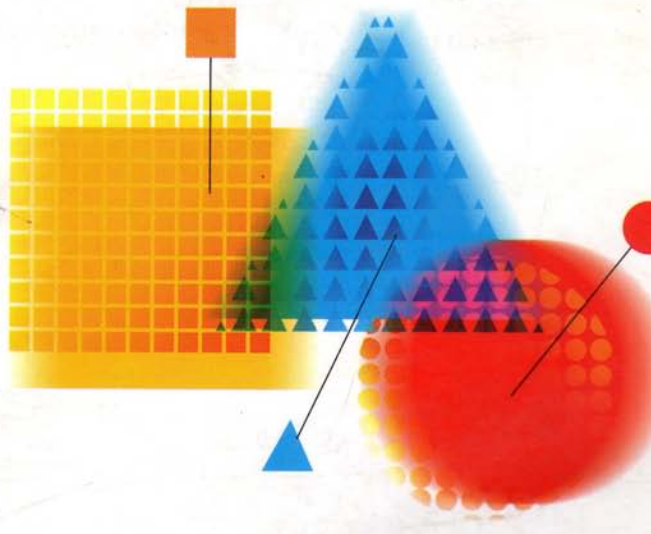


*Alicia R. W. de Camilloni, Susana Celman, Edith Litwin
y María del Carmen Palou de Maté*

La evaluación de los aprendizajes en el debate didáctico contemporáneo

 **Paidós Educador**



SISTEMAS DE CALIFICACIÓN Y REGÍMENES DE PROMOCIÓN*

Alicia R. W. de Camilloni

LA CALIFICACIÓN Y LA PROMOCIÓN DE LOS ESTUDIANTES

Los sistemas de calificación empleados en la evaluación de los aprendizajes y los regímenes de promoción de los alumnos constituyen un componente esencial de los sistemas de evaluación que se emplean en la educación formal. Aunque ambos son producto de un conjunto de decisiones prácticas adoptadas en virtud de la adhesión a una u otra entre las diversas corrientes teóricas de la didáctica, siempre ejercen, de hecho, un fuerte efecto normativo sobre las acciones de enseñanza y de aprendizaje ya que, más que aparecer como una consecuencia de los fundamentos teóricos que se suscriben en la materia, tienen un efecto tan poderoso que suelen operar de manera muy efectiva, en reversa, como determinantes de todas las prácticas pedagógicas. De este modo, si los sistemas de calificación y promoción permanecen estáticos en sus modalidades fundamentales y no acompañan las transformaciones de las prácticas de la enseñanza, sólo sirven de obstáculo a esas transformaciones, obligando a los docentes a una permanente adaptación a los moldes pedagógicos que los procesos de renovación o de reforma pretenden reemplazar. Si, por el contrario, se resuelve producir modificaciones en los sistemas de calificación y promoción, éstos

* Este artículo ha sido posible gracias al subsidio otorgado a la investigación dirigida por la autora por la Secretaría de Ciencia y Técnica de la UBA (FI215).

tienden a desarrollar una indudable influencia sobre las estrategias de enseñanza de los docentes y, particularmente, sobre las estrategias de aprendizaje de los alumnos, con un efecto denominado de "validez retroactiva". Sobre estos procesos mucho se ha estudiado en el campo del análisis del denominado "currículo oculto". Aunque esta influencia puede ser abusiva, porque es necesario trabajar de manera integral sobre la enseñanza y la evaluación ya que ésta es subsidiaria de la primera, debemos reconocer que el papel de los sistemas de calificación y de los regímenes de promoción es central en la educación escolar, más allá de la pretendida desvalorización que de ellos se suele hacer en algunas teorías de la enseñanza, en tanto se les atribuye el constituirse en una manifestación no siempre adecuada o necesaria del ejercicio del control o de la autoridad interna o externa a la institución escolar sobre los aprendizajes de los alumnos y, en consecuencia, en una imposición que recorta la libertad personal de estos últimos.

La importancia que tienen, pues, los sistemas de calificación y los regímenes de promoción, sea evaluándolos en un sentido pretendidamente positivo o sea éste negativo, obliga a efectuar un estudio muy cuidadoso de las características que ofrecen las diversas modalidades que presenta cada uno de los modelos, de manera de seleccionar aquellos que resulten más apropiados para los propósitos de la programación didáctica, de la planificación de la institución y del sistema escolar.

SISTEMAS DE CALIFICACIÓN

Es preciso considerar, en primer lugar, dos cuestiones básicas. El primer aspecto relativo a la constitución de un sistema de calificación es el relacionado con la construcción y la fijación de una escala de calificación, y el segundo aspecto está referido a la decisión fundamental relativa a la adopción de sólo una o más de una de estas escalas en el sistema de calificación que se diseña. Veamos la primera cuestión.

En el proceso de construcción de una escala es necesario tener en cuenta las características que las escalas de calificación adquieren en términos de las condiciones y posibilidades que brinda cada una de las grandes modalidades de las escalas de medición en general.

Escalas de medición

Las escalas de medición pueden ser clasificadas en cuatro grupos a partir de las características que presentan respecto de la relación de ordenamiento progresivo u horizontal de las categorías que las constituyen, de la distancia igual o desigual que existe entre los diferentes grados de la escala y de la existencia o no de un punto de partida (cero absoluto) que permita operar de manera particular con tramos parciales de la escala.

