

Diseño de rúbricas

ANAYA

Monografías
didácticas

- **Contenidos:** Javier Cortés De las Heras.
- **Diseño:** Dirección de arte: Javier Serrano. Cubierta: Patricia Gómez Serrano. Interiores: Marta Gómez Peso.
- **Corrección:** Sergio Borbolla.

Las **normas ortográficas** seguidas en este libro son las establecidas por la Real Academia Española en la *Ortografía de la lengua española*, publicada en el año 2010.

Nuestras publicaciones mantienen el rigor en el uso y en la selección de los contenidos, en las imágenes y en el lenguaje, para cumplir con la **no discriminación** por razón de género, cultura u opinión.

© GRUPO ANAYA, S.A., 2015 - C/ Juan Ignacio Luca de Tena, n.º 15 - 28027 Madrid.

Reservados todos los derechos. El contenido de esta obra está protegido por la Ley, que establece penas de prisión y/o multas, además de las correspondientes indemnizaciones por daños y perjuicios, para quienes reprodujeren, plagiaren, distribuyeren o comunicaren públicamente, en todo o en parte, una obra literaria, artística o científica, o su transformación, interpretación o ejecución artística fijada en cualquier tipo de soporte o comunicada a través de cualquier medio, sin la preceptiva autorización.

Índice de contenidos

0. Introducción	4
1. La evaluación de ejecuciones	5
• Pasos en la evaluación de ejecuciones	
• Delimitación del propósito de la evaluación	
• Especificación de criterios para la evaluación	
2. Diseño de rúbricas	14
• Qué es una rúbrica	
- Qué definición de rúbrica vamos a adoptar	
- Ventajas y desventajas del uso de rúbricas	
- Elementos esenciales de una rúbrica	
• Tipo de rúbricas	
- Rúbricas holísticas	
- Rúbricas analíticas	
• Proceso de elaboración de una rúbrica	
- Número de niveles y «etiquetado»	
- Estrategia general de desarrollo de los descriptores	
- Uso de testigos o muestras de anclaje en las rúbricas holísticas	
- En resumen	
3. Revisión de la rúbrica	28
• Validez de la rúbrica	
• Fiabilidad de la rúbrica	
• Otros aspectos: validez aparente y ecológica	
4. Bibliografía	31

Introducción

En este cuaderno se abordarán los pasos necesarios para diseñar rúbricas como instrumento primordial en la evaluación de ejecuciones o del desempeño. Así, 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 un currículo oficial de Educación Primaria del territorio nacional que, por su diseño, ha permitido extraer con mayor facilidad los elementos esenciales para elaborar rúbricas referidas al criterio, como no puede ser de otra manera.

Desde esta premisa, nuestros elementos de trabajo deberían ser los **criterios de evaluación** y, en la medida de lo posible, los **indicadores de logro** o **estándares de aprendizaje evaluables** que se han definido en las distintas comunidades autónomas derivados de estos criterios. 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. Estos resultados de aprendizaje 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** o **evaluación del desempeño**.

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.
- Los criterios de calidad de una rúbrica, que tienen que ver con su validez y fiabilidad.

1 La evaluación de ejecuciones

La **evaluación auténtica** forma parte de una 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 el artículo de Frey, Schmitt y Allen (2012) sobre la revisión de la literatura publicada hasta el momento acerca del concepto de evaluación auténtica, se delimitan nueve características que tienen en común las definiciones sobre este tipo de evaluación realizadas por cuarenta y siete autores. De ellas, las más citadas para alumnado de Educación Primaria, y que reproducimos aquí, se refieren a:

1. Contexto

- Plantea una actividad o un 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%).

Los mismos autores señalan que en educación secundaria o formación profesional destacan más o menos las mismas características.

Así para Wiggins (1998, pág. 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.

La cuestión de que la evaluación auténtica ha de estar basada en el tipo de trabajo real que hace la gente lleva a pensar en que el alum-

nado ha de mostrar su desempeño en un contexto en el que ha de desarrollar un abanico amplio de conocimientos, destrezas y actitudes de manera eficiente. Es decir, ha de mostrar que es competente. Este tipo de evaluación cobra especialmente importancia en Educación Primaria (23%) y en mayor medida en áreas como la educación física o la educación artística (plástica y música), aunque puede extenderse a otras áreas del ámbito lingüístico o científico.

