebas objetivas), que constituye el cuaderno de instrumentos de evaluación #1, los contenidos que se van a tratar tienen un carácter eminentemente práctico que, esperamos, sirvan de guía en la construcción de los instrumentos que van a ser utilizados en la evaluación de los aprendizajes en el aula.

Como se verá, los ejemplos que van a ir apareciendo se basan en el currículo oficial de la Comunitat Valenciana: Decreto 108/2014, para Educación Primaria y el todavía vigente Decreto 112/2007, para Educación Secundaria Obligatoria. Pero fundamentalmente nos centraremos más en el de Educación Primaria, por tener un diseño más útil para el tema de fondo que nos ocupa, la evaluación referida al criterio.

Desde esta premisa, y recordando lo dicho en el documento *Guía de evaluación educativa para el profesorado de educación primaria y secundaria* (Cortés de las Heras, 2014), nuestros elementos de trabajo son los **criterios de evaluación** y en la medida de lo posible, los **indicadores de logro** que se hayan establecido a partir de éstos o, como en otros contextos de evaluación se denominan, descriptores. Siempre y cuando estos criterios de evaluación estén redactados en forma de **resultados de aprendizaje** (lo que se espera que el estudiante alcance a conocer al finalizar un período formativo) con su sintaxis completa: proceso, contenido y naturaleza de la ejecución, estaremos en condiciones óptimas de diseñar nuestros instrumentos de evaluación. Ellos serán los que nos ayuden a determinar qué tipo de pruebas e instrumentos son los más idóneos para evaluar los aprendizajes de nuestro alumnado y, para el caso de las rúbricas, serán aquellos que tengan que ver con desempeños complejos o productos. De aquí que comenzaremos este cuaderno *práctico* con algo de teoría, describiendo qué se entiende por **evaluación de ejecuciones**.

Una vez descritas las principales características y pasos que se siguen en la evaluación de ejecuciones, el cuaderno se centra en las rúbricas:

- Su concepto,
- tipos,
- los pasos para su diseño, desde la determinación de los criterios de realización y atributos hasta la elección de una escala y,
- los criterios de calidad de una rúbrica, que tienen que ver con su validez y fiabilidad.

# 1. La evaluación de ejecuciones

En la guía de evaluación anteriormente mencionada (Cortés de las Heras, 2014) se dieron unas pinceladas sobre el concepto de **evaluación auténtica**, corriente evaluativa que ha surgido como rechazo a la evaluación tradicional, centrada en el desarrollo de pruebas objetivas que, aunque técnicamente impecables, poco o nada han ayudado en el diseño de intervenciones educativas. En la evaluación auténtica cobraban fuerza tres grandes aspectos: a) el planteamiento de tareas complejas en contextos reales o simulados, b) su carácter formativo y c) el uso amplio de indicadores y además compartido con los estudiantes.

Así para Wiggins (1998, p. 22 y ss.):

*Una evaluación es auténtica cuando la anclamos en el tipo de trabajo real que la gente hace. Es auténtica porque vemos si los estudiantes pueden usar inteligentemente lo que han aprendido en situaciones que paulatinamente se aproximan a situaciones adultas y si pueden presentar innovaciones ante nuevas situaciones.*

Un término que suele asimilarse al de evaluación auténtica en muchos contextos es el de **evaluación de ejecuciones** o del desempeño (del inglés: *performance assessment*) y es porque en este enfoque se concentran una variedad de estrategias evaluativas que pretenden emular contextos reales, uno de los requisitos de la evaluación auténtica. Sin embargo, no necesariamente ha de compartir su carácter formativo o compartir con los estudiantes los criterios de realización. Es decir, la evaluación de ejecuciones es un requisito indispensable de la evaluación auténtica, pero no suficiente para que ésta se dé. Piénsese en el acceso a un puesto de trabajo de comercial donde se evalúe la capacidad de argumentación delante de una audiencia, en este caso no es necesario compartir con los candidatos los criterios de puntuación y desde luego no tiene un carácter formativo, pero no por ello deja de ser una evaluación de un desempeño en un contexto real (o simulado).

Por tanto, la evaluación de ejecuciones es idónea cuando aquello que se quiere evaluar son contenidos procedimentales, habilidades de alto nivel, hábitos sociales, trabajo cooperativo, conductas en la toma de decisiones e incluso productos complejos. También, cuando el objetivo es obtener información relevante sobre el proceso o razonamiento utilizado en la resolución de determinadas tareas. Asimismo, es un enfoque que puede ser muy válido en contextos donde no hay tiempos o condiciones determinadas y que incluyan incluso aspectos de la vida real cuya concreción puede no estar establecida formalmente en el currículo oficial.

Las posibles preguntas que ahora podríamos hacernos son:

1. **¿Es posible una evaluación del desempeño desde nuestras aulas?** No solo es posible, si no que ya se da en algunos casos. Otra cuestión será que dispongamos de los instrumentos válidos y fiables para evaluar ese desempeño.
2. **¿Qué elementos del currículo permiten discernir cuándo es apropiada una evaluación de este estilo?** Aquellos elementos en los que se observe claramente que el aprendizaje es sobre habilidades complejas o de alto nivel cognitivo, de productos, etc. Y esto se puede observar, en nuestro caso, en los criterios de evaluación, y sus especificaciones en *indicadores de logro*, que servirán como base para el diseño de los instrumentos apropiados.

En cualquier caso, siempre partimos de una alineación de los elementos curriculares, tanto prescriptivos como no prescriptivos; y entre ellos, como parte central están los criterios de evaluación. Éstos guían también la instrucción y el diseño de ejercicios, actividades y tareas en las programaciones didácticas, pudiéndose aprovechar algunas de estas actividades como *pruebas de evaluación*, bien sea en situaciones formales o

informales. O diseñar específicamente situaciones de prueba para valorar la consecución de los aprendizajes fijados en uno o varios criterios de evaluación.

## Pasos en la evaluación de ejecuciones

Tal como ocurre en el diseño de pruebas objetivas (véase Cuaderno de evaluación educativa #01) o en el punto 4. *Fases en la evaluación referida al criterio para clase* del documento *Guía de evaluación educativa para el profesorado de educación primaria y secundaria*, en el contexto de la evaluación del desempeño se pueden seguir las mismas fases. Veamos la tabla que compara el esquema general con el de la evaluación de ejecuciones.







	Evaluación referida al criterio (proceso general)	Evaluación del desempeño
	1. Análisis del dominio de aprendizaje a evaluar.	1. Delimitación del propósito de la evaluación
	2. Especificación de criterios de evaluación, prueba, instrumentos y escalas.	2. Especificación de criterios para la evaluación del desempeño: prueba, instrumentos y escalas.
	3. Elaboración del instrumento.	3. Desarrollo de la rúbrica.
	4. Revisión del instrumento.	4. Revisión de la rúbrica.
	5. Realización de la prueba y evaluación del aprendizaje.	5. Realización de la prueba y evaluación del aprendizaje.
	6. Información y toma de decisiones.	6. Información y toma de decisiones.

Tabla 1: Fases en la evaluación referida al criterio vs. evaluación del desempeño.

