

Autoevaluación y co-evaluación: estrategias para facilitar la evaluación continuada

Miguel Valero-García, Luis M. Díaz de Cerio

Dept. d'Arquitectura de Computadors
Escola Politècnica Superior de Castelldefels, Universitat Politècnica de Catalunya
Avda. del Canal Olímpic, s/n
08860 Castelldefels (Barcelona)
{miguel, ldiaz}@ac.upc.edu

Resumen

Allá donde se intentó implantar de forma generalizada, la evaluación continuada acabó por provocar frustración entre el profesorado, por diferentes razones, una de las cuales es el importante incremento de carga de trabajo debida a la corrección de los ejercicios de los alumnos. En nuestra opinión, se suelen mezclar dos aspectos que es importante separar: (a) Evaluación Formativa: evaluar e informar frecuentemente al alumno de forma cualitativa, sobre lo bien o lo mal que va en el curso (cuestión fundamental para el aprendizaje) y (b) Evaluación Calificativa: evaluar cuantitativamente y asignar calificaciones a las tareas, ejercicios, etc., a lo largo del curso (cuestión secundaria respecto al aprendizaje). Cuando se separan estos aspectos, es posible plantear sistemas de evaluación continuada que satisfacen el requerimiento fundamental: mantener informado al alumno, a un coste razonable para el profesor. En este trabajo se propone el uso de la autoevaluación y la co-evaluación como sistemas de evaluación formativa a coste razonable, y se analizan las implicaciones de diferentes esquemas.

1. Introducción: La promesa de la evaluación continuada (la teoría y la realidad)

Está ampliamente aceptado que un ingrediente de la docencia de calidad es un sistema de retroalimentación que permita al alumno mantenerse puntualmente informado sobre su progreso (o falta de progreso) en el plan de aprendizaje [1]. Después de resolver un ejercicio, por ejem-

plo, un alumno necesita saber si lo ha hecho bien o mal, y por qué. Y necesita saberlo pronto, para poder tomar rápidamente las acciones correctoras necesarias. En otras palabras, se trata de evitar que el alumno descubra el día del examen final (o peor aún, al conocer las notas del examen final) que ha entendido bien poca cosa del material del curso.

A pesar de esta obviedad, no es fácil encontrar asignaturas con un buen sistema de evaluación continuada, especialmente cuando el número de alumnos por profesor es elevado. De hecho, una de las quejas más frecuentes de los alumnos es que no tienen una buena información sobre su progreso a lo largo del curso, y que con frecuencia se ven sorprendidos en los exámenes, cuando se pone en evidencia que su rendimiento real está muy por debajo de su predicción.

La evaluación continuada ha sido un estandarte en materia de mejora de la docencia en muchos sitios. Así por ejemplo, fue uno de los ingredientes principales de la reforma de los planes de estudio en la Universidad Politécnica de Cataluña (UPC), a inicios de los 90. Sin embargo, en muchos casos, y por razones de coste tanto desde el punto de vista del profesor como desde el punto de vista del alumno, los sistemas de evaluación continuada han ido simplificándose hasta convertirse en poco más de un examen parcial a mitad del cuatrimestre, y un último examen al final, con lo que el objetivo básico de la evaluación continuada (mantener bien informado al alumno sobre el estado de su aprendizaje) no se cumple.

En este artículo proponemos una serie de ideas que pueden resultar de utilidad para orga-

nizar un sistema de evaluación continuada a un coste asumible. En concreto, en la siguiente sección separamos los dos conceptos clave que intervienen en la evaluación continuada: información continuada (evaluación formativa) y calificación continuada (evaluación calificativa). En la sección 3 proponemos dos técnicas para realizar la evaluación formativa con bajo coste y por tanto disminuyendo en global el coste de la evaluación continuada. En la sección 4 veremos cómo corresponden las técnicas propuestas respecto a los niveles de competencia según la conocida taxonomía de Bloom. En las secciones 5 y 6 presentamos diferentes metodologías para llevar a cabo las técnicas propuestas y finalmente, en la sección 7, presentamos las conclusiones.

2. Información continuada versus calificación continuada (separemos las cosas)

En esta sección consideramos algunas cuestiones teóricas sobre la evaluación y sus características. Estas consideraciones son la base en la que se apoyan las propuestas que se realizan en este artículo.

Cuando se trata de la evaluación del nivel de aprendizaje de nuestros alumnos, solemos distinguir entre dos tipos, según el uso que se hace de la información:

- Evaluación formativa: se usa para guiar y mejorar los procesos de enseñanza y aprendizaje
- Evaluación calificativa: se usa para determinar la calificación que acredita el nivel de aprendizaje conseguido por el alumno.

Por otra parte, un sistema de evaluación (tanto calificativa como formativa) puede tener (entre otros) los siguientes atributos:

- Precisión y fiabilidad: El resultado de la evaluación es el mismo con independencia de la persona que realice la evaluación, o del momento en que se realice la evaluación (fiabilidad), y ese resultado tiene poco margen de error (preciso).
- Prontitud: El resultado de la evaluación está en manos del alumno y del profesor lo antes posible, después de haber realizado el acto a evaluar.

Es interesante ahora analizar la relación que existe entre los tipos de evaluación y los atributos de la evaluación. En particular, puede afirmarse que:

La evaluación calificativa debe ser precisa y fiable, puesto que está en juego el expediente del alumno, que es un documento oficial que puede tener mucha trascendencia a la hora de buscar trabajo, optar a becas, etc. Sin embargo, no es especialmente crítico que los resultados de la evaluación calificativa estén disponibles con prontitud¹.

El atributo más importante de la evaluación formativa es la prontitud. En este caso, también es deseable que el sistema sea preciso y fiable, pero estos atributos son secundarios. En otras palabras, cuando un alumno hace, por ejemplo, un ejercicio, lo importante es que el alumno sepa pronto si las decisiones importantes han sido acertadas. En cambio, no es excesivamente importante saber si el resultado merece una nota de 6.5 o una de 7 (sí que lo es, insistimos, en el caso de la evaluación calificativa).

Finalmente, es importante observar que lo que hace que un sistema de evaluación sea costoso en términos de tiempo de profesor es que el sistema sea preciso y fiable. Efectivamente, el tener que determinar una calificación, típicamente con una precisión de 0.5 puntos, implica un tiempo de análisis del trabajo realizado por cada alumno, que puede llegar a suponer una carga total de trabajo importante cuando hay muchos alumnos por profesor, y muchos actos de evaluación a lo largo del curso.

Lo importante ahora es observar que, en el caso de la evaluación formativa (la que nos interesa en el contexto de este trabajo), es aceptable sacrificar precisión y fiabilidad (y por tanto, reducir coste para el profesor), siempre y cuando se mantenga el requisito de prontitud. Este es el principio subyacente en las propuestas que se desarrollan en las secciones siguientes.

En resumen, cuando se habla de evaluación continuada, conviene distinguir entre información continuada y calificación continuada. La

¹ No obstante, casi siempre la prontitud acaba siendo necesaria para satisfacer los requisitos del sistema administrativo, y del calendario académico, que debe cerrar procesos en una fecha determinada para dar inicio al siguiente cuatrimestre.

información continuada es imprescindible para que el proceso de enseñanza-aprendizaje tenga salud. La calificación continuada es costosa y no es en absoluto necesaria. Incluso en muchos casos, especialmente en el caso de la organización cuatrimestral, no es conveniente calificar (no confundir con evaluar) al alumno antes de la última fase del curso, debido a que parte de su nota queda determinada cuando todavía no ha tenido la oportunidad de asimilar y ejercitar los conceptos básicos. Éste es, por ejemplo, el caso de un primer curso de enseñanza de la programación, en la que todo lo que puede calificarse en la primera mitad de un cuatrimestre tiene poca relevancia comparado con la fase final del curso, que es cuando el alumno está en condiciones de resolver ejercicios de una mínima entidad.