De este modo, las escalas de medición se agrupan en los siguientes tipos:

a) *Escalas nominales*. En ellas las categorías que componen la escala no se encuentran en un orden de seriación que implique que un grado es mayor o mejor que el o los otros grados. No hay grados superiores e inferiores. Todos se encuentran en el mismo nivel. Se emplean para medir la frecuencia con la que los ejemplares evaluados se incluyen en cada una de las clases que la escala presenta. La construcción de una escala nominal requiere que se cumplan las siguientes condiciones: la escala debe incluir todas las categorías necesarias para diferenciar los ejemplares que se evalúan, dependiendo su carácter y número de la necesidad de obtener la información que se requiere con un fin determinado, aquel que dio lugar a la construcción de la escala. Para ello, las categorías se determinan en función del número indispensable de clases que permitan dar cuenta del universo que se estudia. El número debe ser el mínimo posible, sin que se pierda información útil. La escala, además, debe ser completa, es decir que debe hacer posible la inclusión en alguna categoría de la escala de todos los ejemplares analizados. Por otra parte, todas las categorías deben ser excluyentes entre sí, lo que implica que ningún ejemplar podría ser incluido simultáneamente en más de una categoría. La definición de las categorías, para ello, debe ser tan clara que no genere ninguna interpretación ambigua que pueda prestarse a error de clasificación. Una escala nominal se construye *ad hoc*, en los términos necesarios para el estudio que se desea realizar. Si se trata, por ejemplo, de clasificar a las personas según su nacionalidad, el número de categorías para incluir en la escala dependerá de la utilización que se quiera dar a las informa-

ciones que se recojan. En un caso puede ser suficiente contar con dos grados (Argentino; Extranjero) o con tres (Argentino nativo; Argentino naturalizado; Extranjero) o con cuatro (Argentino nativo; Argentino naturalizado; Natural de país limítrofe; Extranjero natural de país no limítrofe), etcétera.

b) *Escalas ordinales*. Son aquellas que presentan una serie de categorías ordenadas de acuerdo con un criterio definido por un sentido u orientación determinada que no permite introducir alteraciones en la secuencia. Los grados de la escala se ordenan, por ejemplo, de mayor a menor o de mejor a peor. Deben cumplir con los mismos requisitos que las escalas nominales, a los que se suma, como se ha dicho, la definición del criterio de ordenamiento de la serie. Se caracterizan, además, porque las distancias entre los pares de categorías o grados consecutivos no deben ser necesariamente iguales. Así, el intervalo entre "Muy bueno" y "Bueno" no tiene que ser igual al intervalo entre "Bueno" y "Regular". Las escalas ordinales son específicas y se usan para medir aspectos determinados. Las escalas ordinales no son y, en consecuencia, no deberían ser utilizadas como escalas universales. Sin embargo con cierta frecuencia adquieren un alto grado de formalismo y se usan indistintamente para evaluar cuestiones muy heterogéneas, perdiendo de este modo la riqueza posible de su valor descriptivo que deriva ya sea de su construcción *ad hoc*, ya de la especificidad de la definición de cada uno de los grados de la escala.

c) *Escalas de intervalos*. Son aquellas en las que se debe responder a los mismos requisitos que hemos enumerado para las escalas ordinales, con la particularidad que se agrega de que los intervalos entre pares de grados consecutivos deben ser iguales. Este último rasgo permite operar con las mediciones en las que se han utilizado las unidades medidas por estas escalas efectuando operaciones aritméticas que no pueden realizarse con las escalas ordinales. Entre ellas, por ejemplo, es posible obtener la media aritmética (promedio) de las mediciones, lo cual no tendría sentido con unidades de medición con escalas ordinales. Las escalas numéricas, siempre que los grados estén definidos sin ambigüedad de tal modo que aseguren la

igualdad de los intervalos, son un ejemplo de estas escalas intervalares.

d) *Escalas de razones o proporciones*. Son aquellas en las que no sólo se cumplen las condiciones establecidas para la construcción de una escala de intervalos sino que se agrega a éstas la de poseer un cero absoluto. Esta característica permite efectuar operaciones en las que se suman mediciones obtenidas independientemente o se comparan mediciones en las que, por ejemplo, el grado "2" de la escala se puede considerar equivalente a la mitad del grado "4". Si no existiera un cero absoluto, esta interpretación de la relación entre los grados 2 y 4 no tendría sentido.

Tipos de escalas de calificación

Para la construcción o selección de una escala y para las decisiones relativas a la adopción de una escala de calificación es indispensable tener en cuenta las características que hemos mencionado respecto de las escalas de medición en general, ya que ellas son la matriz básica a partir de la cual se pueden conceptualizar las escalas de calificación de los aprendizajes escolares.

La norma general que se ha de aplicar en la construcción de una escala de calificación es que el número de grados de la escala debe ser elegido de acuerdo con la posibilidad real que tiene el evaluador de operar con una técnica apropiada y disponible para él, discriminando los diferentes grados de calidad de aquello (rasgo, conducta, realización) que ha de ser evaluado con dicha escala.

Las escalas de calificación pueden presentarse en diversas formas. Las clasificaremos en escalas numéricas, escalas conceptuales o descriptivas y escalas gráficas.

La escala numérica consiste en cierto número de categorías establecidas en términos descriptivos a las que se atribuye de antemano valores numéricos convencionales.