### 1. Delimitación del propósito de la evaluación

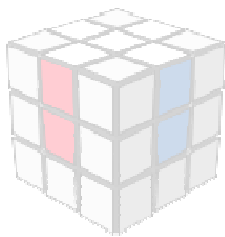
El punto central de esta primera fase es el análisis de lo que queremos evaluar. Es lo que en otros contextos se denomina análisis del dominio. Así que, deberíamos responder a una serie de preguntas, si antes no lo hemos hecho.

- ¿qué conceptos, destrezas o conocimientos en general se quiere evaluar?
- ¿qué deberían saber los alumnos sobre esto que quiero evaluar?
- ¿a qué nivel van a realizar los alumnos el desempeño?
- ¿qué tipo de razonamiento será evaluado: razonamiento, procesos, memorización, creación...?

Si tenemos la suerte de encontrarnos con un currículo donde se especifica qué es lo que los alumnos deben ser capaces de hacer al finalizar una determinada unidad didáctica, trimestre, curso..., es decir hay resultados de aprendizaje bien definidos, como puede ser el caso de **nuestro currículo** con los **criterios de evaluación**, ya tenemos el camino bastante andado.

Los criterios de evaluación, y sus posibles especificaciones (indicadores de logro, micro-criterios de evaluación, descriptores...) si están redactados en términos de resultado de aprendizaje encierran una información valiosísima, no solo para determinar las actividades de enseñanza, sino también las actividades de evaluación o pruebas y por extensión el tipo de instrumento más adecuado.

## Ejemplo de delimitación del dominio desde los criterios de evaluación (resultados de aprendizaje)



Unidad didáctica: El mundo en cifras – 6º de Ed. Primaria

Criterios de evaluación (resultados de aprendizaje):

– 6ºMAT BL5.1. Recoger y registrar informaciones cuantificables a partir de: encuestas, medios de comunicación, INE, internet,..., utilizando la media, la moda y el rango, para analizar información de interés humano o relacionada con su vida cotidiana.

Indicadores de logro:

– 6ºMAT BL5.1.1. Recoge y registra informaciones cuantificables a partir de: encuestas, medios de comunicación, INE, internet,...

– 6ºMAT BL5.1.2. Interpreta y analiza datos sobre informaciones cuantificables a partir de: encuestas, medios de comunicación, INE, internet,..., utilizando la media, la moda y el rango.

Análisis del dominio:

*Nivel de ejecución*

Recoger (observar)-registrar, analizar e interpretar

*Contenido*

Información cuantificable de interés humano

Moda, media, rango

*Naturaleza de la ejecución*

Encuestas, medios de comunicación, INE, Internet

A partir de este punto se podría ir pensando en el tipo de actividad de evaluación:

- Son acciones complejas que no se suelen dar a la vez, si no en un proceso a lo largo de un determinado tiempo.
  - **¿qué actividades de evaluación podría plantear?** Una exposición oral con presentación donde se den unas conclusiones, un informe escrito donde plasmen el proceso, una prueba escrita donde identifiquen datos y los analicen e interpreten...
  - **¿cuándo voy a evaluar el desempeño?** Un día, varios días, en clase, en una actividad extraescolar, desde mi casa con un producto que me entreguen...
  - **¿cómo y con qué voy a evaluar?** Evalúo solo yo, que se valoren entre ellos... utilizando algún instrumento dependiendo de la prueba de evaluación: prueba objetiva, ensayo de opinión, lista de cotejo, una rúbrica...
- Necesitan haber trabajado conceptos como la moda, media y rango
- Necesitan haber manejado fuentes de información analógica y digital (encuestas, medios de comunicación diferentes, INE, Internet...)

## 2. Especificación de criterios para la evaluación

Una vez analizado el dominio, se puede tener alguna pista más clara sobre qué tipo de prueba se quiere utilizar para evaluar el desempeño que está incluido en los criterios de evaluación. Se trata ahora de elegir esta prueba y de especificar aquellas condiciones que permitan delimitar las acciones que el evaluador (docente) debe tener en cuenta y aquellas condiciones de la prueba que se indicarán al alumnado. Es ideal que estas condiciones sean públicas y compartidas, incluso elaboradas o consensuadas entre el docente y el alumnado.

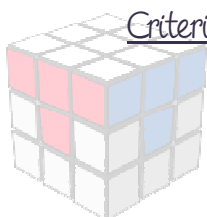
En resumen, se debe:

1. Elegir la prueba de evaluación.
2. Especificar si la situación de evaluación será formal o informal.
3. Especificar las condiciones de la prueba.

En cualquier caso se debe ser muy preciso a la hora de delimitar las condiciones de realización: han de ser iguales para todos, salvo adaptaciones necesarias; con los mismos materiales, tiempo de ejecución, recursos humanos, con las mismas instrucciones...

4. Definir los **criterios de realización**: se dedicará un apartado especial, ya que es una parte fundamental para el desarrollo de las rúbricas.
5. Elegir el tipo de rúbrica (las veremos más adelante) y escala de medida.

## Ejemplo de especificación de criterios para la evaluación



Criterios para la evaluación:

- Se acota el tema dentro de la UDI: p.ej. Indicadores de países del mundo (salud, economía, educación...)
- Se delimita la prueba: **Informe escrito**
- Y sus límites o condiciones: la estructura, extensión y contenido del informe (p.ej. el número de indicadores a registrar); agrupamiento: en grupo; fecha de entrega; materiales a proporcionar (páginas web); visita al Instituto Valenciano de Estadística; sesiones donde se podrá trabajar en clase y consultar al profesor...
- Instrumento: rúbrica (holística)
- Escala de medida: de 1 a 4 (Necesita mejorar, Adecuado, Bueno, Excelente)

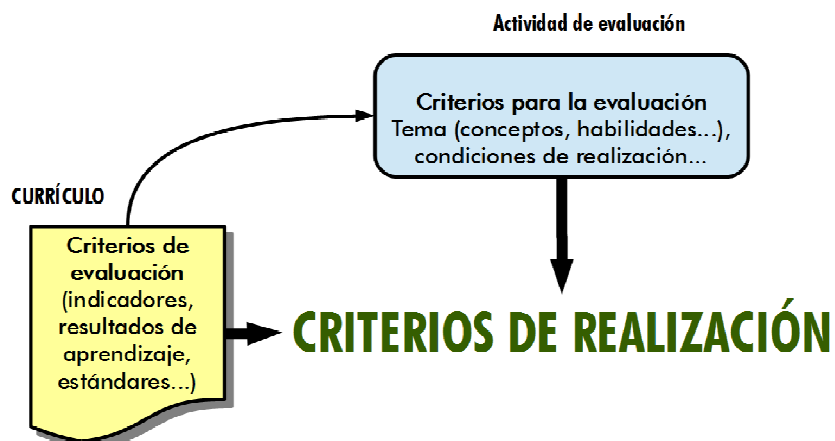
Como se ha dicho, estos criterios para la evaluación deberían ser públicos y consensuados, en la medida de lo posible, con el alumnado.