3. Autoevaluación y co-evaluación: información con prontitud y a bajo coste

Las estrategias de autoevaluación y co-evaluación pueden usarse como base para la organización de un sistema de evaluación que proporcione información con prontitud, aunque para ello tengamos que renunciar a la precisión y a la fiabilidad. La idea básica de estas estrategias es que los alumnos pueden ser colaboradores del profesor, en este caso, en las tareas de evaluación; ya que los profesores, por lo general, no disponen de ayudantes para realizar la labor evaluadora. Si los alumnos han de colaborar en la evaluación, las únicas dos opciones de las que disponemos son: evaluarse a sí mismos (autoevaluación) o evaluar a otros compañeros (co-evaluación). En concreto, cuando se usa la estrategia de la autoevaluación, es el propio alumno el que determina en qué medida su trabajo está bien o mal siguiendo las instrucciones del profesor. En el caso de la co-evaluación, cada alumno evalúa el trabajo de uno o varios de sus compañeros, también siguiendo las instrucciones del profesor.

La autoevaluación y la co-evaluación proporcionan información con prontitud, puesto que si el profesor tiene preparadas las instrucciones con antelación, los alumnos pueden realizar la evaluación inmediatamente después de realizar el trabajo y obtener las conclusiones rápidamente.

Lógicamente, la evaluación será menos precisa y fiable que si la hubiese realizado el profesor, puesto que el profesional es él, y no los alumnos.

Además de resolver el problema básico que nos concierne en este trabajo, la autoevaluación y la co-evaluación tienen otras virtudes que conviene tener bien presentes [2]. En concreto, en el caso de la autoevaluación:

- Los alumnos van interiorizando los criterios de corrección que el profesor hace explícitos a través de las instrucciones para la autoevaluación. Esto permite a los alumnos ajustar cada vez más sus respuestas a lo que el profesor espera.
- Los alumnos desarrollan el hábito de la reflexión, y la identificación de los propios errores, cuestión fundamental cuando se trata de formar personas con capacidad para aprender de forma autónoma.

En el caso de la co-evaluación, además de las virtudes anteriores, podemos mencionar también las siguientes:

- Los alumnos se esfuerzan más, impulsados por la motivación de quedar bien ante los ojos de sus compañeros (este tipo de motivación suele ser superior a la de quedar bien ante los ojos del profesor).
- Los alumnos desarrollan el hábito de criticar de forma constructiva el trabajo realizado por compañeros con los que van a tener que continuar colaborando. Ésta es también una habilidad fundamental que se echa en falta con frecuencia en el mundo profesional.

Es posible que al plantear un sistema de evaluación continuada basado en la autoevaluación y en la co-evaluación los alumnos manifiesten unas ciertas reticencias, e incluso puedan acusar al profesor de cargarles a ellos con un trabajo y una responsabilidad que no les corresponde. En estos casos, conviene admitir que, efectivamente, uno de los objetivos de estas estrategias es reducir el tiempo que el profesor dedica a la evaluación para poder dedicar ese tiempo a otras tareas igualmente importante para el aprendizaje. Pero además, conviene tener a mano todo el repertorio de virtudes de estas estrategias que, bien planteadas, proyectarán ante los alumnos la idea de que el profesor ha pensado en lo que es

bueno para sus alumnos a la hora de establecer los elementos de su programa.

4. La evaluación y los niveles de competencia

Otro elemento importante a la hora de establecer un sistema de evaluación continuada es el análisis del nivel de competencia de los objetivos formativos que se van a evaluar. La taxonomía de referencia para el estudio del nivel de competencia de los objetivos formativos es la taxonomía de Bloom [3], según la cual un objetivo formativo pertenece a uno de los siguientes niveles, en orden creciente de competencia: *conocimiento, comprensión, aplicación, análisis, síntesis o evaluación*.

En [4] se hizo un trabajo de adaptación de la taxonomía de Bloom al ámbito de la enseñanza de la ingeniería, y se plantearon algunas reflexiones sobre el equilibrio que debe existir (y el desequilibrio que muchas veces existe) entre nivel de competencia, métodos docentes y métodos de evaluación.

En el contexto de este trabajo, en el que lo que nos interesa es la evaluación, proponemos una taxonomía simplificada, basada en tres niveles de competencia. A continuación se describen cada uno de estos niveles y se da un ejemplo de ejercicio (o tipo de ejercicio) de cada nivel, perteneciente al ámbito de la enseñanza de la programación de ordenadores.

- *Conocimiento*: Requiere que el alumno recuerde datos, hechos, información que le ha sido suministrada con anterioridad.
 1. Describe la sintaxis y la semántica del bucle “for” del lenguaje C
 2. Indica cuál es el rango de representación de los enteros de tipo short, en lenguaje C
- *Comprensión*: Requiere que el alumno utilice un determinado procedimiento sistemático (una “receta”) sobre un caso particular.
 3. Determinar el valor final de una variable después de ejecutar una secuencia de sentencias de lenguaje C
 4. Escribir el código en lenguaje C para ordenar, mediante el método de la burbuja, un vector de caracteres.

- *Aplicación*: Requiere que el alumno elija, de entre las “recetas” que ha comprendido previamente, la más apropiada para resolver un determinado problema.
 5. Determinar cuál es la estructura de datos más adecuada para resolver un problema
 6. Determinar qué operaciones de un determinado programa pueden ser realizadas mediante procedimientos y funciones de librerías ya existentes.

Es importante hacer notar que los nombres que Bloom da a los niveles de competencia muchas veces no corresponden directamente al significado de la palabra en castellano. Por tanto, muchos ejemplos que nosotros podamos considerar de comprensión o de aplicación no corresponden directamente con los respectivos niveles de competencia.

En la taxonomía simplificada que se propone, los niveles de conocimiento y comprensión se corresponden directamente con los dos primeros niveles de la taxonomía de Bloom. Sin embargo, el nivel de aplicación de la propuesta corresponde a una fusión de los niveles de aplicación y superiores de la taxonomía de Bloom. Consideramos que esta simplificación es válida en el contexto de asignaturas iniciales de primer ciclo, en las que rara vez se supera el nivel de aplicación en la taxonomía de Bloom.

Cuando el ejercicio que debe realizar el alumno es de nivel de conocimiento o de comprensión, la respuesta es única o admite muy pocas variaciones. Además, en el caso de comprensión, los resultados intermedios en el proceso de aplicar la “receta” también son únicos. Por ejemplo, en el caso del ejercicio 4, no hay en realidad muchas formas distintas de codificar el algoritmo de la burbuja para ordenar un vector de caracteres. En estas condiciones, la evaluación es muy sencilla: la solución del alumno debe asemejarse a la solución “oficial”, y cualquier diferencia es, en potencia, un error del alumno. Este principio es base para los esquemas de autoevaluación propuestos en la siguiente sección.

Cuando el ejercicio es de nivel de aplicación, diferentes alumnos pueden dar respuestas distintas, todas ellas válidas. Ahora no tiene sentido usar una solución “oficial” como base para la evaluación (aunque sí puede ser útil

Para cada uno de los ejercicios siguientes, debes preparar un informe de autoevaluación que se compone de dos partes:

- a) *Solución al ejercicio.*
- b) *Errores que nunca más volveré a cometer.*

Para escribir la parte (b) debes usar la solución oficial del ejercicio. Puesto que cada uno de los ejercicios tiene una solución única, cualquier diferencia entre tu solución y la solución oficial puede ser:

- Un error en la solución oficial (cosa poco probable).
- Un error tuyo.
- Una diferencia admisible entre las soluciones.