El número de grados comprendidos por la escala puede variar. Con fines escolares se han utilizado escalas de muchos grados (100; 75; 60; 50; 25 o 20 grados) y escalas de pocos grados (11; 10; 6; 5; 4; 3 o 2 grados). En la escuela secundaria y en la educación superior argentina se ha utilizado tradicionalmente una escala de 11 grados

(0 a 10). De ella encontramos dos versiones. En una, la calificación mínima de aprobación es 4 puntos (con 4 grados o niveles de calidad diferentes para la desaprobación y 7 niveles de calidad para la aprobación). En la otra, la calificación mínima de aprobación es 6 o 7 puntos, con un número superior de grados para la desaprobación (de 0 a 5 o 6, respectivamente) que para la aprobación (6 a 10 o 7 a 10).

Se observa actualmente, sin embargo, una tendencia en la teoría de la evaluación a apoyar la propuesta de reducción del número de grados de las escalas de calificación a una cantidad que no supere los cinco grados. Este principio de reducción es fruto de las críticas provenientes de las corrientes anticuantitativistas en educación, las que no pueden aceptar aquello que, con toda evidencia, consideran un exceso en la diferenciación cuantitativa aplicada al análisis de las realizaciones de los alumnos. En la práctica se comprueba que, a pesar de la apariencia en contrario, las escalas integradas por un gran número de grados no permiten una discriminación rigurosa y clara de cada una de las categorías de calidad correspondientes. Por esta razón, lejos de perfeccionar la exactitud de la medición, contribuyen a acrecentar la imprecisión de las calificaciones asignadas por los docentes. También desde un punto de vista estrictamente práctico, trabajar con una escala de pocos grados facilita la tarea del docente ya que disminuye la posibilidad de error o injusticia y, por ende, la ansiedad en la elección de la calificación. Por otro lado, también ayuda a lograr consensos en los criterios de calificación empleados por diferentes docentes. De este modo, parece conveniente construir escalas con un número de categorías pequeño con el fin de mejorar la definición de cada grado y con el objetivo de contribuir al perfeccionamiento de la confiabilidad de las calificaciones.

Pero una escala de dos grados ("1" y "2" o "suficiente" e "insuficiente", por ejemplo) puede resultar poco descriptiva de los progresos de los estudiantes, porque cada grado debe cubrir un campo muy amplio y variado de rasgos. Si bien se aumenta la confiabilidad en las escalas de pocos grados, Robert E. Thorndike y Elizabeth P. Hagen (1991) señalan:

Desde luego, esta ventaja se logra a expensas de la mayor parte de la información que el sistema de calificación puede suministrar acerca del individuo. En un sistema de aprobado-reprobado, sólo sabemos que el

estudiante ha sido juzgado como aprobado en un curso, excepto por la pequeña fracción que se juzgó como fracasada. [...] A medida que se incrementa el número de discriminaciones más allá del sistema de dos o tres categorías, las distinciones empiezan a ser importantes para todos los estudiantes (pág. 577).

Así pues, si la capacidad descriptiva de la escala es pobre se perderán muchas de las funciones de la evaluación. Aumentando en pocos grados más la escala, con uno o dos puntos como mínimo por el lado de la no aprobación, y con dos o tres puntos por el costado de la aprobación, la escala permite una mejor definición de las categorías empleadas por el evaluador. Según los casos, sin embargo, puede ser útil discriminar con mayor fineza el nivel de logro en las realizaciones actuales de los alumnos. Las escalas de cuatro o cinco grados suelen servir de base para la elección de escalas. En ellas se emplean uno o dos grados para diferenciar la calidad en la desaprobación y dos, tres o cuatro para distinguir niveles de calidad en la aprobación. Por ejemplo:

1) 1 (desaprobación); 2, 3 y 4 (aprobación).

2) 1 y 2 (desaprobación); 3, 4 y 5 (aprobación).

Se puede afirmar que más allá del número que el análisis de los casos particulares demuestre ser necesario y suficiente, al aumentar el número de grados de la escala es más lo que se pierde en confiabilidad que lo que se gana en información.

Las escalas numéricas, según la calidad de la definición de cada uno de los grados que las componen, aplicando la categorización que antes se presentó con respecto a los tipos de escalas de medición, pueden ser ordinales, intervalares o de razones.

Las escalas conceptuales o descriptivas tienen algunos aspectos semejantes a las escalas numéricas pero presentan otros rasgos que los diferencian claramente de ellas. Ambos tipos presentan un escalonamiento de la calidad de lo que es evaluado desde un grado mínimo hasta un grado máximo. Se distinguen, empero, porque la escala numérica, al ser formalmente siempre la misma, sin importar el carácter sustantivo de lo que se evalúa, es mucho menos descrip-

tiva que la escala conceptual, se presta a distintas interpretaciones e induce a que rasgos heterogéneos sean tratados como equivalentes o equiparables. Una vez asignada una calificación, la frialdad de los números no permite discernir el tipo de rasgo, producto o comportamiento evaluado.

Las escalas conceptuales, por el contrario, se ajustan a las necesidades de cada instancia evaluativa. Varían su forma según las características de aquello que se desea evaluar.