## Los criterios de realización

Los criterios de realización son el paso previo al desarrollo de rúbricas y nacen de:

- a) los límites impuestos por los criterios de evaluación (aquello que se debe evaluar de forma prescriptiva) y,
- b) Los límites impuestos por la prueba o actividad de evaluación, que a su vez estaba limitada por los criterios de evaluación.

Especifican aún más si cabe cuáles son los atributos del desempeño que van a ser evaluados, y resultan de una riqueza inestimable para el proceso de enseñanza y aprendizaje porque, siendo compartidos, proporcionan al alumnado un horizonte de desempeño.





Para el desarrollo de los criterios de realización se recomiendan una serie de pasos, que por otra parte resultan bastante lógicos:

### **1. Identificar globalmente la destreza o producto.**

Quizá este paso ya lo hayamos realizado en la especificación de los criterios para la evaluación. Se trataría de pensar en la ejecución de la actividad de evaluación que va a realizar nuestro alumnado (sea un producto o una destreza) de forma global. Qué pasos realizará un alumno tipo en el desempeño, con qué dificultades se encontrará, qué utilizará (recursos materiales), si es un producto, qué características tendrá...Posiblemente sea interesante que realicemos nosotros la actividad o, si tenemos experiencia, pensar también en esas otras veces en las que se ha encomendado esta actividad a otro tipo de alumnado.

### **2. Listar sus aspectos importantes.**

Una vez identificada la destreza o el producto, su desempeño, se trataría de listar aquellos aspectos, al menos aquellos que se alinean con los criterios de evaluación del currículo, fácilmente identificables, y otros más específicos que no se alejan de ellos respecto al proceso o procesos cognitivos implícitos, los contenidos y la naturaleza de la ejecución.

### **3. Limitar el número de criterios (observables).**

La lista anterior debe reducirse de manera que eliminemos aspectos redundantes y no observables. Es bueno quedarse con una lista que en cualquier caso no supere un número determinado de criterios. Se ha de pensar en que éstos son los que van a servir de base para elaborar la rúbrica y una rúbrica excesivamente analítica va a dificultar la evaluación.

### **4. Pensar sobre conductas importantes incluidas en la actividad de evaluación.**

Como prueba definitiva se someterá la lista limitada de criterios a una última revisión, donde se comprobará si son criterios que representan las conductas más importantes de la actividad que ha de realizar el alumnado y pueden describirse lo suficientemente como para poder evaluarlas de forma fiable. De nada serviría incluir, por ejemplo, una actitud de cooperación con los compañeros si no se dan las condiciones para observarla o simplemente no la podemos describir o es accesorio.

### **5. Expresar los criterios de realización en términos de conductas (alumnos) o características (producto).**

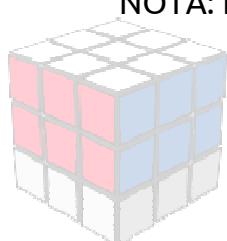
Por último, se hace necesario expresar los criterios de realización en términos de conductas (para el caso de destrezas) o de características (para el caso de productos). Estas descripciones son fundamentales para el desarrollo de la rúbrica y cada una de ellas vendrá identificada con un atributo o variable. Es la etiqueta que describe la característica principal del criterio de realización.

Además, si el desempeño sigue una secuencia temporal, podremos ordenar los criterios en el orden en el que van a ser observados. Esto facilitará el proceso de calificación del desempeño con la rúbrica, especialmente si se realiza a la par que el desempeño (p.ej.: una exposición oral sobre un tema).

Todos estos pasos que están expuestos linealmente no tienen por qué ocurrir en ese orden. De hecho es un proceso que requiere de revisiones continuas, ir hacia adelante y hacia atrás hasta conseguir el propósito deseado. De hacerlo bien, se está incidiendo en la construcción de un instrumento válido respecto al contenido a evaluar.

## **Ejemplo de descripción de criterios de realización y sus atributos**

NOTA: Los atributos van en negrita.



### **Búsqueda**

Consulta diversos medios fiables en la recogida de información.

**Relevancia**

Incluye datos de indicadores relevantes en educación, economía y salud de distintos países.

**Claridad**

Refleja con claridad los datos de los diferentes países pudiéndose identificar y asociar fácilmente a cada país al que pertenecen y distinguirse de otros indicadores.

**Análisis**

Analiza de forma cuantitativa el fenómeno estudiado a partir de los datos estableciendo comparaciones entre los diferentes países usando la media, la moda y el rango.

**Interpretación**

Establece una interpretación clara y congruente a partir de los datos obtenidos, dando una visión global a partir del análisis.

En los siguientes apartados se desarrollarán las fases 3. y 4. del proceso de evaluación del desempeño expuesto en la Tabla 1. Las fases 5. y 6. no difieren del uso de otros instrumentos en el marco de la evaluación referida al criterio y no serán tratadas en este documento.

## 2. Diseño de rúbricas

### Definición

El término rúbrica, en el contexto de la evaluación, proviene del inglés *scoring rubric*, que traducido literalmente es “rúbrica para puntuar”. Efectivamente, se tiene claro que es un instrumento de evaluación que sirve para puntuar o al menos para calificar al estudiante en el ejercicio de un desempeño...y quizá los angloparlantes que propusieron el término le añadieron este adjetivo (*scoring*) al sustantivo (*rubric*) para denotar un tipo especial de rúbrica.

Según el *Longman Dictionary of Contemporary English*, el término *rubric* es un conjunto de instrucciones o una explicación en un libro o examen, etc. Con el adjetivo, y tomado de manera literal, se está diciendo que es un conjunto de instrucciones que sirve para puntuar.

En el diccionario de la Real Academia de la Lengua Española, en su 23ª edición, rúbrica está definido como:

1. *Rasgo o conjunto de rasgos de forma determinada, que como parte de la firma pone cada cual después de su nombre o título, [...].*
2. *Epígrafe o rótulo.*
3. *Cada una de las reglas que enseñan la ejecución y práctica de las ceremonias y ritos de la Iglesia católica en los libros litúrgicos.*

Es decir, nada que ver con el concepto inglés moderno. Quizá porque aún no se haya incorporado plenamente su uso como ese conjunto de reglas o instrucciones que se utiliza en un examen. Así, las referencias remiten al uso en la firma (esos rallajos que hacemos encima de nuestro nombre, quien lo utilice, cuando firmamos algún documento) o a un uso más antiguo, que remite además a la etimología de la palabra, derivada del latín *ruber* (rojo)<sup>1</sup>, que era el color utilizado en los epígrafes o rótulos de los libros y probablemente en el conjunto de las reglas de la liturgia.

Dejando la cuestión estrictamente léxica y centrándose en la cuestión pedagógica, en distintos libros, artículos y otro tipo de documentos que tratan las rúbricas, suelen aparecer diferentes definiciones. Por ejemplo, la del glosario de términos de pruebas estandarizadas del *Educational Testing Service* (ETS), encargados de la evaluación de PISA últimamente (<https://www.ets.org/understanding-testing/glossary/>) que define rúbrica como “un conjunto de reglas para puntuar las respuestas a un ítem de respuesta construida<sup>2</sup>, a veces denominado guía de puntuación o valoración (*scoring guide*). Personalmente, creo que no es una definición de aplicación general por estar demasiado contextualizada en pruebas estandarizadas. Así, las “guías de valoración” utilizada en las pruebas PISA, por ejemplo, no son estrictamente una rúbrica. La rúbrica requiere de algunas características más específicas.