En no pocas ocasiones, el término de evaluación auténtica se asimila al de **evaluación de ejecuciones** o del desempeño (del inglés: *performance assessment*) y es porque, como se ha visto, 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 colaborar con el alumnado en los criterios de realización. Es decir, la evaluación de ejecuciones es un requisito indispensable de la evaluación auténtica, pero no suficiente para que esta se dé. Piénsese en el acceso a un puesto de trabajo de comercial en el que 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; sin embargo, 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 el 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 el aula?** 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. Y es aquí donde entran en juego las rúbricas.
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 la mayor parte de currículos, en los criterios de evaluación, y sus especificaciones en indicadores de logro o estándares de aprendizaje, 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. Estos 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 otros instrumentos, en el contexto de la evaluación del desempeño se pueden seguir unas fases de desarrollo que abarcan no solo el diseño propiamente dicho, sino además todo el proceso de evaluación. Veamos la tabla que compara el esquema general de la evaluación referida al criterio frente a la evaluación del desempeño.

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.

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 quieren evaluar?
- ¿Qué deberían saber los alumnos sobre lo 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 en el que se especifica qué es lo que los escolares deben ser capaces de hacer al finalizar una determinada unidad didáctica, trimestre, curso..., es decir, si hay resultados de aprendizaje bien definidos, ya tenemos recorrido gran parte del camino. Si los criterios de evaluación están redactados en términos de resultado de aprendizaje (y sus posibles especificaciones como indicadores de logro, microcriterios de evaluación, descriptores...), 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. A saber:

- Utilizan uno o más verbos a un nivel jerárquico en alguno de los dominios cognitivo, motor o afectivo.
- Especifican un contenido (hechos, conceptos, principios, procedimientos, estrategias cognitivas, destrezas, habilidades, valores, actitudes, conductas...).
- Y, en algunas ocasiones, una naturaleza de ejecución o contextos de realización, referidos al fin, los recursos humanos o materiales, las estrategias, los modos, los agrupamientos, el tiempo...

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

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

Criterios de evaluación (resultados de aprendizaje):

- 6.º MAT Bloque 5.1. 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:

- 6.º MAT Bloque 5.1.1. Recoge y registra informaciones cuantificables a partir de encuestas, medios de comunicación, INE, Internet...
- 6.º MAT Bloque 5.1.2. 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, como docente, en el tipo de actividad de evaluación, e ir respondiendo a determinadas preguntas:

- ¿Son **acciones complejas** que pueden observarse en un proceso o en un momento puntual?
- **¿Qué actividades de evaluación, pruebas, podría plantear?** Una exposición oral con presentación en la que se llegue a unas conclusiones, un informe escrito donde los estudiantes plasmen el proceso, una prueba escrita en la que identifiquen datos, los analicen y los interpreten...
- **¿Cuándo voy a evaluar el desempeño?** Un día, varios días; en clase; en una actividad extraescolar; desde casa con un documento que entreguen...
- **¿Cómo y con qué voy a evaluar?** ¿Evalúa solo el docente? ¿Que se valoren entre ellos... utilizando algún instrumento dependiendo de la prueba de evaluación: prueba objetiva, ensayo de opinión, lista de cotejo, una rúbrica...?
- ¿Qué **conceptos, procedimientos y actitudes** necesitan haber trabajado antes para realizar la prueba de evaluación en condiciones? Por ejemplo: la moda, la media y el rango; el acceso y el manejo de fuentes de información, el conocimiento de una encuesta...

Especificación de criterios para la evaluación

Una vez analizado el dominio que se va a evaluar, los aprendizajes concretos, 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 (profesorado) 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 realización 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 idénticas 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 la escala de medida.