Veamos algunos ejemplos:

Para apreciar la frecuencia con la que se produce un cierto fenómeno, la escala puede ser la siguiente:

- Siempre
- Generalmente
- A veces
- Nunca

Para apreciar la intensidad:

- Muy fuerte
- Fuerte
- Moderadamente fuerte
- Regular
- Débil
- Muy débil

Para apreciar la cantidad y la calidad de los conocimientos:

- | | |
|-----------------|----------------|
| • Sobresaliente | • Excelente |
| • Distinguido | • Muy Bueno |
| • Bueno | • Bueno |
| • Aprobado | • Regular |
| • Insuficiente | • Insuficiente |
| • Reprobado | • Reprobado |

Cada una de estas categorías constituye un grado de la escala y debe ser claramente definida. Con frecuencia se hace corresponder

a cada grado conceptual un símbolo numérico. Si esto ocurre, la escala conceptual se convierte en una escala numérica y presenta los mismos problemas de que adolece este último tipo de escalas. La delicadeza de las evaluaciones se pierde en favor de una formalidad presuntamente favorable a la exactitud, cómoda, por cierto, pero generadora del vaciamiento de la dimensión cualitativa de la evaluación.

En el mismo caso se encuentran las escalas alfabéticas como la que sigue:

A; B; C; D; E o F..., etcétera.

En ellas también suele establecerse una correspondencia con una escala numérica, pero esto no las convierte en escalas intervalares o de razones porque las escalas conceptuales son, por su misma naturaleza y básicamente, escalas ordinales.

La escala gráfica consiste en representar, sobre referencias lineales, categorías descriptivas determinadas.

Tiene la característica de presentar la intensidad de un rasgo como un continuo desde un grado mínimo hasta un grado máximo. La intensidad deseable del rasgo puede encontrarse señalada y definida en cualquiera de los puntos de la línea de evaluación, en ambos extremos y en un número determinado de puntos intermedios. Cada uno de los grados de la escala, cuyo número se recomienda que no supere los siete para no disminuir la confiabilidad de la evaluación, debe estar especificado en términos de lo que el docente está en condiciones de observar, conceptualizar e inferir realmente. En ninguno de los dos extremos se colocarán categorías que no puedan ser alcanzadas efectivamente o que se estima casi imposible que se presenten en los alumnos que integran la clase. Pero, en compensación, se habrá de tener en cuenta también que el apoyo gráfico lineal de la escala corresponde a una recta y no a un segmento. Será posible, por lo tanto, utilizar todos los puntos de la recta, aun los intermedios, entre los puntos definidos y los que se encuentran más allá de los últimos puntos definidos en los extremos.

Veamos algunos ejemplos de escalas gráficas.

INTERVENCIÓN EN CLASE

Interviene siempre, preguntando y respondiendo. Aporta nueva información.	Responde voluntariamente cuando es interrogado. A veces hace preguntas.	Responde voluntariamente cuando es interrogado.	Responde sólo cuando es interrogado.	No responde ni interviene nunca.

LECTURAS SOBRE LA MATERIA

Estudia a veces y superficialmente en el libro de texto.	Estudia siempre pero superficialmente en el libro de texto.	Estudia siempre y cuidadosamente en el libro de texto.	Estudia en el libro de texto y consulta algunos materiales recomendados.	Estudia en el libro de texto y consulta bibliografía recomendada y no recomendada por el docente.

Al igual que las escalas ordinales, las escalas gráficas se construyen *ad hoc*, no son escalas universales o formales sino escalas específicas.

Elección de una escala de calificación

Para determinar cuál ha de ser la escala que se ha de emplear en un programa de evaluación se deben tomar en consideración tres aspectos: el tipo de escala, el número de grados de la escala y la definición que se dará a cada uno de los grados.

En cuanto al número de grados, la escala debe estar compuesta por el número mínimo de categorías que resulte suficiente para que la evaluación no pierda sensibilidad ante las diferencias individuales. Un número razonable es, por lo general, de 4 o 5 grados, con uno

o dos grados para la desaprobación y tres para la aprobación. Si bien se han utilizado escalas de dos grados (aprobado-desaprobado), éstas parecen ser más adecuadas para los niveles más altos de la educación, por ejemplo, la enseñanza universitaria, pareciendo, sin embargo, insuficientes para diferenciar niveles de calidad en los trabajos realizados en los niveles primario o básico y secundario, en razón de que las calificaciones cumplen una doble función respecto del aprendizaje. En primer lugar, las escalas con diferencias entre diversos grados de calidad en la aprobación proveen información más precisa acerca de los resultados obtenidos y, en segundo lugar, influyen fuertemente, sin duda, sobre la motivación de los estudiantes frente a los requerimientos de estudio. El reconocimiento al esfuerzo realizado por el alumno aparece, en este caso, como un componente de la calificación y cuanto mayores son los niveles del sentimiento de éxito, mayor el nivel de aspiración del alumno. Ambos efectos, más y mejor información recibida por el alumno respecto de su aprendizaje y mayor motivación, actúan entre los determinantes más importantes del rendimiento.

Pero cuando las escalas tienen, en cambio, un número excesivo de grados, tanto por el lado de la desaprobación como por el de la aprobación, las escalas mismas pueden obligar a un exceso de cuantificación y convertirse en posible causa de error en la evaluación. El número de grados de la escala debe elegirse muy cuidadosamente. Requiere justificación explícita y la adopción de medidas complementarias que permitan un uso confiable de la escala elegida.