En otros documentos se encontrarán otras definiciones, aquí se propone una que tiene sentido con lo que se va a exponer seguidamente.

---

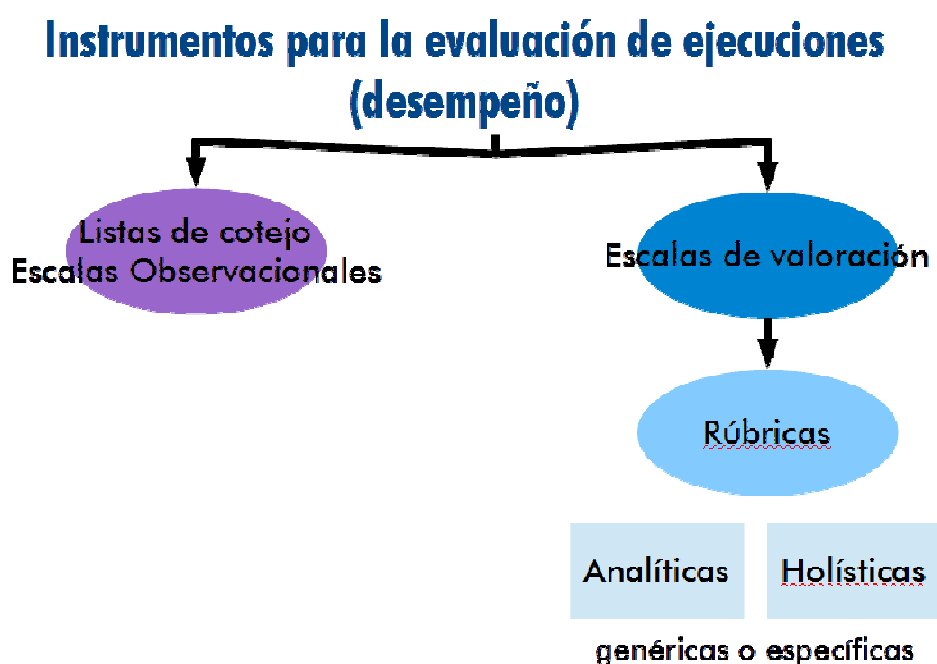
<sup>1</sup> Como curiosidad, Jose Antonio Millán en el primer capítulo de su libro *El candidato melancólico* hace mención a rúbrica y otras derivadas de la misma raíz: rubor, rubí, rubeola...

<sup>2</sup> Una pregunta de una prueba objetiva que requiere de una respuesta abierta, y generalmente larga. En el cuaderno de evaluación sobre pruebas objetivas aparecen este tipo de ítems.

## Rúbrica

*Instrumento de calificación que utiliza la descripción cualitativa de los criterios de realización de una tarea o actividad compleja a evaluar de forma graduada. Así, proporciona pautas con respecto a cómo se va a apreciar el trabajo de un estudiante.*

Como ya se ha dicho, la rúbrica es un excelente instrumento para la evaluación del desempeño, aunque no es el único, existen otras alternativas como la lista de cotejo o las escalas observacionales, quizá más apropiadas cuando la información a registrar se limita a la determinación de si se cumple o no con unos criterios específicos. En el esquema siguiente se hace una clasificación de los distintos tipos reconocidos para la evaluación del desempeño, y las rúbricas se sitúan como un tipo de escala o guía de valoración. La explicación sobre el tipo de rúbricas (analíticas u holísticas) se da más adelante.



## Ventajas y desventajas del uso de rúbricas

Las rúbricas nacen sin duda como una alternativa al uso de pruebas objetivas para la evaluación de aprendizajes que son complejos, que suponen interacción o el uso de destrezas cognitivas de alto nivel. Aprendizajes que no se pueden valorar como correctos o incorrectos.

Otra de las grandes ventajas es que su uso se puede extender a todos los niveles educativos, desde la Educación Infantil hasta la Educación Superior, y pueden ser utilizadas en diferentes materias y situaciones. Son tremendamente versátiles.

Son útiles en un enfoque de *evaluación para el aprendizaje* o evaluación con propósitos formativos: apoyan la exploración de hasta qué punto unos determinados criterios de realización se han alcanzado y permiten proporcionar retroalimentación a los estudiantes sobre cómo mejorar sus realizaciones. Es más, en su construcción, desde la definición de los criterios de realización, puede participar el alumnado con mayor o menor apoyo del docente según el nivel educativo en el que se encuentre.

Sin embargo, muchas de las rúbricas no son beneficiosas para el aprendizaje, principalmente por problemas de diseño: el proceso de construcción no se hace adecuadamente y puede haber inconsistencias en las descripciones de los criterios de realización en los diferentes niveles de la escala. Por ejemplo, las más comunes y accesibles en Internet contienen fallos de diseño y no sólo afectarían a la instrucción, sino también a la

validez de los resultados.

Por otra parte, no todo aprendizaje es evaluable con rúbricas, especialmente aquellos que se refieren a aspectos muy concretos del currículo, y que pueden ser más fácilmente evaluables con una lista de cotejo o incluso con una prueba objetiva. Por el hecho de “estar de moda” no deben desdeñarse otros instrumentos y no se debe forzar la máquina”, por ejemplo: no contar con el contexto de una actividad de evaluación y rubricar criterios de evaluación o intentar “rubricar” cada ítem de una prueba objetiva...Son ejemplos que suelen verse habitualmente y cada vez más.

## Tipos de rúbricas

Habitualmente se suele hacer una clasificación de las rúbricas en: a) holísticas o b) analíticas, atendiendo a cómo consideran la evaluación del desempeño, si como un todo o por partes respectivamente.

También existe otra clasificación atendiendo al tipo de desempeño que se evalúa, si es específico o es general, quedando así rúbricas que son específicas o rúbricas que son genéricas, y esto dependerá del tipo de criterios por los que se evaluará en ese desempeño. En esta clasificación no se profundizará demasiado, simplemente sirva decir que las **rúbricas genéricas** se dan cuando existen desempeños que reúnen características similares independientemente del contenido, que son interdisciplinarias, por ejemplo, la presentación de una información (siempre y cuando el contenido de esa presentación no sea objeto de evaluación). Por el contrario, las **rúbricas específicas** de la tarea, enfatizan las características peculiares de una tarea concreta que no volverá a repetirse y están centradas en conceptos, hechos o destrezas y habilidades muy específicas.

Obviamente tanto con un tipo de clasificación como con otra, aunque más con la última desarrollada, puede haber enfoques mixtos donde se combinan aspectos genéricos y específicos o dos rúbricas que se combinen para formar una sola, con una evaluación holística y otra analítica de diferentes partes del desempeño. En definitiva, se puede hacer este tipo de combinaciones obteniendo cuatro tipos de rúbricas: holística específica o genérica y analítica específica o genérica.