Ejemplo de especificación de criterios para la evaluación

Criterios para la evaluación:

- Se acota el tema dentro de la unidad didáctica; por ejemplo, indicadores de países del mundo (salud, economía, educación...).
- Se delimita la prueba: **informe escrito**.
- Y sus límites o condiciones: la estructura, la extensión y el contenido del informe (por ejemplo, el número de indicadores a registrar); agrupamiento: en grupo; fecha de entrega; materiales a proporcionar (páginas web); visita al Instituto 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 de
- 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:

1. Identificar globalmente la destreza o el producto.

Quizás 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 efectuará un estudiante tipo en el desempeño; con qué dificultades se encontrará; qué utilizará (recursos materiales); si es un producto, qué ca-

racterísticas 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 elaborar una lista con aquellos aspectos, al menos los 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 estos 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, en la que 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 accesoría.

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 (por ejemplo, una exposición oral sobre un tema).

Todos estos pasos que están expuestos linealmente no tienen por qué seguirse en ese orden. De hecho, es un proceso que requiere 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

Búsqueda¹

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 tres y cuatro del proceso de evaluación del desempeño expuesto en la Tabla 1. Las fases cinco y seis no difieren del uso de otros instrumentos en el marco de la evaluación referida al criterio y no serán tratadas en este documento.

¹ Los atributos van en negrita.

2 Diseño de rúbricas

Qué es una rúbrica

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 fueron los que propusieron el término) añadieron el adjetivo *scoring* al sustantivo *rubric* para denotar un tipo especial de rúbrica.

Según el Longman Dictionary of Contemporary English, *rubric* es un conjunto de instrucciones o una explicación en un libro, examen, etc. *Scoring rubric*, de manera literal, significaría «conjunto de instrucciones que sirven para puntuar».

En el diccionario de la Real Academia de la Lengua Española, en su 23.^a 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 ha incorporado plenamente su uso con la acepción que tienen en el entorno anglosajón. Las referencias remiten al uso en la firma (a esos rallajos que hacemos encima de nuestro nombre cuando firmamos algún documento) o a un uso más antiguo, que remite además a la etimología de la palabra, derivada del latín *ruber* = rojo², que era el color utilizado en los epígrafes o los rótulos de los libros y probablemente en el conjunto de las reglas de la liturgia.

² 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.

Dejando la cuestión estrictamente léxica y centrándose en la cuestión pedagógica, en distintos libros, artículos y otros tipos de documentos que tratan las rúbricas, suelen aparecer diferentes definiciones. Veamos un ejemplo:

El glosario de términos de pruebas estandarizadas del Educational Testing Service (ETS) —el encargado, últimamente, de la evaluación de PISA (https://www.ets.org/understanding_testing/glossary/)— define rúbrica como «un conjunto de reglas para puntuar las respuestas a un ítem de respuesta construida³, a veces denominado guía de puntuación o valoración (*scoring guide*)».

Personalmente, creo que esta no es una definición de aplicación general, pues responde en exceso a un contexto de pruebas estandarizadas; es más, las «guías de valoración» utilizadas en las pruebas PISA no son estrictamente una rúbrica, ya que la rúbrica requiere algunas características más específicas.

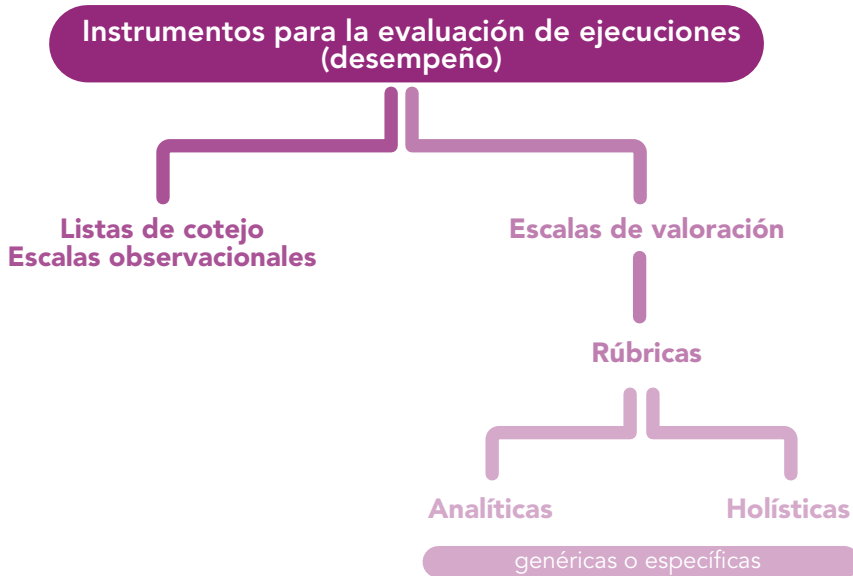
El ejemplo nos sirve para comprender que el análisis de las distintas definiciones de rúbrica empleadas en diferentes documentos podría llevarnos a una larga discusión, y que es preciso que prefijemos la definición que vamos a adoptar antes de entrar en otros análisis.