Compara tu solución con la oficial, identifica las diferencias y clasifica cada una de ellas según los tipos anteriores. Escribe tus conclusiones en la sección "Errores que nunca más volveré a cometer", de tu informe de autoevaluación.

Figura 1: Instrucciones para la autoevaluación.

como ejemplo). Lo que se necesita es explicitar los atributos que deben tener las soluciones correctas, de manera que la evaluación consiste en identificar los atributos propuestos en las respuestas a evaluar. Por ejemplo, en el caso del ejercicio 5, uno de los atributos de una solución correcta es que la estructura de datos propuesta ocupe un espacio razonable de memoria, o en otras palabras, que no haya soluciones alternativas que ocupen mucho menos espacio de memoria. Así pues, la identificación clara de los atributos de las soluciones correctas será el principio básico para el sistema de co-evaluación que se propone en la sección 6.

5. Autoevaluación y el nivel de comprensión

En esta sección se propone un esquema de autoevaluación que puede usarse cuando los objetivos formativos a evaluar corresponden al nivel de conocimiento o comprensión.

La idea básica es partir de una solución "oficial" al ejercicio, y basar la autoevaluación en la comparación de la propia solución con la solución "oficial". Puesto que, dada la naturaleza del ejercicio, no hay muchas variaciones válidas posibles en la respuesta, cualquier diferencia entre la respuesta del alumno y la solución "oficial" es un error potencial. Por tanto, la autoevaluación consiste en identificar las dife-

rencias, reconocer aquellas que corresponden a errores, y justificar aquellas que son variaciones admisibles. Las instrucciones para los alumnos podrían ser las mostradas en la figura 1. Es muy importante que el alumno entienda la importancia de hacer con rigor la comparación de la propia solución con la "oficial". Al realizar la autoevaluación por primera vez, es habitual que el alumno escriba, en la sección de *Errores que nunca más volveré a cometer*, cosas como: "Mi solución es claramente distinta de la oficial, pero creo que también es correcta". En un caso así, lo adecuado es devolver la autoevaluación al alumno y exigirle que realice correctamente la identificación y clasificación de las diferencias.

Los aspectos de logística para organizar el sistema de autoevaluación también son muy importantes. Se proponen, tres posibilidades:

1. Los alumnos no entregan las autoevaluaciones al profesor. Es un material de uso propio. Si bien esta estrategia cumple con el propósito perseguido (informar al alumno con prontitud, con una carga de trabajo baja para el profesor), no es recomendable porque:

- El profesor no tiene ocasión de intervenir para asegurarse que los alumnos hacen bien la autoevaluación (y las primeras veces, seguro que no la harán bien).
- El profesor no tiene información de cómo van los alumnos. No puede realizar accio-

nes correctoras, ni a nivel individual ni colectivo.

2. El profesor pide, de vez en cuando, los informes de autoevaluación. De esta forma, puede asegurarse de que los alumnos están haciendo bien la tarea, y tiene información sobre las dificultades de cada uno de los individuos, y de las dificultades generales de la clase. Idealmente, el profesor explica a sus alumnos, al inicio del curso, que uno de sus objetivos para el curso es que aprendan a identificar sus propios errores, de manera que calificará los informes de autoevaluación (en función de si se han identificado bien los errores), y usará esa calificación para obtener parte de la nota final del alumno. Naturalmente, esta opción implica algo más de trabajo para el profesor. La opción admite dos variantes:

- De vez en cuando, el profesor recoge el informe de autoevaluación de todos los alumnos, correspondiente a un ejercicio.
- De vez en cuando, el profesor recoge todos los informes de autoevaluación de un determinado alumno o grupo de alumnos.

La primera opción tiene la ventaja de que el profesor puede tener una idea clara de las dificultades generales del grupo con un determinado tipo de ejercicios. Tiene el inconveniente de que se trabaja con muchas hojas sueltas, cosa que puede resultar incómoda. Con la segunda opción, el alumno entrega al profesor una carpeta bien organizada con sus informes de autoevaluación (y, posiblemente, otros materiales del curso). Además, al analizar esta carpeta, el profesor puede obtener más fácilmente una visión nítida de la evolución del alumno (hasta qué punto su rendimiento va mejorando). Otra ventaja adicional es que el profesor puede pedir la carpeta con más frecuencias a aquellos alumnos que necesitan más ayuda.

3. Además de calificar la calidad de la autoevaluación (como en la opción anterior) el profesor usa la calificación que se asignan los alumnos para obtener la nota final de la asignatura. En este caso, en la guía para la autoevaluación, además de la solución oficial deben aparecer los criterios precisos para asignar una calificación a la solución. La ventaja de esta opción es que el alumno tiene una mayor responsabilidad en el proceso, y tiende a tomarse más en serio la autoevaluación, al ver que hay un impacto direc-

to mayor en la calificación final. Sin embargo, también hay inconvenientes importantes: el alumno puede centrarse más en la nota que en la identificación de errores, la elaboración de las instrucciones para la autoevaluación se complica², y el profesor debe, probablemente, dedicar más tiempo a analizar los informes de autoevaluación. Por otra parte, no es una alternativa apropiada cuando la organización y contenidos del curso no recomiendan el asignar calificaciones hasta la fase final en la que los alumnos están en condiciones óptimas para ser calificados.

Cualquiera que sea el esquema adoptado, es muy importante tener en cuenta que los alumnos necesitan repetir el proceso con una cierta frecuencia para aprender a autoevaluarse, y a sacar provecho de la actividad. Las experiencias puntuales (por ejemplo, un único ejercicio de autoevaluación en todo el curso) suelen generar más frustración y desorientación que otra cosa, y sólo son recomendables si se trata de probar la mecánica, de cara a una implementación generalizada en el futuro.

6. Co-evaluación y el nivel de aplicación

Cuando se trata ejercicios de nivel de aplicación, una solución “oficial” no es suficiente para realizar la evaluación, puesto que el ejercicio admite varias soluciones correctas y distintas entre sí. Lo importante es explicitar las características que debe tener una determinada solución para que pueda ser considerada correcta.

Una forma de presentar a los alumnos las características de las soluciones correctas es usar rúbricas [5] como las que se muestra en la figura 2. En este caso, se trata de evaluar el código para resolver un determinado problema de programación. En la columna de la izquierda se identifican diferentes criterios de evaluación del

² Muchos profesores insisten en lo complicado que puede resultar el explicitar los criterios para que el alumno pueda calcular una calificación precisa. Sin embargo, eso no es más difícil que explicitar los criterios de corrección que debe usar un grupo de profesores que se reparten la corrección de un ejercicio del examen final, cosa que hacemos de forma habitual.

Código			
Criterio	Nivel de calidad		
	3 Notable	2 Suficiente	1 Insuficiente
Correcto	La aplicación funciona bien en todos los casos. No he encontrado ningún fallo.	Hay (como máximo) un par de situaciones en las que el programa no ha funcionado bien.	La aplicación falla constantemente.
Robusto	La aplicación resiste sin bloquearse todos los errores típicos que puede cometer un usuario "poco hábil". No he conseguido que se cuelgue.	Es razonablemente robusto. No es fácil que se quede colgado, pero en uno o dos casos se bloqueó.	La aplicación no es robusta en absoluto. Se queda colgada con frecuencia ante errores típicos del usuario al entrar datos.
Amigable	El usuario no tiene ninguna duda, en ningún momento, sobre cómo interactuar con la aplicación, qué datos debe suministrar y cómo interpretar los resultados y mensajes de la aplicación.	Los mensajes e información que da la aplicación son suficientes para trabajar bien. Sin embargo, en alguna ocasión he tenido algunas dudas sobre lo que hay que hacer o cómo hay que hacerlo.	El usuario tiene dudas constantes sobre lo que le está pidiendo la aplicación, y es difícil interpretar los resultados y mensajes en pantalla.
Comparado con el nuestro	Este código es mejor.	Este código es similar.	Este código es peor.