Especial descripción merecen las rúbricas holísticas y analíticas. Pero antes mencionemos los elementos esenciales con los que cualquier rúbrica debe presentarse.

- 1) Criterios de realización (opcional).
- 2) Atributos, asociados a cada criterio de realización.
- 3) Niveles de desempeño.
- 4) Descripción de los niveles de desempeño.

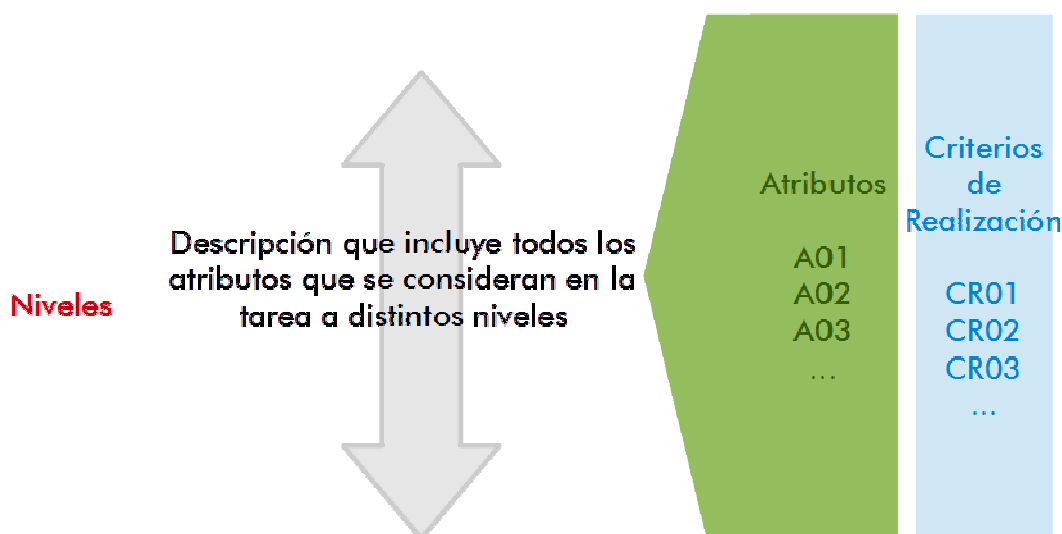
La forma en la que lo hagan determinará si estamos ante una rúbrica holística o analítica. Estos elementos se distribuyen en forma de matriz, filas por columnas, donde niveles y atributos fundamentalmente se reparten los papeles.

## Rúbricas holísticas

Cuando se da la superposición de criterios o son criterios muy asociados, es mejor utilizar este tipo de rúbrica. Los criterios se consideran en combinación en una sola escala descriptiva. Apoyan valoraciones más amplias sobre la calidad del proceso o del producto. Una valoración holística puede también formar parte de una analítica como un factor más, aunque este tipo de métodos tiene la dificultad de sobreponer los criterios de la rúbrica holística con los otros factores.

Su uso resulta más práctico, más rápido que las rúbricas analíticas. Sin embargo, al proporcionar una valoración global del desempeño se pierden detalles específicos y su uso formativo está limitado, al contrario que las analíticas, usándose preferiblemente en evaluaciones con propósito sumativo. También dada su limitada escala de puntuación este uso sumativo se circunscribe casi al hecho de proporcionar una calificación (excelente, notable, bien...), más que una puntuación.

En la rúbrica holística los niveles suelen aparecer delimitando las filas y los criterios de realización y sus atributos se combinan para proporcionar una descripción para cada uno de estos niveles.



### Ejemplo de rúbrica holística

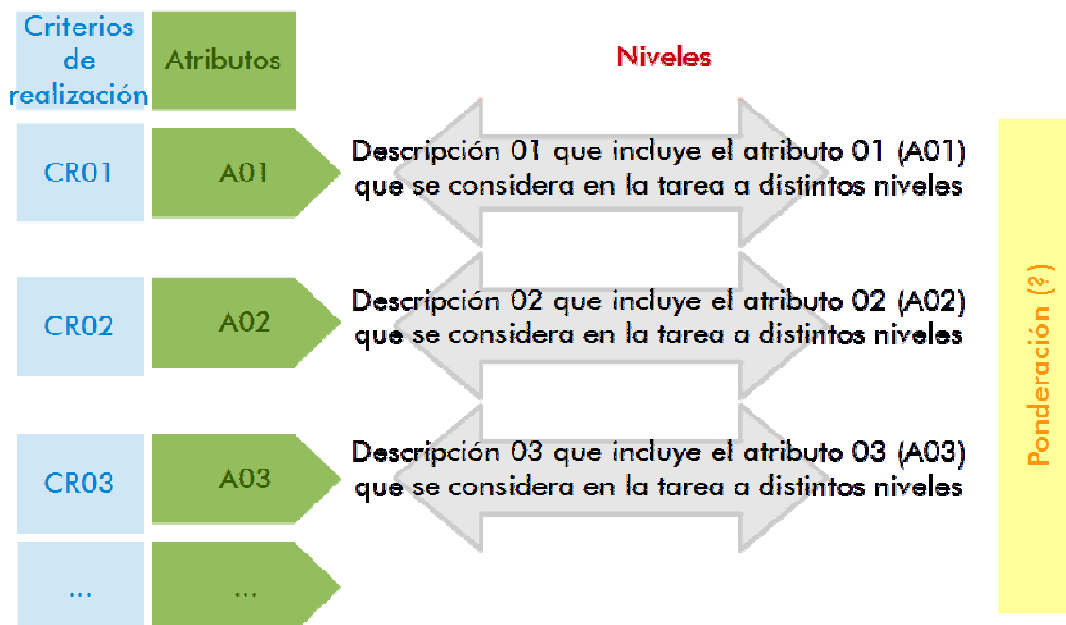
<b>1</b> <i>Necesita mejorar</i>	Se consulta una única fuente de información no fiable. Los indicadores no son relevantes y se dificulta la lectura de los mismos. Además, los análisis y la interpretación no son congruentes con los datos obtenidos.
<b>2</b> <i>Suficiente</i>	Se consulta una única fuente de información aunque fiable. Algún indicador de los presentados es relevante en todas o alguna de las áreas y se refleja de forma clara, aunque no es suficiente para permitir una lectura sencilla de los datos. El análisis y la interpretación es congruente con los datos en alguna ocasión.
<b>3</b> <i>Bueno</i>	Se consultan diversas fuentes de información, aunque alguna de ellas puede no ser fiable. La mayor parte de indicadores y los datos asociados son relevantes y permiten una lectura sencilla de casi toda la información. El análisis y la interpretación se mantiene casi siempre congruente con los datos obtenidos.
<b>4</b> <i>Excelente</i>	Se han consultado diversas fuentes de información fiables. Todos los indicadores y datos son relevantes y están reflejados de forma clara, permitiendo una lectura directa de los datos. El análisis de los datos es completo y permite una interpretación totalmente congruente.