En cuanto a la definición de cada grado, debe ser clara y comprensible para todos los implicados en los procesos de evaluación y en el uso de la información que ella provee. Todos —docentes, alumnos, directivos, padres y público en general— deben poder interpretar las calificaciones de la misma manera. Cuando se emplea una escala numérica debe explicarse previamente qué significado se atribuye a cada uno de los símbolos numéricos que integran la escala. De otro modo, la evaluación estaría impedida de cumplir con sus funciones orientativas.

Se sugiere también la necesidad de completar la calificación numérica con descripciones que sirvan de guía a alumnos y docentes para el diagnóstico de logros y dificultades y para la orientación

del tratamiento apropiado para solucionar estas últimas. Una calificación numérica que no especificara en qué consisten los logros y los yerros no tendría utilidad para el alumno más que como pauta simbólica de su éxito o fracaso. No es de extrañar que con este tipo de calificaciones, la evaluación suela quedar al margen de los procesos de enseñanza y de aprendizaje y pierda, en consecuencia, sus funciones educativas. La calificación debe enriquecerse mediante la localización de los errores y el señalamiento aprobatorio de los resultados apropiados en el trabajo evaluado.

Otro difícil problema queda por resolver. Toda calificación lleva implícito un conflicto básico en las funciones de la evaluación. Hay dos maneras de calificar que responden a diferentes criterios, según se contrasten las realizaciones del estudiante con sus propias posibilidades o con un patrón ajeno a él, sea el rendimiento de sus compañeros (una norma) o una pauta establecida por el docente como objetivo a alcanzar (un estándar). Es evidente que cada una de estas formas dará lugar a calificaciones diversas para un mismo trabajo y un mismo alumno. También el éxito y el fracaso pueden tener dos tipos de interpretación. Tienen una realidad subjetiva cuando se refieren al acuerdo o desacuerdo existente entre el nivel de desempeño del alumno con su nivel de aspiración (lo que para un alumno es un éxito para otro puede ser un fracaso) y pueden tener una realidad objetiva cuando se compara ese nivel de desempeño con el patrón de rendimiento esperado y que ha sido adoptado a priori por el docente en su programación didáctica. Este conflicto de funciones se refleja, con frecuencia, en una pregunta común entre los docentes: ¿debemos calificar al alumno por su esfuerzo o debemos hacerlo en función de sus realizaciones efectivas? Ante esta pregunta se pueden ofrecer diversas respuestas porque, en efecto, el rendimiento del alumno puede ser apreciado, ya no de dos, sino de tres maneras diferentes, según se lo compare con los logros realmente alcanzados por otros alumnos, con una pauta fijada a priori por el docente o con el rendimiento que podría esperarse de él en condiciones conocidas.

Sin incursionar todavía en la problemática de los regímenes de promoción y restringiéndonos a la cuestión de las escalas de calificación, podemos establecer, siguiendo el primer criterio mencionado arriba, una relación entre el rendimiento de un alumno en una

asignatura determinada y el rendimiento medio de un grupo representativo de alumnos de edad similar en el mismo nivel, ciclo y año de enseñanza de la asignatura. Con este fin se han acuñado las nociones de edad educativa y de cociente educativo. De la misma forma en que se deriva el concepto de "edad mental" de un test de inteligencia, se deriva la edad educacional de un test de rendimiento. Utilizando un instrumento diseñado para evaluar conocimientos en un dominio determinado o para evaluar el desarrollo de una habilidad, se mide el rendimiento de una población de alumnos de características definidas. Cuando se trabaja con sujetos de distintas edades y niveles de aprendizaje en una disciplina, se obtiene, de este modo, una escala de normas de edad referida al instrumento utilizado. La edad educativa representa la edad en años y meses de las personas que obtuvieron un determinado puntaje en la administración del instrumento empleado, medido en términos de su relación con el puntaje promedio (en este caso, la mediana es la medida de tendencia central elegida) de la norma de edad de la población de referencia. La mediana es el puntaje por debajo del cual se encuentra la mitad de los puntajes, de todo ese grupo de edad. Con las medianas obtenidas se construye la escala de rendimiento por edad.

Tomemos, por ejemplo, la escala siguiente:

Puntaje	Edad	Año de escolaridad
20	14 años	8.8
21	14 años, 2 meses	8.9
22	14 años, 3 meses	8.9
23	14 años, 4 meses	9.0
24	14 años, 5 meses	9.1

Si un alumno obtuviera 22 puntos, su edad educativa sería de 14 años, 3 meses, la que correspondería a 8.9 años de escolaridad.

Si relacionamos la E.E. (edad educativa) con la edad cronológica (E.C.) del alumno obtendremos su cociente educativo (C.E.).

$$C.E. = \frac{E.E.}{E.C.} (.100)$$

Por ejemplo:
$$C.E. = \frac{14 \text{ años, 8 meses}}{13 \text{ años, 8 meses}}$$

$$C.E. = \frac{176 \text{ meses}}{164 \text{ meses}} (.100)$$

$$C.E. = 107$$

El C.E. nos indicará cuánto tiempo ha necesitado el alumno para alcanzar un rendimiento determinado en comparación con el promedio (mediana) del tiempo requerido por los alumnos de su edad. Así, de la misma forma en que el C.I. se deriva de la noción de edad mental, el C.E. se deriva del concepto de E.E. En ambos casos se trata, y debe siempre puntualizarse, de un C.E. "medido", al igual que de una E.M. y un C.I. "medidos", sólo válidos con referencia a los instrumentos de medición utilizados.