Figura 2: Ejemplo de rúbrica para la evaluación de un código

código. En las columnas de la derecha se identifican, para cada uno de los criterios de evaluación, las características que debe reunir una determinada solución para que pueda ser considerada *notable*, *suficiente* o *insuficiente* (en este ejemplo, se consideran sólo tres niveles de calidad). La tarea del evaluador es identificar en la solución que tiene que evaluar, las características señaladas en la rúbrica, y determinar el nivel de calidad para cada uno de los criterios de evaluación. El informe de evaluación se completa con los argumentos en los que el evaluador se basa para tomar la decisión, como por ejemplo:

- El código falla en los casos de prueba 1 y 4.
- El mensaje que avisa al usuario para que introduzca datos no es claro: tengo dudas sobre el formato que debo usar.
- El código está muy mal indentado. Tengo muchas dificultades para identificar los bloques de código.

Sin duda, la evaluación del nivel de aplicación es más difícil de objetivizar que la evaluación del nivel de conocimiento o comprensión. En el caso, por ejemplo, del argumento (b) de la lista anterior, dos personas distintas pueden tener opiniones diferentes sobre la claridad del mensaje que el programa ofrece al usuario. Además, con toda probabilidad, el autor del código opina-

rá que su mensaje es claro. Por estas razones, la evaluación del nivel de aplicación no acaba de combinarse bien con la estrategia de autoevaluación (cuando se trata de cosas opinables, no somos buenos jueces de nosotros mismos).

En cambio, la estrategia de la co-evaluación se adapta mejor a la naturaleza de este tipo de evaluación. Un alumno puede juzgar de forma más neutral el resultado del trabajo de otros compañeros, naturalmente con la ayuda de una rúbrica bien elaborada. Además, su propia solución es un punto de referencia importante para valorar los méritos y deméritos de las soluciones de los compañeros. Esa es la razón de que, en la rúbrica que se muestra como ejemplo, la última fila haga referencia a una comparación entre la solución propia y la evaluada.

Los aspectos logísticos de la co-evaluación son más complicados que los de la autoevaluación, simplemente porque los resultados de los trabajos de los alumnos deben ser recogidos y redistribuidos para ser evaluados por otros. Las soluciones a este problema dependerán mucho de las circunstancias de cada caso, por lo que poco más puede decirse aquí al respecto con carácter general.

Los comentarios que se han hecho en el apartado anterior en cuanto al uso que hace el

profesor de los resultados de la autoevaluación también son aplicables en el caso de la co-evaluación. En todo caso, conviene ser consciente de que los alumnos no suelen mostrarse favorables a asignar a los compañeros calificaciones que puedan afectar a la nota final. Por tanto, si se opta por esta alternativa, es importante tener a punto el argumento para convencerles de que la habilidad de emitir críticas constructivas, y juzgar el trabajo de los compañeros es importante para el ejercicio profesional.

7. Conclusión

Las técnicas que se plantean de forma teórica en este trabajo se están poniendo en práctica en la asignatura Laboratorio de Programación (LP), perteneciente a la Ingeniería Técnica en Telecomunicaciones, de la Escuela Politécnica Superior de Castelldefels (UPC). En realidad, han sido las experiencias en esta asignatura las que nos han ayudado a poner en orden nuestras ideas, en la forma presentada aquí.

La asignatura ofrece un escenario ideal. En la primera parte, los objetivos formativos son de nivel de comprensión. Los alumnos deben escribir códigos en Visual C++ que resuelvan casos particulares mediante algoritmos conocidos. En esta primera parte se utiliza la técnica de la autoevaluación, los informes de autoevaluación han de incluir una justificación correcta, no sólo de los errores, sino de las diferencias en general y esta nota se usa en el cálculo de la nota final.

En la segunda parte de la asignatura los objetivos son de nivel de aplicación. Los alumnos deben tomar decisiones relativas a las estructuras de datos y algoritmos que hay que usar para resolver un determinado problema. Las actividades se realizan en modo proyecto, de forma que los alumnos, trabajando en grupo, deben resolver un problema. La técnica de la co-evaluación basada en rúbricas se usa para que cada grupo evalúe el trabajo de otros.

No es objetivo de este trabajo dar detalles de implementación, ni analizar los resultados obtenidos para una asignatura concreta. La orientación teórica que hemos dado al artículo es más general, y trata de mostrar la separación entre Evaluación Formativa y Evaluación Calificativa como una nueva perspectiva al modelo enseñanza-aprendizaje que puede aplicarse a

diferentes asignaturas. En todo caso, sí pueden apuntarse algunas conclusiones, obtenidas a partir de encuestas de opinión (especialmente sobre la autoevaluación, con la que llevamos más tiempo experimentando):

- Los alumnos perciben la autoevaluación como algo positivo. En particular, creen que les permite mantenerse puntualmente informados.
- Los alumnos aprenden pronto la mecánica. El porcentaje de autoevaluaciones que deben ser corregidas es pequeño.
- El tiempo de dedicación de los profesores a la supervisión es asumible. En el caso de la asignatura LP, estamos hablando de 1 hora a la semana, para un grupo de 40 alumnos.
- En versiones previas de la asignatura, la nota de autoevaluación contaba para la nota final. Esto producía que los alumnos pusieran mucho énfasis en cálculo de la nota, y menos en la identificación de los errores cometidos. Actualmente esta nota no cuenta para la nota final y se da más peso a la calidad del informe en cuanto a la justificación de diferencias (no solo de los errores) entre la solución del estudiante y la solución oficial.

Referencias

- [1] A. W. Chickering y Z. F. Gamson, Seven Principles for Good Practice in Undergraduate Education, <http://www.hcc.hawaii.edu/intranet/committees/FacDevCom/guidebk/teachtip/7princip.htm>
- [2] A.W. Bangert, Peer Assessment: A Win-Win Instructional Strategy for Both Students and Teachers, *J. Cooperation & Collaboration in College Teaching*, Vol. 10, No. 2, p. 77.
- [3] B.S. Bloom et al, Taxonomy of Educational Objectives: Handbook I, Cognitive Domain. Nueva York: David McKay, 1956.
- [4] M. Valero-García y J.J. Navarro, Niveles de competencia de los objetivos formativos en las ingenierías, *VII Jornadas sobre la Enseñanza Universitaria de la Informática JENUI 2001*, p. 149
- [5] Ideas and Rubric, Instructional Intranet, Chicago Public Schools. http://intranet.cps.k12.il.us/Assessments/Ideas_and_Rubrics/ideas_and_rubrics.html

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal. ISSN 1531-7714

Copyright 2001, PAREonline.net.

Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited. Please notify the editor if an article is to be used in a newsletter.

Mertler, Craig A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25). Retrieved January 9, 2010 from <http://PAREonline.net/getvn.asp?v=7&n=25> . This paper has been viewed 182,905 times since 12/11/2001.