### Rúbricas analíticas

En las rúbricas analíticas los criterios de realización, con sus atributos, se evalúan en los distintos niveles proporcionando una descripción para cada uno de ellos. Cada atributo recibirá una puntuación por separado, aunque también puede obtenerse con propósitos sumativos una puntuación global del desempeño. Ocasionalmente se asignan pesos a la evaluación de cada atributo.

La principal ventaja en el uso de este tipo de rúbricas es que proporciona tanto a los estudiantes como al docente una imagen muy clara de por qué obtienen la puntuación que obtienen los alumnos y se detecta mejor qué partes del desempeño requieren de una revisión por parte de ambos. Su función es diagnóstica y con propósitos claramente formativos, aunque ya se ha dicho que puede utilizarse una puntuación global del desempeño con propósitos sumativos. Sin embargo, para crearlas se necesita mucho tiempo y su uso en la práctica de la evaluación requiere de mucha destreza, especialmente

en desempeños que se evalúan en el mismo momento de su realización.



En este esquema, se aprecia que los niveles forman ahora las columnas, y los atributos, cada uno de ellos por separado, las filas. Para cada atributo, y en cada nivel, se da una descripción. Adicionalmente, se puede considerar que un atributo pese más en el conjunto de atributos que otro, así pueden utilizarse o bien porcentajes o bien factores de multiplicación de los atributos, cuyo uso es más sencillo, y en el fondo es similar al uso de porcentajes.

## Ejemplo de rúbrica analítica

Atributos	1 Necesita mejorar	2 Suficiente	3 Bueno	4 Excelente	Punt.	Peso %
<b>Búsqueda</b>	Ha consultado <u>una única</u> fuente de información <u>no</u> fiable.	Ha consultado <u>una única</u> fuente de información <u>fiable</u> .	Ha consultado <u>varias</u> fuentes de información aunque alguna no sea fiable.	Ha consultado <u>varias</u> fuentes de información, <u>todas</u> ellas fiables.		(opcional)
<b>Relevancia</b>	<u>Ningún</u> indicador es relevante en las tres áreas (educación, economía y salud).	<u>Algun</u> indicador es relevante, en todas o alguna de las tres áreas (educación, economía y salud).	<u>La mayor parte</u> de indicadores son relevantes en las tres áreas (educación, economía y salud).	<u>Todos</u> los indicadores son relevantes en las tres áreas (educación, economía y salud).		
<b>Claridad</b>	<u>Ningún</u> dato está reflejados de forma clara, dificultando la lectura de éstos.	<u>Algun</u> dato está reflejado de forma clara, aunque no es suficiente para permitir una lectura sencilla de éstos.	<u>La mayor parte</u> de los datos está reflejada de forma clara, aunque puede que alguno no permita una lectura sencilla de éstos.	<u>Todos</u> los datos están reflejados de forma clara permitiendo la lectura directa y sencilla de éstos.		
<b>Análisis</b>	El análisis de los datos no está basado en el establecimiento de comparaciones entre los valores de distintos países.	Una pequeña parte del análisis se basa en el establecimiento de comparaciones entre los valores de distintos países.	La mayor parte del análisis se basa en el establecimiento de comparaciones entre los valores de distintos países.	Se realiza un análisis completo de los datos obtenidos estableciendo comparaciones entre los valores de distintos países.		
<b>Interpretación</b>	No se realiza una interpretación congruente con el análisis de los datos obtenidos.	En alguna parte de la interpretación se puede observar congruencia con los datos obtenidos.	En la mayor parte de la interpretación se puede observar congruencia con los datos obtenidos.	Toda la interpretación es congruente con el análisis de los datos obtenidos.		
<b>TOTAL</b>						

En este ejemplo, considerando la rúbrica en un uso sumativo, como hay cuatro niveles y cinco atributos la puntuación oscilará entre los 5 y los 20 puntos. Si además se añade a esto la utilización de factores de multiplicación, la escala de puntuación variará. Por ejemplo, si se decide que el atributo "Análisis" valga el doble, la puntuación a obtener por cualquier estudiante oscilará entre los 6 y los 24 puntos.

## Proceso de elaboración de una rúbrica

El proceso propiamente dicho de redactar una rúbrica es relativamente simple, aunque no sencillo de llevar a cabo correctamente. Se requiere de práctica para elaborar rúbricas con la suficiente validez y fiabilidad en su amplio sentido, cuestiones que serán tratadas más adelante. Así, una vez descritos los criterios de realización e identificados sus atributos quedan unos pocos pasos más. Si no hemos decidido qué tipo de rúbrica vamos a utilizar, sería ahora el momento de realizarlo: ¿analítica u holística?, ¿específica o genérica? Obviamente esta decisión depende del tipo de desempeño y criterios de evaluación que estén implicados.

### Número de niveles y "etiquetado"

El primer paso que debemos realizar es la determinación del número de niveles de desempeño que vamos a utilizar y el uso de números (puntuación), adjetivos que definan el desempeño (calificación) o letras (A, B, C...) o cualquier combinación de estos elementos.

Si se asignan **numerales** se suele utilizar desde el 1 hasta el máximo nivel (por ejemplo 4). Hay otras rúbricas que utilizan el 0 como puntuación mínima hasta el máximo nivel, y otras variantes más imprecisas como utilizar saltos numéricos (por ejemplo: 10, 15, 20, 25) o rangos en cada nivel (por ejemplo: 1,0-3,9; 4,0-4,9; 5,0-5,9; 6-6,9; 7,0-8,9; 9-10). Éstas últimas opciones no son recomendables ya que, observando que se trata de rúbricas usadas para puntuar, obligan a determinar dentro de un nivel qué puntuación voy a asignar



y en base a qué se hacen distinciones dentro de éste. Y si esto no está especificado, incidirá negativamente en la fiabilidad y equidad de la evaluación.

Con el uso de **calificadores** pasa exactamente lo mismo, hay multitud de formulaciones, aunque no tienen el problema de los numerales y suelen utilizarse para establecer una etiqueta al desempeño que, en el caso de requerir una calificación oficialmente, son fácilmente utilizables (insuficiente, suficiente, bien, notable, sobresaliente). Otras fórmulas:

- malo, regular, bueno, excelente
- comienzo, en desarrollo, realizado, ejemplar
- inadecuado, regular, bien, muy bien
- incipiente, en proceso, avanzado, excelente
- ...

El número de niveles más usual se sitúa entre tres y cinco niveles. Un mayor número de niveles implica generalmente una pérdida de consistencia en las descripciones de éstos y como consecuencia incide en la fiabilidad de la puntuación o calificación otorgada en el desempeño. No obstante, el número de niveles ha de adecuarse al grado de complejidad de la actividad de evaluación a evaluar.

Otra cuestión es decidir si va a ser un número par o impar de niveles. En esto no hay un acuerdo unánime, aunque hay una gran parte de autores que se decantan por un número par, polarizando los desempeños en "buenos/malos", evitando así situaciones intermedias "cómodas" para el evaluador.

## Estrategia general de desarrollo de los descriptores

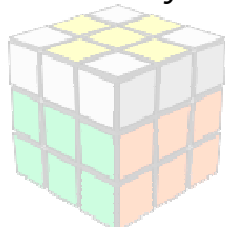
Una vez determinado el número de niveles viene la parte más complicada, la descripción de los niveles en los términos que señala el criterio de realización.