Veamos algunos ejemplos. Un estudiante que tuviera 14 años de E.E. y una E.C. de 15 años, tendría un C.E. de 93.3. Otro estudiante de la misma clase que obtuviera el mismo puntaje en la evaluación pero que tuviera 13 años de edad cronológica, tendría un C.E. de 107. Se ha de tener en cuenta, empero, que para la determinación de la escala de normas de edad se toman como base las medianas de los puntajes respecto de cada grupo de edad y que, por lo tanto, en cada uno hay puntajes superiores y puntajes inferiores a las medianas. Al determinar los C.E. de los alumnos de una clase se ha de esperar, pues, que haya tantos estudiantes que estén por encima del valor 100 como por debajo de él. Hay que señalar, también, que el "C.E. medido" no se correlaciona directa y necesariamente con el "C.I. medido" pues son muchos los determinantes que actúan sobre el C.E. además del C.I.

Si tomamos, en cambio, el segundo criterio mencionado más arriba, que consiste en comparar al estudiante con una pauta establecida a priori por el docente, nos encontramos con la modalidad de la evaluación "referida a criterios". Aquí el verdadero problema a resolver es la determinación de los niveles de logro deseados por el docente, ya que ellos deben encontrarse entre los niveles posibles de alcanzar para los alumnos. El mecanismo que se utiliza para determinarlos es producto, por lo general y entre otros condicionantes, de la experiencia anterior del docente frente a sus alumnos, de la experiencia anterior del docente como alumno, de los niveles de aspiración del docente, de la cultura organizacional de la institución escolar en la que se desarrolla la enseñanza y, por último, de las tradiciones académicas de la disciplina que se enseña. Si las calificaciones del alumno son relativas a estos patrones, en ocasiones rígidos, y en otras oscilantes, la validez de dichas calificaciones puede ser cuestionable. Será siempre relativa a los criterios de base empleados y puede caer en un sensible irrealismo, tanto por lo bajo de la pauta establecida como por la imposibilidad de alcanzarla en razón de su extremada dificultad.

Si se sigue, por fin, el tercer criterio, consistente en comparar la realización del estudiante con lo que según sus condiciones podía esperarse de él, será necesario utilizar otra fórmula, que combine una evaluación de sus aptitudes con la evaluación de sus logros efectivos. Tradicionalmente para este fin se ha empleado el cociente de rendimiento (C.R.), obtenido dividiendo el C.E. por el C.I. Para un alumno que tiene, por ejemplo, un C.E. de 14 años de edad con una E.M. de 17 años, 6 meses, y con un rendimiento "normal" para su edad, el C.R. sería:

$$C.R. = \frac{C.E.}{C.I.} = \frac{\frac{E.E.}{E.C.}}{\frac{E.M.}{E.C.}} = \frac{\frac{14}{14}}{\frac{17,5}{125}} = \frac{100}{125} = 80$$

En este ejemplo, un alumno bien dotado no ha rendido todo lo que podía esperarse de él. La medida de esta diferencia nos la da el C.R. que es significativamente inferior a 100.

En otro ejemplo, un alumno de 15 años de E.C. y 13 años de E.M., retrasado en su E.E. (E.E. = 13 años) con respecto a la mayoría de los jóvenes de su misma E.C., tiene un C.R. de 100:

$$\text{C.R.} = \frac{\text{C.E.}}{\text{C.I.}} = \frac{\text{E.C.}}{\text{E.M.}} = \frac{15 \text{ años}}{13 \text{ años}} = \frac{86,6}{86,6} = 100$$

Este segundo alumno, cuyo cociente de rendimiento es 100, a pesar de sus dificultades, trabaja mejor que el primero en relación con lo que podía lograr cada uno de ellos. El primer alumno requiere mayor atención con el fin de ayudarlo a alcanzar el pleno desenvolvimiento de sus posibilidades. Pero debemos tener siempre en cuenta, también, que se debe actuar sobre todas las potencialidades de los dos para no acentuar sus aparentes diferencias iniciales y lograr el máximo desarrollo, de hecho imprevisible, en ambos alumnos.

Los conceptos de C.E. y de C.R. se encuentran con mucha frecuencia en la literatura de habla inglesa sobre el tema y son, o eran, de aplicación muy frecuente en los países anglosajones. Su empleo ha sido escaso o nulo en América latina, en donde es también poco frecuente el empleo de pruebas estandarizadas de rendimiento. En los conceptos que hemos presentado pueden advertirse, por una parte, la influencia de las corrientes conductistas de la medición en la evaluación, que se manifiestan en el esfuerzo por generar fórmulas que permitan construir indicadores "exactos" y confiables para expresar los resultados de la medición. Pero, por otra parte, es interesante observar que, más allá del intento de cuantificar los rendimientos y las habilidades o aptitudes, estos conceptos permiten estudiar la relación existente entre aspectos que

en la evaluación de los aprendizajes con enfoques no conductistas también deben ser tomados en consideración. Son, pues, útiles, aun cuando no se pretenda utilizar las fórmulas para producir calificaciones numéricas y sólo se usen para contrastar criterios diferentes de evaluación. Su valor conceptual es significativo incluso en paradigmas no cuantitativistas de la evaluación en educación.