Qué definición de rúbrica vamos a adoptar

En esta monografía, se propone una definición que tiene sentido en el marco de lo que se va a exponer seguidamente. Definimos rúbrica como un 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 que se va a registrar se limita a la determinación de si se cumple o no con unos criterios específicos.

³ Una pregunta de una prueba objetiva que requiere una respuesta abierta y, generalmente, larga.



En el esquema anterior se hace una clasificación de los distintos instrumentos 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 realizará 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, y que no se pueden valorar como correctos o incorrectos.

Una de sus grandes ventajas es que su empleo se puede extender a todos los niveles educativos, desde la Educación Infantil hasta la Educación Superior, y a diferentes materias y situaciones. Son tremendamente versátiles.

Son útiles en un enfoque de evaluación para el aprendizaje o la evaluación con propósitos formativos: nos ayudan a explorar hasta qué punto se han alcanzado unos determinados criterios de realización y permiten proporcionar una retroalimentación a los estudiantes sobre cómo mejorar sus realizaciones. Es más, el alumnado puede colaborar

en su construcción, participando en la definición de los criterios de realización, con mayor o menor apoyo del docente según el nivel educativo en el que se encuentre.

Como contrapartida, cabe decir que una rúbrica mal diseñada puede ser perjudicial para el aprendizaje. Si su construcción no se hace adecuadamente, puede haber inconsistencias en las descripciones de los criterios de realización en los diferentes niveles de la escala. Las más comunes y accesibles en Internet contienen fallos de diseño que no solo afectan a la instrucción, sino también a la validez de los resultados.

Por otra parte, no todos los aprendizajes han de ser evaluados 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. No debemos desdeñar ningún instrumento de evaluación ni forzarnos a emplear rúbricas cuando la situación no lo requiera. 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 cada vez más frecuentes de un uso forzado de las rúbricas.

Elementos esenciales de una rúbrica

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 se presenten estos elementos determinará si estamos ante una rúbrica holística o analítica. Estos elementos se distribuyen en forma de matriz (filas por columnas). En esta matriz, son los niveles y los atributos los que fundamentalmente se reparten los papeles.

Tipos de rúbricas

Habitualmente se suele elaborar 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 que atiende al tipo de desempeño que se evalúa (si es específico o si es general), que depende del tipo de criterios por los que se evalúa dicho desempeño, y que divide las rúbricas en específicas o genéricas. En esta monografía, no profundizaremos demasiado en esta clasificación; sirva decir que las **rúbricas genéricas** se dan cuando existen desempeños que reúnen características similares, independientemente del contenido, que son, por ejemplo, interdisciplinarios o la presentación de una información (siempre y cuando el contenido de esta no sea objeto de evaluación).

Las **rúbricas específicas** 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 una clasificación como con la otra, puede haber enfoques mixtos donde se combinan aspectos genéricos y específicos, o dos rúbricas para formar una sola, con una evaluación holística y otra analítica de diferentes partes del desempeño.

En definitiva, se pueden hacer combinaciones y obtener cuatro tipos de rúbricas: holística específica, holística genérica, analítica específica y analítica genérica.

Especial descripción merecen las rúbricas holísticas y las analíticas.