Para mantener la consistencia de las descripciones (y por tanto obtener un instrumento válido) se recomienda mantener todos los elementos que señala el criterio de realización, que está redactado en términos de conducta observable. Lo que variaría de nivel a nivel es el grado de desempeño de esa conducta, y esos grados se basan en tres tipos de calificadores:

- **de cantidad:** ninguno, alguno, la mayoría, todos
- **de frecuencia:** nunca, raramente, a veces, usualmente, siempre
- **de intensidad:** ligeramente, moderadamente, principalmente, extremadamente

## Ejemplo de uso de calificadores en los descriptores

Esta es la descripción de un criterio de realización y su atributo (precisión) en una rúbrica analítica con cuatro niveles (los otros atributos se obvian en el ejemplo: amplitud, relevancia y claridad)



Actividad de evaluación: Dibujar un mapa de la comarca con los elementos geográficos proporcionados.

Criterio de realización: Las convenciones de los mapas (título, leyenda, puntos cardinales y elementos geográficos como ríos, montañas, ciudades, carreteras...) se usan correctamente y los elementos geográficos se sitúan con precisión.

Atributo: Precisión

Nivel 1: Las convenciones de los mapas y los elementos geográficos son **raramente** precisos.

Nivel 2: Las convenciones de los mapas y los elementos geográficos son **a veces** precisos.

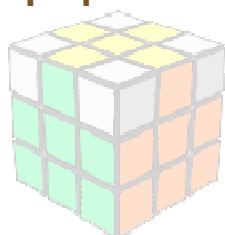
Nivel 3: Las convenciones de los mapas y los elementos geográficos son **usualmente** precisos.

Nivel 4: Las convenciones de los mapas y los elementos geográficos son **siempre** precisos.

El ejemplo anterior es sencillo. En algunas rúbricas se mantiene durante un nivel más o dos una cierta parte de la conducta y la otra va variando. Por ejemplo, en el caso del ejemplo anterior podría darse que usara siempre las convenciones de los mapas en el segundo nivel ya pero todavía raramente situase con precisión los elementos geográficos. No obstante, nos moveríamos siempre con la descripción de los elementos que integran la conducta en diferentes grados de desempeño utilizando los calificadores anteriormente mencionados.

Lo que no es justificable, por ser un error común de consistencia de las descripciones, es añadir conductas o elementos conforme se van describiendo niveles más altos de desempeño que no se han descrito en el criterio de realización, o viceversa, ir quitando elementos de la conducta que desaparecen de la descripción conforme se describen niveles más bajos de desempeño. Otro error común es introducir elementos o conductas que están fuera del criterio de realización, que ni siquiera se han contemplado, o cambiar el nivel de destreza cognitiva (por ejemplo, para un atributo de identificación de elementos pedir en niveles altos de desempeño que reflexionen sobre los elementos identificados).

### Ejemplo de una rúbrica inconsistente



Actividad de evaluación: Lectura en silencio de un libro de lectura a su elección.

Criterio de realización: El alumno/a lee independientemente y está en la tarea durante el período de lectura silenciosa.

Atributo: Lectura silenciosa

Nivel 1: Está fuera de la tarea y se comporta de manera disruptiva durante el período de lectura silenciosa.

Nivel 2: Tiene dificultad en la elección de libros para la lectura silenciosa.

Nivel 3: Lee independientemente durante el periodo de lectura silenciosa.

Nivel 4: Elige libros con entusiasmo y lee independientemente durante el proceso de lectura silenciosa.

Las correcciones sugeridas por Robin y Simon (2004) son:

1. Si la habilidad lectora es el objetivo, pensar el criterio de realización para asegurarse de que el atributo tiene sentido.

2. Si las conductas de aprendizaje van a ser medidas, y la autonomía y la atención son los atributos deseados, reelaborar los descriptores como se muestra:

Atributo: Autonomía y atención

Nivel 1: El estudiante **raramente** lee independientemente y permanece en la tarea **muy poco** tiempo durante el período de lectura silenciosa.

Nivel 2: El estudiante **a veces** lee independientemente y permanece en la tarea **algo** de tiempo durante el período de lectura silenciosa.

Nivel 3: El estudiante **usualmente** lee independientemente y permanece en la tarea **la mayor parte** de tiempo durante el período de lectura silenciosa.

Nivel 4: El estudiante **siempre** lee independientemente y permanece en la tarea **todo** el tiempo durante el período de lectura silenciosa.

Finalmente, para acabar ya con este apartado, ¿por dónde se debería **comenzar a redactar una rúbrica?**

Se recomienda por tanto comenzar bien por el nivel superior y luego describir el inferior, o viceversa. Seguidamente se describirían los niveles intermedios. En el caso de la rúbrica holística se tendrá que hacer una vez únicamente, y en el caso de las rúbricas analíticas se tendrá que hacer para cada atributo.

## Uso de testigos o muestras de anclaje en las rúbricas holísticas

Un aspecto importante o deseable en el uso de rúbricas holísticas es el uso de **testigos** o muestras ejemplares, cuando esto sea posible (por ejemplo, redacciones). Esto significa ni más ni menos que escoger de entre las realizaciones de los estudiantes aquellas que estén en el límite superior de cada nivel, de manera que nos sirvan como elemento global de comparación.

En subsiguientes evaluaciones, si se han establecido bien las rúbricas y los testigos, al evaluar un desempeño de forma global tendremos tanto elementos verbales (descripciones), como elementos visuales (testigos) para situar la ejecución de un estudiante determinado en un nivel.

La pregunta a realizar sería: ¿esta redacción está mejor o peor que este modelo de nivel 1? Si está mejor, pasaría a formar parte del siguiente nivel, el 2, y lo compararíamos de nuevo con el modelo límite del nivel 2 para hacer la misma pregunta. Si está peor, lo dejaríamos en el nivel 1 (el mínimo).

Tener buenos modelos o testigos requiere de tiempo, de haber trabajado suficientemente con la rúbrica y tener producciones de los alumnos que sean significativamente clarificadoras y útiles para valorar globalmente un desempeño determinado.

En resumen, el proceso de construcción de una rúbrica, contemplado desde su inicio, queda reflejado en este esquema. Quedaría por último, abordar la cuestión de la revisión de una rúbrica.



### 3. Revisión de la rúbrica

En este apartado se muestran aquellos aspectos que se han debido tener en cuenta a lo largo del proceso de elaboración de una rúbrica y que ahora servirán como elementos de revisión. Todos ellos inciden en dos características fundamentales de cualquier instrumento, y sobre ellas ya se ha hablado en anteriores documentos de esta serie, la fiabilidad y la validez. Dos conceptos íntimamente relacionados pero no en la misma dirección. Obtener un instrumento fiable no implica que sea válido. Ahora bien, si el esfuerzo está en obtener un instrumento válido es seguro que será en cierta forma fiable.