En estos términos, como hemos visto, la construcción de un sistema de calificación implica la elección entre alguno de los siguientes criterios: la calificación por las dotes que el alumno manifiesta poseer para la realización de tareas en dominios determinados, la calificación por sus logros efectivos en comparación con una pauta predefinida por el docente, o la calificación otorgada por comparación con los logros de los compañeros de su grupo de clase o de una población seleccionada. Cada uno marca una dirección distinta para la evaluación.

Al calificar al alumno con referencia a sus dotes y a su esfuerzo, se utiliza una noción correspondiente al campo del C.R. Un alumno voluntarioso, que logra superar sus propias dificultades, se vería mejor calificado que, por ejemplo, otro estudiante inteligente pero indolente. Las calificaciones de ambos estudiantes ya no serían semejantes en el caso de realizaciones iguales, sino que ocurriría —como con el C.R., si hubiera medio de calcularlo cuantitativamente— que el alumno esforzado obtendría mejor calificación.

Para calificar al alumno por su rendimiento efectivo en relación con una pauta predefinida, es menester expresar esta última en términos mensurables o apreciables para el docente. Se trata, entonces, de establecer previamente una descripción ajustada del rendimiento deseado. A este fin respondían los llamados "objetivos operacionales" que tenían como función principal facilitar una medición numérica ajustada. Esos objetivos pueden ser reemplazados, sin embargo, desde un marco teórico diferente, por la descripción cualitativa de los logros esperados. No se estaría lejos, en este caso, de la aplicación de la noción de C.E. porque, aun si no se desea trabajar con una cuantificación exacta, es obvio que el criterio que se aplica en una evaluación cualitativa puede coincidir con el de este concepto.

Cuando se adopta el criterio de calificar al alumno a partir de la comparación de sus aprendizajes con los logros efectivamente

demostrados por su grupo de clase u otro criterio tomado como referencia, los problemas que surgen son muy específicos y serán analizados en el examen que haremos posteriormente de los regímenes de promoción.

Aunque muchas veces ajenos al espíritu rigurosamente cuantitativista que sustenta el uso de calificaciones asignadas mediante sistemas asociados al uso de escalas de calificación numéricas de intervalos o razones, los docentes siempre comparan y eligen criterios. Es menester que se extremen los cuidados para que las calificaciones tengan origen y fundamentación clara, en particular, para los alumnos.

Considerando que, entre los criterios que hemos expuesto más arriba, el equivalente al C.E. y el semejante al C.R. evalúan en función de aspectos diferentes, ambos interesantes por su significatividad, se puede sostener que sería conveniente utilizarlos en conjunto, con el propósito de mejorar las apreciaciones parciales y de no perder información útil para el docente y el alumno. El sistema que se crea de este modo debiera conservar la identidad de cada uno y preservarlos de la confusión en la interpretación. Una sola calificación numérica, apoyada en uno solo de esos criterios o como fruto de una combinación de ambos, sería siempre insuficiente para el cumplimiento de las funciones de la evaluación. Se recomienda adoptar, entonces, un sistema de doble calificación. En él se complementan dos calificaciones: una, referida a los progresos realizados por el escolar en relación con una norma establecida previamente, y la otra, referida a sus progresos contrastados con sus aptitudes consideradas individualmente. Este sistema permite combinar las ventajas de ambos criterios y permite manejar simultáneamente los dos sentidos, objetivo y subjetivo. Para realizaciones semejantes, por ejemplo, las dobles calificaciones podrían resultar las siguientes:

Alumno A: 4 (muy bueno)

Alumna B: 4 (regular)

Alumno C: 4 (excelente)

Aquí se han empleado dos escalas: para el rendimiento efectivo en términos de una pauta predefinida: escala numérica de 1 a 5. Para

el rendimiento en términos de capacidad y esfuerzo, escala conceptual con grados con definición clara: malo, regular, bueno, muy bueno, excelente.

La primera calificación procura ser objetiva. La segunda, más subjetiva, se puede emplear en la construcción de una calificación nueva atendiendo al sentido de la progresión que se manifiesta en estas segundas calificaciones. Servirá en cada caso de advertencia o de estímulo y reconocimiento del docente ante la actuación del alumno, incidiendo sobre posteriores aprendizajes y, por tanto, como proveedora de información y estímulo de la motivación. Las segundas calificaciones pueden convertirse en calificación numérica. Pero ésta no resulta del cálculo del promedio sino de la tendencia que ha manifestado el alumno respecto de su esfuerzo en el trabajo de aprendizaje.

La necesidad de completar las evaluaciones numéricas con anotaciones referentes a los errores y sus posibles causas, a la adecuación de la velocidad impresa a la realización del trabajo, que puede ser excesivamente lenta o apresurada, y la referencia a la dirección e intensidad de los esfuerzos del alumno, nos muestra la insuficiencia de un sistema de calificaciones que pretenda manejarse sólo con una simple escala numérica.

La enseñanza opera como una fuerza en un sistema de fuerzas. Es por sí misma, ella también, un conjunto dinámico que actúa sobre y con sujetos dinámicos. Las magnitudes escalares, desprovistas de orientación y en las que sólo interesan los valores numéricos, no pueden describir el tipo de cambio producido por la educación. Por esto, pensamos, el aprendizaje escolar debiera ser conceptualizado a los fines de su medición como magnitud vectorial en el marco de la evaluación, representándose los progresos por segmentos o semirrectas con dirección, y no sólo por sus valores numéricos asignados en instancias puntuales aisladas entre sí.