Designing Scoring Rubrics for Your Classroom

Craig A. Mertler

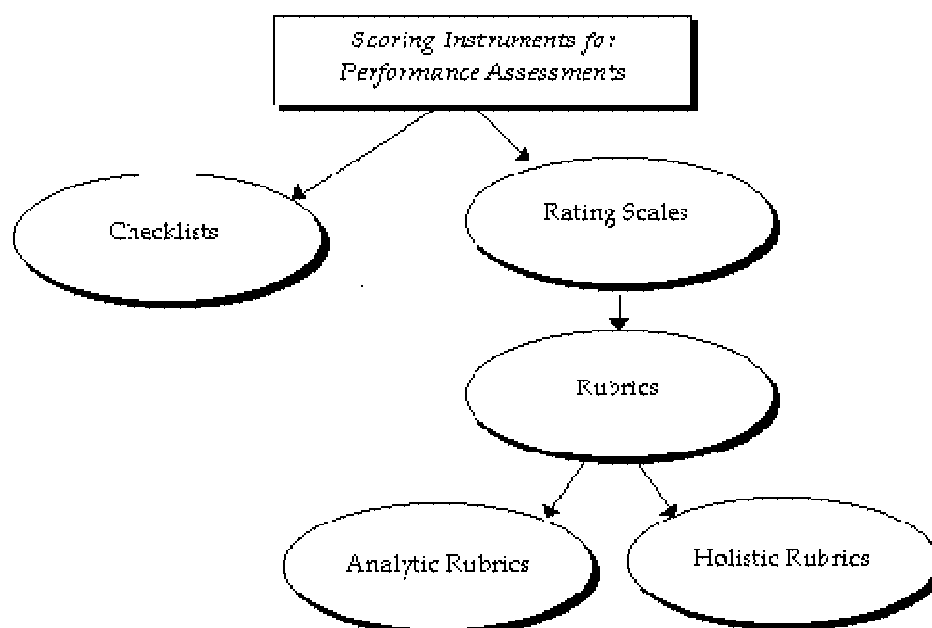
Bowling Green State University

Rubrics are rating scales—as opposed to checklists—that are used with performance assessments. They are formally defined as scoring guides, consisting of specific pre-established performance criteria, used in evaluating student work on performance assessments. Rubrics are typically the specific form of scoring instrument used when evaluating student performances or products resulting from a performance task.

There are two types of rubrics: holistic and analytic (see Figure 1). A **holistic rubric** requires the teacher to score the overall process or product as a whole, without judging the component parts separately (Nitko, 2001). In contrast, with an **analytic rubric**, the teacher scores separate, individual parts of the product or performance first, then sums the individual scores to obtain a total score (Moskal, 2000; Nitko, 2001).

Figure 1:

Types of scoring instruments for performance assessments



Holistic rubrics are customarily utilized when errors in some part of the process can be tolerated provided the overall quality is high (Chase, 1999). Nitko (2001) further states that use of holistic rubrics is probably more appropriate when performance tasks require students to create some sort of response and where there is no definitive correct answer. The focus of a score reported using a holistic rubric is on the overall quality, proficiency, or understanding of the specific content and skills-it involves assessment on a unidimensional level (Mertler, 2001). Use of holistic rubrics can result in a somewhat quicker scoring process than use of analytic rubrics (Nitko, 2001). This is basically due to the fact that the teacher is required to read through or otherwise examine the student product or performance only once, in order to get an "overall" sense of what the student was able to accomplish (Mertler, 2001). Since assessment of the overall performance is the key, holistic rubrics are also typically, though not exclusively, used when the purpose of the performance assessment is summative in nature. At most, only limited feedback is provided to the student as a result of scoring performance tasks in this manner. A template for holistic scoring rubrics is presented in Table 1.

<u>Score</u>	<u>Description</u>
5	Demonstrates complete understanding of the problem. All requirements of task are included in response.
4	Demonstrates considerable understanding of the problem. All requirements of task are included.
3	Demonstrates partial understanding of the problem. Most requirements of task are included.
2	Demonstrates little understanding of the problem. Many requirements of task are missing.
1	Demonstrates no understanding of the problem.
0	No response/task not attempted.

Analytic rubrics are usually preferred when a fairly focused type of response is required (Nitko, 2001); that is, for performance tasks in which there may be one or two acceptable responses and creativity is not an essential feature of the students' responses. Furthermore, analytic rubrics result initially in several scores, followed by a summed total score-their use represents assessment on a multidimensional level (Mertler, 2001). As previously mentioned, the use of analytic rubrics can cause the scoring process to be substantially slower, mainly because assessing several different skills or characteristics individually requires a teacher to examine the product several times. Both their construction and use can be quite time-consuming. A general rule of thumb is that an individual's work should be examined a separate time for each of the specific performance tasks or scoring criteria (Mertler, 2001). However, the advantage to the use of analytic rubrics is quite substantial. The degree of feedback offered to students-and to teachers-is significant. Students receive specific feedback on their performance with respect to each of the individual scoring criteria-something that does not happen when using holistic rubrics (Nitko, 2001). It is possible to then create a "profile" of specific student strengths and weaknesses (Mertler, 2001). A template for analytic scoring rubrics is presented in Table 2.

Table 2:
Template for analytic rubrics

	Beginning 1	Developing 2	Accomplished 3	Exemplary 4	Score
Criteria #1	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	
Criteria #2	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	
Criteria #3	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	
Criteria #4	Description reflecting beginning level of performance	Description reflecting movement toward mastery level of performance	Description reflecting achievement of mastery level of performance	Description reflecting highest level of performance	

Prior to designing a specific rubric, a teacher must decide whether the performance or product will be scored holistically or analytically (Airasian, 2000 & 2001). Regardless of which type of rubric is selected, specific performance criteria and observable indicators must be identified as an initial step to development. The decision regarding the use of a holistic or analytic approach to scoring has several possible implications. The most important of these is that teachers must consider first how they intend to use the results. If an overall, summative score is desired, a holistic scoring approach would be more desirable. In contrast, if formative feedback is the goal, an analytic scoring rubric should be used. It is important to note that one type of rubric is not inherently better than the other—you must find a format that works best for your purposes (Montgomery, 2001). Other implications include the time requirements, the nature of the task itself, and the specific performance criteria being observed.

As you saw demonstrated in the templates (Tables 1 and 2), the various levels of student performance can be defined using either quantitative (i.e., numerical) or qualitative (i.e., descriptive) labels. In some instances, teachers might want to utilize both quantitative and qualitative labels. If a rubric contains four levels of proficiency or understanding on a continuum, quantitative labels would typically range from "1" to "4." When using qualitative labels, teachers have much more flexibility, and can be more creative. A common type of qualitative scale might include the following labels: master, expert, apprentice, and novice. Nearly any type of qualitative scale will suffice, provided it "fits" with the task.

One potentially frustrating aspect of scoring student work with rubrics is the issue of somehow converting them to "grades." It is not a good idea to think of rubrics in terms of percentages (Trice, 2000). For example, if a rubric has six levels (or "points"), a score of 3 should not be equated to 50% (an "F" in most letter grading systems). The process of converting rubric scores to grades or categories is more a process of logic than it is a mathematical one. Trice (2000) suggests that in a rubric scoring system, there are typically more scores at the average and above average categories (i.e., equating to grades of "C" or better) than there are below average categories. For instance, if a rubric consisted of nine score categories, the equivalent grades and categories might look like this:

Table 3:
Sample grades and categories

<i>Rubric Score</i>	<i>Grade</i>	<i>Category</i>
8	A+	Excellent
7	A	Excellent
6	B+	Good
5	B	Good
4	C+	Fair
3	C	Fair
2	U	Unsatisfactory
1	U	Unsatisfactory
0	U	Unsatisfactory

When converting rubric scores to grades (typical at the secondary level) or descriptive feedback (typical at the elementary level), it is important to remember that there is not necessarily one correct way to accomplish this. The bottom line for classroom teachers is that they must find a system of conversion that works for them and fits comfortably into their individual system of reporting student performance.

Steps in the Design of Scoring Rubrics

A step-by-step process for designing scoring rubrics for classroom use is presented below. Information for these procedures was compiled from various sources (Airasian, 2000 & 2001; Mertler, 2001; Montgomery, 2001; Nitko, 2001; Tombari & Borich, 1999). The steps will be summarized and discussed, followed by presentations of two sample scoring rubrics.