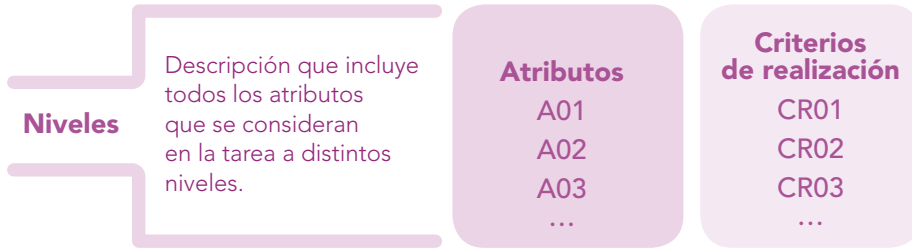
Rúbricas holísticas

El uso de rúbricas holísticas se recomienda cuando se da la superposición de criterios o los criterios están muy asociados. 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.

Las rúbricas holísticas son más fáciles de construir y de aplicar que las analíticas, pero solo proporcionan una valoración global del desempeño (con ellas se pierden detalles específicos) y su uso formativo es limitado, por lo que se emplean preferiblemente en evaluaciones con propósito sumativo. Además, 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

<p>1</p> <p>Necesita mejorar</p>	<p>Se consulta una única fuente de información no fiable. Los indicadores no son relevantes y se dificulta su lectura. Además, los análisis y la interpretación no son congruentes con los datos obtenidos.</p>
<p>2</p> <p>Suficiente</p>	<p>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.</p>
<p>3</p> <p>Bueno</p>	<p>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.</p>
<p>4</p> <p>Excelente</p>	<p>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.</p>

Rúbricas analíticas

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

La principal ventaja del uso de este tipo de rúbricas es que proporciona, tanto al alumnado como al profesorado, una imagen muy clara de por qué los estudiantes obtienen determinada puntuación, y permiten detectar con facilidad qué partes del desempeño requieren una revisión por parte de cada uno. Su función es diagnóstica y con propósitos claramente formativos, aunque ya se ha indicado 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 mucha destreza; especialmente en desempeños que se evalúan en el mismo momento de su realización.

Atributos	Criterios de realización	Niveles	Ponderación
A01	CR01	Descripción 01 que incluye el atributo 01 (A01) que se considera en la tarea a distintos niveles.	
A02	CR02	Descripción 02 que incluye el atributo 02 (A02) que se considera en la tarea a distintos niveles.	
A03	CR01	Descripción 03 que incluye el atributo 03 (A03) que se considera en la tarea a distintos niveles.	
...	

En el esquema anterior, 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 proporciona una descripción.

Adicionalmente, se puede considerar que un atributo pese más en el conjunto de atributos que otro; en tal caso, 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	Puntuación	Peso
Búsqueda	Ha consultado una única fuente de información no fiable.	Ha consultado una única fuente de información fiable.	Ha consultado varias fuentes de información aunque alguna no sea fiable.	Ha consultado varias fuentes de información, todas ellas fiables.		(opcional)
Relevancia	Ningún indicador es relevante en las tres áreas (educación, economía y salud).	Algún indicador es relevante, en todas o alguna de las tres áreas (educación, economía y salud).	La mayor parte de indicadores son relevantes en las tres áreas (educación, economía y salud).	Todos los indicadores son relevantes en las tres áreas (educación, economía y salud).		
Claridad	Ningún dato está reflejado de forma clara, dificultando la lectura de estos.	Algún dato está reflejado de forma clara, aunque no es suficiente para permitir una lectura sencilla de estos.	La mayor parte de los datos está reflejada de forma clara, aunque puede que alguno no permita una lectura sencilla de estos.	Todos los datos están reflejados de forma clara permitiendo la lectura directa y sencilla de estos.		
Análisis	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.		
Interpretación	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.		
Total						

En el ejemplo anterior, si se considera la rúbrica en un uso sumativo, como hay cuatro niveles y cinco atributos, la puntuación estará comprendida 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 que puede obtener un estudiante quedará comprendida entre los 6 y los 24 puntos.