#### Validez de la rúbrica

En principio se destacan características que se refieren a cómo el instrumento mide lo que el evaluador pretende medir (validez) en base al contenido que recoge el instrumento: es ese contenido y no otro el que se quiere evaluar; y en base a la construcción del propio instrumento respetando la estructura o constructo del contenido (dominio) a evaluar. Así, una rúbrica bien construida debe:

a) describir fielmente el contenido que se quiere evaluar y en toda su extensión. Es decir, están todos los atributos del desempeño o producto a evaluar y están bien descritos (p.ej.: si se quiere evaluar el conocimiento sobre historia, pero se describe en términos de exposición del conocimiento sobre historia, no se está evaluando el conocimiento sino la exposición; por otra parte también se deben evaluar todos aquellos aspectos observables que se entienden por "conocimiento sobre historia" o al menos los más relevantes).

b) contener atributos apropiados que se corresponden con los de la respuesta del alumno en el desempeño o producto que es evaluado (p.ej.: si se evalúa el razonamiento, los errores de pronunciación en la explicación del estudiante son irrelevantes al propósito de la evaluación y no deberían incluirse en los criterios de realización).

#### Fiabilidad de la rúbrica

Este concepto tiene que ver con la precisión de la medida o calificación, en algunos casos, de los instrumentos de evaluación. De ahí que es importante como primera característica que los niveles de la rúbrica están bien definidos y permitan una separación clara en la ejecución o producto. Si no fuese así, cualquier asignación de nivel a un desempeño evaluado tendría más de arbitrario de lo deseable.

Otra característica tiene que ver con lo que ya se ha comentado en el apartado de *Estrategia general de desarrollo de los descriptores*, y es que aunque los descriptores para cada nivel de la escala sean diferentes de los anteriores o los siguientes, esa diferencia la deben dar los cambios producidos en la cualidad de los criterios con el uso de esos calificadores que se han mostrado anteriormente, y no la debería dar la introducción explícita o implícita de nuevos criterios o cambios en la importancia que toman los mismos.

Desde el punto de vista del uso de las rúbricas es deseable que además se hagan comprobaciones del funcionamiento de las rúbricas desde el punto de vista de la fiabilidad. Se puede hacer de dos modos:

- a) Valorar los mismos productos o desempeños (si está la posibilidad) por un mismo docente dos o más veces con las rúbricas elaboradas, para comprobar si se llega a las mismas conclusiones. Entre la primera y segunda valoración se recomienda un período largo de "lavado", es decir, dejar pasar el tiempo para no condicionar la valoración de la segunda vez por el recuerdo de la primera valoración.
- b) Valorar los mismos productos o desempeños por varios docentes con las mismas rúbricas.

Es interesante anotar qué porcentaje de coincidencias se ha obtenido. Si está por

debajo del 80%, la rúbrica se puede mejorar todavía un poco más. Menos del 60% de coincidencias requieren de una revisión profunda de las rúbricas tal y como se ha señalado en este apartado de fiabilidad, tanto en la fiabilidad intrajueces (un mismo valorador en dos o más tiempos diferentes), como en la fiabilidad interjueces (distintos evaluadores valoran varios desempeños). Esto se puede hacer sin necesidad de recurrir a técnicas estadísticas más o menos complejas.

También cabe destacar que el uso de testigos (muestras de desempeño o de anclaje) entre los diferentes niveles de la rúbrica va a contribuir a la consecución de un instrumento fiable.

## Otros aspectos: validez aparente y ecológica

Existen otros aspectos destacables en una rúbrica, que no hacen más que mejorar su apariencia o su uso en un determinado contexto:

Por ejemplo, que las descripciones en cada nivel sean comprensibles, usen un lenguaje claro y directo que las hace prácticas y fáciles de usar. Que tengan una estructura clara, tal como se expone y se ejemplifica en este documento... Son cuestiones relacionadas con la llamada validez aparente, que no es más que un tipo de validez de contenido que viene a destacar, la comodidad o motivación en el uso, en este caso por los docentes, de las rúbricas elaboradas.

En otro orden de cosas, una rúbrica que sea respetuosa con los aprendizajes, que no sea una injerencia sobre el modo de aprender y de enseñar habitual en una clase, es una rúbrica con validez ecológica, un tipo de validez externa (todas las anteriores son internas, del instrumento). Esto se puede conseguir por ejemplo:

- a) Haciendo que las descripciones en los distintos niveles de la rúbrica, o en general, ponen el foco en el logro positivo (éxito) de la realización, no en la falta de ese logro (p. ej.: inaudible vs. necesita proyectar la voz para que se le oiga...). Así, cuando se propone retroalimentación a los alumnos, se hace en términos positivos. Es una característica que cuesta mucho de conseguir, se está acostumbrado a sancionar la falta de logro.
- b) Compartiendo la rúbrica antes de su uso con los estudiantes, de forma que ellos conocen previamente sobre qué se les va a evaluar, o elaborarla directamente con ellos.

Siempre nos quedará el consuelo del cubo de "Rubrik", si no nos sale, siempre podemos intentarlo de otra forma:



Aunque estoy convencido de que con empeño y un procedimiento, acaba por salir.

## 4. Bibliografía

Brualdi Timmins, A.C. (1998). Implementing performance assessment in the classroom. *Practical Assessment, Research & Evaluation*, 6(2). Recuperado 13 junio 2014 de <http://PAREonline.net/getvn.asp?v=6&n=2>.

ERIC/AE (2014). *Scoring Rubrics – Definitions and Construction*. Disponible on-line en: [http://ericae.net/faqs/rubrics/scoring\\_rubrics.htm](http://ericae.net/faqs/rubrics/scoring_rubrics.htm)].

Gonzalez, J. (2014). Know your terms: holistic, analytic and single-point rubrics. Recuperado 1 junio 2014 de <http://www.cultofpedagogy.com/holistic-analytic-single-point-rubrics/>.

Mertler, C.A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25). Recuperado 13 junio 2014 de <http://PAREonline.net/getvn.asp?v=7&n=25>.

Moskal, B. (2000). Scoring rubrics: what, when and how?. *Practical Assessment, Research & Evaluation*, 7(3). Recuperado 13 junio 2014 de <http://PAREonline.net/getvn.asp?v=7&n=3>.

Moskal, B. (2003). Recommendations for developing classroom performance assessments and scoring rubrics. *Practical Assessment, Research & Evaluation*, 8(14). Recuperado 13 junio 2014 de <http://PAREonline.net/getvn.asp?v=8&n=14>.

Moskal, B. M. y Leydens, J.A. (2000). Scoring rubric development: validity and reliability. *Practical Assessment, Research & Evaluation*, 7(10). Recuperado 13 junio 2014 de <http://PAREonline.net/getvn.asp?v=7&n=10>.

Simon, M. y Forgette R. (2001). A rubric for scoring postsecondary academic skills. *Practical Assessment, Research & Evaluation*, 7(18). Recuperado 13 junio 2014 de <http://PAREonline.net/getvn.asp?v=7&n=18>.

Tiemey, R. y Simon, M. (2004). What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research & Evaluation*, 9(2). Recuperado 13 junio 2014 de <http://PAREonline.net/getvn.asp?v=9&n=2>.

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