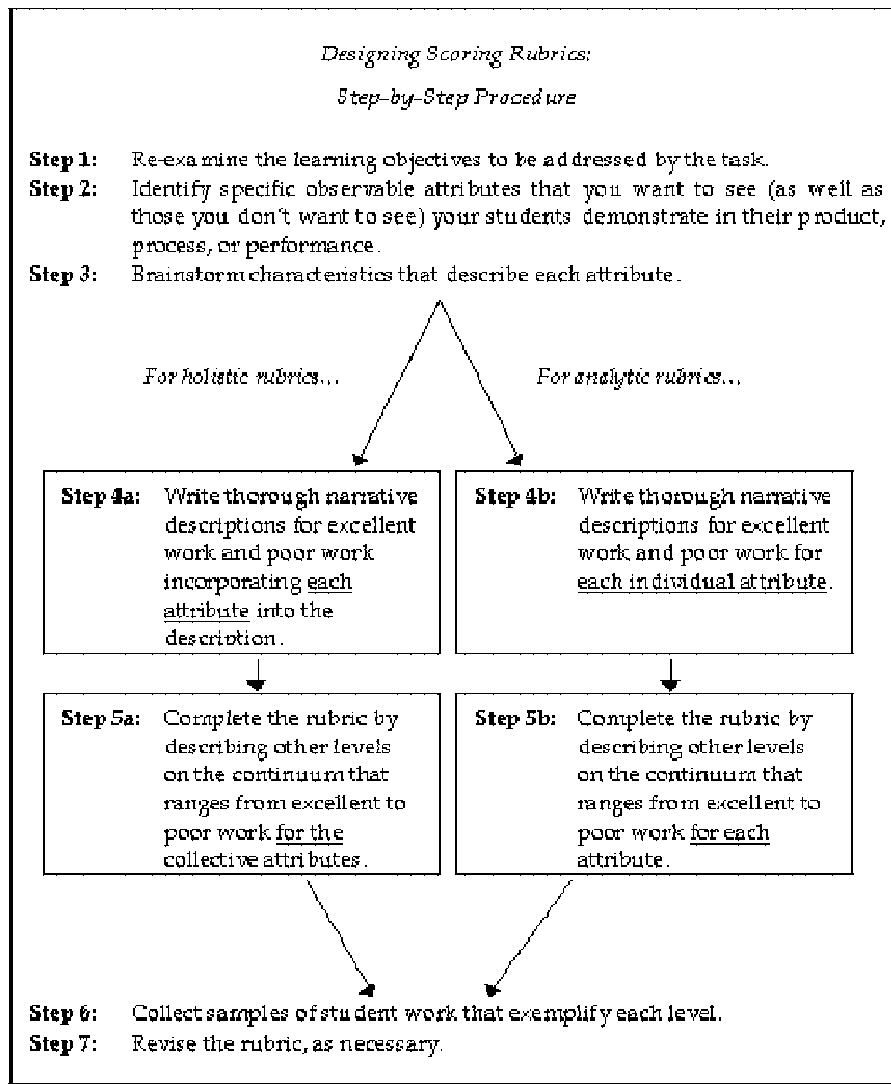
- Step 1:** *Re-examine the learning objectives to be addressed by the task.* This allows you to match your scoring guide with your objectives and actual instruction.
- Step 2:** *Identify specific observable attributes that you want to see (as well as those you don't want to see) your students demonstrate in their product, process, or performance.* Specify the characteristics, skills, or behaviors that you will be

looking for, as well as common mistakes you do not want to see.

- Step 3:** *Brainstorm characteristics that describe each attribute.* Identify ways to describe above average, average, and below average performance for each observable attribute identified in Step 2.
- Step 4a:** *For holistic rubrics, write thorough narrative descriptions for excellent work and poor work incorporating each attribute into the description.* Describe the highest and lowest levels of performance combining the descriptors for all attributes.
- Step 4b:** *For analytic rubrics, write thorough narrative descriptions for excellent work and poor work for each individual attribute.* Describe the highest and lowest levels of performance using the descriptors for each attribute separately.
- Step 5a:** *For holistic rubrics, complete the rubric by describing other levels on the continuum that ranges from excellent to poor work for the collective attributes.* Write descriptions for all intermediate levels of performance.
- Step 5b:** *For analytic rubrics, complete the rubric by describing other levels on the continuum that ranges from excellent to poor work for each attribute.* Write descriptions for all intermediate levels of performance for each attribute separately.
- Step 6:** *Collect samples of student work that exemplify each level.* These will help you score in the future by serving as benchmarks.
- Step 7:** *Revise the rubric, as necessary.* Be prepared to reflect on the effectiveness of the rubric and revise it prior to its next implementation.

These steps involved in the design of rubrics have been summarized in Figure 2.

Figure 2:
Designing Scoring Rubrics: Step-by-step procedures



Two Examples

Two sample scoring rubrics corresponding to specific performance assessment tasks are presented next. Brief discussions precede the actual rubrics. For illustrative purposes, a holistic rubric is presented for the first task and an analytic rubric for the second. It should be noted that either a holistic or an analytic rubric could have been designed for either task.

*Example 1:
Subject - Mathematics
Grade Level(s) - Upper Elementary*

Mr. Harris, a fourth-grade teacher, is planning a unit on the topic of data analysis, focusing primarily on the skills of estimation and interpretation of graphs. Specifically, at the end of this unit, he wants to be able to assess his students' mastery of the following instructional objectives:

- Students will properly interpret a bar graph.
- Students will accurately estimate values from within a bar graph. (step 1)

Since the purpose of his performance task is summative in nature - the results will be incorporated into the students' grades, he decides to develop a holistic rubric. He identifies the following four attributes on which to focus his rubric: estimation, mathematical computation, conclusions, and communication of explanations (steps 2 & 3). Finally, he begins drafting descriptions of the various levels of performance for the observable attributes (steps 4 & 5). The final rubric for his task appears in Table 4.

Table 4:
Math Performance Task – Scoring Rubric
Data Analysis

<u>Score</u>	<u>Description</u>
4	Makes accurate estimations. Uses appropriate mathematical operations with no mistakes. Draws logical conclusions supported by graph. Sound explanations of thinking.
3	Makes good estimations. Uses appropriate mathematical operations with few mistakes. Draws logical conclusions supported by graph. Good explanations of thinking.
2	Attempts estimations, although many inaccurate. Uses inappropriate mathematical operations, but with no mistakes. Draws conclusions not supported by graph. Offers little explanation.
1	Makes inaccurate estimations. Uses inappropriate mathematical operations. Draws no conclusions related to graph. Offers no explanations of thinking.
0	No response/task not attempted.

Example 2:
Subjects - Social Studies; Probability & Statistics
Grade Level(s) - 9 - 12

Mrs. Wolfe is a high school American government teacher. She is beginning a unit on the electoral process and knows from past years that her students sometimes have difficulty with the concepts of sampling and election polling. She decides to give her students a performance assessment so they can demonstrate their levels of understanding of these concepts. The main idea that she wants to focus on is that samples (surveys) can accurately predict the viewpoints of an entire population. Specifically, she wants to be able to assess her students on the following instructional objectives:

- Students will collect data using appropriate methods.
- Students will accurately analyze and summarize their data.
- Students will effectively communicate their results. (step 1)

Since the purpose of this performance task is formative in nature, she decides to develop an analytic rubric focusing on the following attributes: sampling technique, data collection, statistical analyses, and communication of results (steps 2 & 3). She drafts descriptions of the various levels of performance for the observable attributes (steps 4 & 5). The final rubric for this task appears in Table 5.