Proceso de elaboración de una rúbrica

El proceso en sí de redacción de una rúbrica es relativamente simple, aunque no sencillo de llevar a cabo de manera correcta. Se requiere práctica para elaborar rúbricas con la suficiente validez y fiabilidad en su amplio sentido, cuestiones que serán tratadas más adelante. Una vez descritos los criterios de realización e identificados sus atributos, deberemos:

- Decidir qué tipo de rúbrica vamos a utilizar (¿analítica u holística?, ¿específica o genérica?). Obviamente esta decisión depende del tipo de desempeño y de los criterios de evaluación que estén implicados.
- Determinar el número de niveles y realizar el «etiquetado».
- Desarrollar los descriptores.
- Recoger testigos o muestras ejemplares, si la rúbrica es holística.

Número de niveles y «etiquetado»

El primer paso que debemos realizar es decidir el número de niveles de desempeño que vamos a determinar y si en estos vamos a emplear números (puntuación), adjetivos que definan el desempeño (calificación), letras (A, B, C...) o cualquier combinación de estos elementos.

Cuando se asignan **numerales** se suele utilizar una puntuación que va desde el 1 hasta el máximo nivel (por ejemplo, el 4). Hay rúbricas que utilizan el 0 como puntuación mínima, y rúbricas más imprecisas que emplean 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).

Estas últimas opciones no son recomendables, ya que complican el proceso de asignación de las puntuaciones.

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 asignar una etiqueta al desempeño que puede emplearse para realizar una calificación oficial (insuficiente, suficiente, bien, notable, sobresaliente).

Otras fórmulas posibles son:

- 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. Un mayor número de niveles implica generalmente una pérdida de consistencia en las descripciones de estos y, como consecuencia, incide en la fiabilidad de la puntuación o la calificación otorgada en el desempeño. No obstante, el número de niveles ha de adecuarse al grado de complejidad de la actividad que se va a evaluar.

Otra cuestión es decidir optar por 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 conservar 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 (amplitud, relevancia y claridad) se obvian en el ejemplo.

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 uno o dos niveles una cierta parte de la conducta y la otra va variando. Por ejemplo, en el caso del ejemplo anterior, podría formularse el nivel dos de este otro modo:

- **Nivel 2:** Las convenciones de los mapas son **siempre** precisas, y los elementos geográficos son **raramente** precisos.

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, y que es un error común en la consistencia de las descripciones, es añadir conductas o elementos conforme se van describiendo niveles más altos de desempeño; 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 o la alumna 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 período 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.

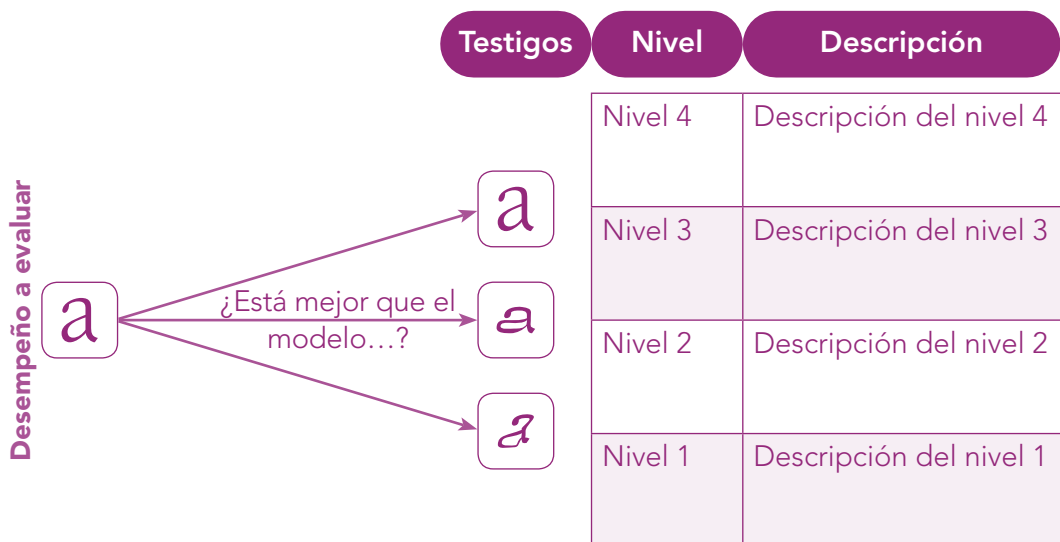
- **Nivel 1:** El estudiante raramente lee independientemente y permanece en la tarea muy poco tiempo durante el período de lectura silenciosa.
- **Nivel 2:** El estudiante a veces lee independientemente y permanece en la tarea parte del tiempo durante el período de lectura silenciosa.
- **Nivel 3:** El estudiante usualmente lee independientemente y permanece en la tarea la mayor parte del 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.

Como recomendación final sugerimos que, para redactar una rúbrica:

- Se redacte el nivel superior y, a continuación, el inferior, o viceversa.
- Se describan seguidamente 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

También cabe destacar que el uso de testigos (muestras de desempeño o de anclaje) entre los diferentes niveles de la rúbrica, especialmente en las holísticas, va a contribuir a la consecución de un instrumento fiable. Estas muestras de desempeño se sitúan en los límites de cada nivel de la rúbrica. Es decir, el mejor desempeño del nivel 1 es el mejor modelo o muestra de ese nivel; el siguiente sería el mejor modelo de desempeño del nivel 2..., y así sucesivamente.



Se trata entonces de evaluar un desempeño cualquiera con las descripciones que se proporcionan en cada uno de los niveles de la rúbrica, como es habitual, pero ayudándose en este caso de los modelos representativos de cada nivel, que no son más que desempeños reales tipo. Se compara un desempeño a evaluar con cada uno de esos modelos: «¿está mejor que el modelo del nivel 1?». Si fuera así, se compararía con el modelo del nivel 2, y así sucesivamente. Si no, obviamente está en el nivel 1.

Disponer de buenos modelos o testigos es una cuestión que depende del tiempo: requiere que hayamos trabajado considerablemente con la rúbrica y que dispongamos de un número suficiente de producciones de los alumnos y de las alumnas 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 el esquema siguiente. Faltaría, por último, abordar la cuestión de la revisión de la 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, 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 se centra 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 aquel recoge; en este caso, la rúbrica.

La pregunta principal a la que hay que responder es si esa rúbrica que se acaba de diseñar o que se ha cogido prestada de otras fuentes:

- a) Describe 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 (por ejemplo, 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) Contiene atributos apropiados que se corresponden con los de la respuesta del alumno o de la alumna en el desempeño o producto que es evaluado (por ejemplo, si se evalúa el razonamiento, los errores de pronunciación en la explicación del estudiante son irrelevantes para el 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 sea importante como primera característica que los niveles de la rúbrica estén 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 (cantidad, frecuencia, intensidad), y no la debería proporcionar la introducción explícita o implícita de nuevos criterios o cambios en la importancia que estos toman.

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 la segunda valoración se recomienda un período prolongado de «lavado»; es decir, dejar pasar el tiempo para no condicionar la valoración de la segunda vez por el recuerdo de la primera.
- 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 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.

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 cuaderno... 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 la 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, pongan el foco en el logro positivo (éxito) de la realización, no en la falta de ese logro (por ejemplo, «inaudible frente a necesita proyectar la voz para que se le oiga...»). Así, cuando se propone una retroalimentación a los alumnos, se hace en términos positivos. Es una característica que cuesta mucho conseguir; se está acostumbrado a señalar la falta de logro.
- b) Compartiendo la rúbrica antes de su uso con los estudiantes, de forma que ellos conozcan previamente sobre qué se les va a evaluar, o elaborarla directamente con ellos.

4 Bibliografía

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

ERIC/AE (2014). *Scoring Rubrics – Definitions and Construction*. Disponible online en http://ericae.net/faqs/rubrics/scoring_rubrics.htm

Gonzalez, J. (2014). «Know your terms: holistic, analytic and single-point rubrics». Recuperado el 1 de 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 el 13 junio de 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 el 13 junio de 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 el 13 junio de 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 el 13 junio de 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 el 13 junio de 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 el 13 junio de 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.