Table 5:
Performance Task – Scoring Rubric
Population Sampling

Name _____		Date _____			
	Beginning 1	Developing 2	Accomplished 3	Exemplary 4	Score
Sampling Technique	Inappropriate sampling technique used	Appropriate technique used to select sample; major errors in execution	Appropriate technique used to select sample; minor errors in execution	Appropriate technique used to select sample; no errors in procedures	
Survey/ Interview Questions	Inappropriate questions asked to gather needed information	Few pertinent questions asked; data on sample is inadequate	Most pertinent questions asked; data on sample is adequate	All pertinent questions asked; data on sample is complete	
Statistical Analyses	No attempt at summarizing collected data	Attempts analysis of data, but inappropriate procedures	Proper analytical procedures used, but analysis incomplete	All proper analytical procedures used to summarize data	
Communication of Results	Communication of results is incomplete, unorganized, and difficult to follow	Communicates some important information; not organized well enough to support decision	Communicates most of important information; shows support for decision	Communication of results is very thorough; shows insight into how data predicted outcome	
Total Score = _____					

Resources for Rubrics on the Web

The following is just a partial list of some Web resources for information about and samples of scoring rubrics.

- "Scoring Rubrics: What, When, & How?" (<http://pareonline.net/getvn.asp?v=7&n=3>). This article appears in Practical Assessment, Research, & Evaluation and is authored by Barbara M. Moskal. The article discusses what rubrics are, and distinguishes between holistic and analytic types. Examples and additional resources are provided.
- "Performance Assessment-Scoring" (<http://www.pgcps.pg.k12.md.us/~elc/scoringtasks.html>). Staff in the Prince George's County (MD) Public Schools have developed a series of pages that provide descriptions of the steps involved in the design of performance tasks. This particular page provides several rubric samples.
- "Rubrics from the Staff Room for Ontario Teachers" (<http://www.quadro.net/~ecoxon/Reporting/rubrics.htm>) This site is a collection of literally hundreds of teacher-developed rubrics for scoring performance tasks. The rubrics are categorized by subject area and type of task. This is a fantastic resource...check it out!
- "Rubistar Rubric Generator" (<http://rubistar.4teachers.org/>)
- "Teacher Rubric Maker" (http://www.teach-nology.com/web_tools/rubrics/) These two sites house Web-based rubric generators for teachers. Teachers can customize their own rubrics based on templates on each site. In both cases, rubric templates are organized by subject area and/or type of performance task. These are wonderful resources for teachers!

References

<http://pareonline.net/getvn.asp?v=7&n=25>

- Airasian, P. W. (2000). *Assessment in the classroom: A concise approach* (2nd ed.). Boston: McGraw-Hill.
- Airasian, P. W. (2001). *Classroom assessment: Concepts and applications* (4th ed.). Boston: McGraw-Hill.
- Chase, C. I. (1999). *Contemporary assessment for educators*. New York: Longman.
- Mertler, C. A. (2001). Using performance assessment in your classroom. Unpublished manuscript, Bowling Green State University.
- Montgomery, K. (2001). *Authentic assessment: A guide for elementary teachers*. New York: Longman.
- Moskal, B. M. (2000). Scoring rubrics: what, when, and how?. *Practical Assessment, Research, & Evaluation*, 7(3). Available online: <http://pareonline.net/getvn.asp?v=7&n=3>
- Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Upper Saddle River, NJ: Merrill.
- Tombari, M. & Borich, G. (1999). *Authentic assessment in the classroom: Applications and practice*. Upper Saddle River, NJ: Merrill.
- Trice, A. D. (2000). *A handbook of classroom assessment*. New York: Longman.

Contact information

Craig A. Mertler
Educational Foundations & Inquiry Program
College of Education & Human Development
Bowling Green State University
Bowling Green, OH 43403

mertler@bgnet.bgsu.edu
Phone: 419-372-9357 Fax: 419-372-8265

Descriptors: *Rubrics; Scoring; *Student Evaluation; *Test Construction; *Evaluation Methods; Grades; Grading;
*Scoring

Understanding Rubrics

by Heidi Goodrich Andrade

Authentic assessments tend to use *rubrics* to describe student achievement. At last, here's clarity on the term.

Every time I introduce rubrics to a group of teachers the reaction is the same – instant appeal (“Yes, this is what I need!”) followed closely by panic (“Good grief, how can I be expected to develop a rubric for everything?”). When you learn what rubrics do – and why – you can create and use them to support and assess student learning without losing your sanity.

What Is a Rubric?

A rubric is a scoring tool that lists the criteria for a piece of work, or “what counts” (for example, purpose, organization, details, voice, and mechanics are often what count in a piece of writing); it also articulates gradations of quality for each criterion, from excellent to poor. The term defies a dictionary definition, but it seems to have established itself, so I continue to use it.

The example in Figure 1 (adapted from Perkins et al 1994) lists the criteria and gradations of quality for verbal, written, or graphic reports on student inventions – for instance, inventions designed to ease the Westward journey for 19th century pioneers for instance, or to solve a local environmental problem, or to represent an imaginary culture and its inhabitants, or anything else students might invent.

This rubric lists the criteria in the column on the left: The report must explain (1) the purposes of the invention, (2) the features or parts of the invention and how they help it serve its purposes, (3) the pros and cons of the design, and (4) how the design connects to other things past, present, and future. The rubric could easily include criteria related to presentation style and effectiveness, the mechanics of written pieces, and the quality of the invention itself.

The four columns to the right of the criteria describe varying degrees of quality, from excellent to poor. As concisely as possible, these columns explain what makes a good piece of work good and a bad one bad.

Figure 1

Rubric for an Invention Report

Criteria	Quality			
Purposes	The report explains the key purposes of the invention and points out less obvious ones as well.	The report explains all of the key purposes of the invention.	The report explains some of the purposes of the invention but misses key purposes.	The report does not refer to the purposes of the invention.
Features	The report details both key and hidden features of the invention and explains how they serve several purposes.	The report details the key features of the invention and explains the purposes they serve.	The report neglects some features of the invention or the purposes they serve.	The report does not detail the features of the invention or the purposes they serve.
Critique	The report discusses the strengths and weaknesses of the invention, and suggests ways in which it can be improved.	The report discusses the strengths and weaknesses of the invention.	The report discusses either the strengths or weaknesses of the invention but not both.	The report does not mention the strengths or the weaknesses of the invention.
Connections	The report makes appropriate connections between the purposes and features of the invention and many different kinds of phenomena.	The report makes appropriate connections between the purposes and features of the invention and one or two phenomena.	The report makes unclear or inappropriate connections between the invention and other phenomena.	The report makes no connections between the invention and other things.

Why Use Rubrics?

Rubrics appeal to teachers and students for many reasons. First, they are powerful tools for both teaching and assessment. Rubrics can improve student performance, as well as monitor it, by making teachers' expectations clear and by showing students how to meet these expectations. The result is often marked improvements in the quality of student work and in learning. Thus, the most common argument for using rubrics is they help define "quality." One student actually *didn't* like rubrics for this very reason: "If you get something wrong," she said, "your teacher can prove you knew what you were supposed to do!" (Marcus 1995).

A second reason that rubrics are useful is that they help students become more thoughtful judges of the quality of their own and others' work. When rubrics are used to guide self- and peer-assessment, students become increasingly able to spot and solve problems in their own and one another's work. Repeated practice with peer-assessment, and especially self-assessment, increases students' sense of responsibility for their own work and cuts down on the number of "Am I done yet?" questions.

Third, rubrics reduce the amount of time teachers spend evaluating student work. Teachers tend to find that by the time a piece has been self- and peer-assessed according to a rubric, they have little left to say about it. When they do have something to say, they can often simply circle an item in the rubric, rather than struggling to explain the flaw or strength they have noticed and figuring out what to suggest in terms of improvements. Rubrics provide students with more informative feedback about their strengths and areas in need of improvement.

Fourth, teachers appreciate rubrics because their "accordion" nature allows them to accommodate heterogeneous classes. The examples here have three or four gradations of quality, but there is no reason they can't be "stretched" to reflect the work of both gifted and those with learning disabilities.

Finally, rubrics are easy to use and to explain. Christine Hall, a fourth grade teacher, reflected on how both students and parents responded to her use of rubrics:

Students were able to articulate what they had learned, and by the end of the year could be accurate with their evaluations. Parents were very excited about the use of rubrics. During parent conferences I used sample rubrics to explain to parents their purpose, and how they were used in class. The reaction of parents was very encouraging. They knew exactly what their child needed to do to be successful.

How Do You Create Rubrics?

Rubrics are becoming increasingly popular with educators moving toward more authentic, performance-based assessments. Recent publications contain some rubrics (Brewer 1996; Marzano et al 1993). Chances are, however, that you will have to develop a few to reflect your own curriculum and teaching style. To boost the learning leverage of rubrics, the rubric design process should engage students in the following steps:

1. *Look at models:* Show students examples of good and not-so-good work. Identify the characteristics that make the good ones good and the bad ones bad.
2. *List criteria:* Use the discussion of models to begin a list of what counts in quality work.
3. *Articulate gradations of quality:* Describe the best and worst levels of quality, then fill in the middle levels based on your knowledge of common problems and the discussion of not-so-good work.
4. *Practice on models:* Have students use the rubrics to evaluate the models you gave them in Step 1.
5. *Use self- and peer-assessment:* Give students their task. As they work, stop them occasionally for self- and peer-assessment.
6. *Revise:* Always give students time to revise their work based on the feedback they get in Step 5.
7. *Use teacher assessment:* Use the same rubric students used to assess their work yourself.

Step 1 may be necessary only when you are asking students to engage in a task with which they are unfamiliar. Steps 3 and 4 are useful but time-consuming; you can do these on your own, especially when you've been using rubrics for a while. A class experienced in rubric-based assessment can streamline the process so that it begins with listing criteria, after which the teacher writes out the gradations of quality, checks them with the students, makes revisions, then uses the rubric for self-, peer-, and teacher assessment.

Ann Tanona, a second grade teacher, went through the seven-step process with her students. The result was a rubric for assessing videotaped *Reading Rainbow*-style "book talks" (fig. 2).

Figure 2

Book Talk Rubric

Criteria	Quality		
Did I get my audience's attention?	Creative beginning	Boring beginning	No beginning
Did I tell what kind of book?	Tells exactly what type of book it is	Not sure, not clear	Didn't mention it
Did I tell something about the main character?	Included facts about character	Slid over character	Did not tell anything about main character
Did I mention the setting?	Tells when and where story takes place	Not sure, not clear	Didn't mention setting
Did I tell one interesting part?	Made it sound interesting — I want to buy it!	Told part and skipped on to something else	Forgot to do it
Did I tell who might like this book?	Did tell	Skipped over it	Forgot to tell
How did I look?	Hair combed, neat, clean clothes, smiled, looked up, happy	Lazy look	Just-got-out-of-bed look, head down
How did I sound?	Clear, strong, cheerful voice	No expression in voice	Difficult to understand— 6-inch voice or screeching

Tips on Designing Rubrics

Ann's rubric is powerful because it articulates the characteristics of a good "book talk" in students' own words. It also demonstrates some of the difficulties of designing a good rubric.

Perhaps the most common challenge is avoiding unclear language, such as "creative beginning." If a rubric is to teach as well as evaluate, terms like these must be defined for students. Admittedly, *creative* is a difficult word to define. Ann handled this problem by having a discussion of what the term "creative beginning" meant in the book talks.

Patricia Crosby and Pamela Heinz, both seventh grade teachers, solved the same problem in a rubric for oral presentations by actually listing ways in which students could meet the criterion (fig. 3). This approach provides valuable information to students on how to begin a talk and avoids the need to define elusive terms like *creative*.

Figure 3

Rubric for an Oral Presentation

Criterion	Quality		
Gains attention of audience.	Gives details or an amusing fact, a series of questions, a short demonstration, a colorful visual or a personal reason why they picked this topic.	Does a two-sentence introduction, then starts speech. Gives a one-sentence introduction, then starts speech.	Does not attempt to gain attention of audience, just starts speech.

A second challenge in rubric design is avoiding unnecessarily negative language. The excerpt from the rubric in Figure 3 avoids words like *boring* by describing what was done during a so-so beginning to a talk and implicitly comparing it with the highest level of quality. Thus, students know exactly what they did wrong and how they can do better next time, not just that the opening to their talk was boring.

Articulating gradations of quality is often a challenge. It helps if you spend a lot of time thinking about criteria and how best to chunk them before going on to define the levels of quality. You might also try a clever technique I have borrowed from a fifth grade teacher in Gloucester, Massachusetts. She describes gradations of quality as: "Yes," "Yes but," "No but," and "No." For example, Figure 4 shows part of a rubric for evaluating a scrapbook that documents a story. This approach tends to work well, as long as you aren't too rigid about it. Rigidity can have amusing results: One student wrote out the lowest level of quality for the criterion, "Is it anachronism free?" this way: "No, I did not remember to not use anachronism"!

Figure 4

Rubric for Evaluating a Scrapbook

Criterion	Quality				
Gives enough details.	Yes, I put in enough details to give the reader a sense of time, place, and events.	Yes, I put in some details, but some key details are missing.	No, I didn't put in enough details, but I did include a few.	No, I had almost no details.	

What to Do Once You've Created Rubrics

Creating rubrics is the hard part – using them is relatively easy. Once you've created a rubric, give copies to students and ask them to assess their own progress on a task or project. Their assessments should not count toward a grade. The point is for the rubric to help students learn more and produce better final products, so including self-assessments in grades is unnecessary and can compromise students' honesty.

Always give students time to revise their work after assessing themselves, then have them assess one another's work. Peer-assessment takes some getting used to. Emphasize the fact that peer-assessment, like self-assessment, is intended to help everyone do better work. You may also need to hold students accountable for their assessments of a classmate's work by having them sign off on the rubric they use. You can then see how fair and accurate their feedback is, and you can ask for evidence that supports their opinions when their assessments don't match yours. Again, giving time for revision after peer-assessment is crucial.

Parents can use rubrics to help their children with their homework. Finally, when you assess student work, use the same rubric that was used for self- and peer-assessment. When you hand the marked rubric back with the students' work, they'll know what they did well and what they need to work on in the future.

Grading (if you must) is also relatively easy with rubrics. A piece of work that reflects the highest level of quality for each criterion obviously deserves an *A*, one that consistently falls in the lowest level is a *D* or *F*, and so on. Because one piece of work rarely falls in only one level of quality, many teachers average out the levels of quality, either formally or informally.

Rubrics can also be included in portfolios. However you use them, the idea is to support and to evaluate student learning. Students, as well as teachers, should respond to the use of rubrics by thinking, "Yes, this is what I need!"

<http://learnweb.harvard.edu/alps/thinking/docs/rubricar.htm>

References

Brewer, R. (1996). *Exemplars: A Teacher's Solution*. Underhill, VT: Exemplars.

Marcus, J. (1995). "Data on the Impact of Alternative Assessment on Students." Unpublished manuscript. The Education Cooperative, Wellesley, MA.

Marzano, R., D. Pickering, and J. McTighe (1993). *Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model*. Alexandria, VA: ASCD.

Perkins, D., H. Goodrich, S. Tishman, and J. Mirman Owen (1994). *Thinking Connections: Learning to Think and Thinking to Learn*. Reading, MA: Addison-Wesley.