Decisión sobre el número de escalas de calificación

Del análisis efectuado sobre los distintos tipos de escala de medición y de calificación surge con claridad que cada una de las modalidades califica objetos (caracteres y rasgos, procesos y productos) diversos y lo hace de distinta manera. De ahí que toda

calificación sea relativa no sólo al instrumento que se ha utilizado para medir, como lo afirmamos más arriba, sino también a la escala de calificación empleada.

En la mayor parte de los sistemas se ha intentado construir escalas de carácter universal, esto es, capaces de describir con precisión aceptable los diferentes niveles de calidad de objetos, procesos o habilidades y destrezas de todo tipo. Estas escalas tienen carácter formal y pueden ser ejemplificadas por las escalas numéricas que son de uso habitual en la educación escolar. Cuando las escalas alfabéticas o conceptuales están integradas por grados definidos de manera muy general, también se emplean al modo de escalas universales. Este tipo de escalas presentan la obvia ventaja de su practicidad. Todos los docentes pueden usar la misma escala y se hacen posibles, también, las comparaciones entre calificaciones correspondientes a distintas disciplinas y clases de trabajo. Se produce, de este modo, una homologación formal de la consistencia interna de los logros de un alumno y, también, de los de una cantidad ilimitada de alumnos. Pero se debe remarcar que la homologación es sólo formal y que, por esta causa, pierde la calificación buena parte de su valor descriptivo. Si se quieren registrar y conservar las características diferenciales específicas de las realizaciones o de los aprendizajes de los alumnos, tratándose de uno solo o de muchos, es imprescindible emplear escalas diferentes con menor pretensión de universalidad, ya que ésta se encuentra en correlación inversa con el poder descriptivo de las calificaciones. La decisión de emplear una misma escala, por ejemplo, para todas las disciplinas del currículo o de emplear escalas construidas *ad hoc* para cada una de ellas es, pues, una cuestión de la mayor importancia. De lo expresado hasta ahora al respecto, se infiere que parecería que el problema de las escalas de calificación en un programa escolar estará bien resuelto cuando esté conformado por el menor número de escalas diversas que resulten necesarias para proporcionar la información que sea útil para los alumnos, los docentes, las autoridades escolares y todos aquellos implicados en los procesos de enseñar y aprender en la escuela.

RÉGIMENES DE PROMOCIÓN

El establecimiento de un régimen de promoción es siempre una tarea delicada en la que los riesgos de adoptar decisiones equivocadas son grandes. El régimen elegido debiera hacer imposible que se impida la promoción de un alumno al curso siguiente si éste se encuentra en condiciones de proseguir sus estudios porque ha realizado ya los aprendizajes necesarios. E, igualmente, tampoco debiera permitir la promoción de un alumno que no posee los conocimientos básicos indispensables para realizar los aprendizajes correspondientes al curso superior o que no ha alcanzado las competencias que la acción escolar debe garantizar ante el propio individuo y la sociedad. La generosidad injustificada en materia de promociones es causa muy frecuente de posteriores fracasos escolares, así como el a veces injusto rechazo de promociones se produce, en ocasiones, en el sistema escolar y es causa, a su vez, de grave frustración personal, e incluso de abandonos y deserción escolar. Numerosos estudios comprueban la importancia que tienen las calificaciones obtenidas en la escuela sobre la autoestima de los alumnos. La influencia de las calificaciones es grande y tiene un doble efecto: por un lado y de manera general, sobre la construcción del "sí mismo" del estudiante, y por el otro, y más específicamente, sobre las aspiraciones que tiene frente al estudio. Si el primero es decisivo en el proceso de desarrollo de su personalidad, el segundo tiene un efecto primario en las motivaciones y los niveles de rendimiento deseados y efectivamente alcanzados por él.

La elección de un régimen de promoción presenta, entre sus principales dificultades, la de requerir que se asocien de manera previsible las calificaciones provenientes de diferentes instancias, lo cual implica encontrar una forma de conjunción de lo diverso. Los términos "agregación", "combinación" y "reconciliación" han sido adoptados para referirse a los procesos donde se reúnen y relacionan conjuntos de calificaciones con el propósito de producir una calificación única. Aunque afirma que no es posible producir definiciones perfectas de estos términos, Dylan Wiliam (1992) propone utilizar la palabra "combinación" para describir todos esos procesos, "agregación" cuando se reúnen evaluaciones

de diferentes "cosas" y "reconciliación" cuando las calificaciones se refieren a la misma "cosa" (pág. 19).

Para la fijación de un régimen de promoción son varios los criterios que se han de combinar o complementar, luego de analizar los rasgos fundamentales de las modalidades empleadas con mayor frecuencia en los sistemas escolares.

Analizaremos, en primer lugar, seis de ellos, que se destacan por su uso más frecuente en los sistemas escolares: el sistema de promoción por calificación promedio, el sistema de "logros mínimos exigidos", el sistema de promoción por examen final, el así llamado "sistema de promoción sin examen", los sistemas de puntajes derivados y la evaluación de portafolio.

Régimen de promoción por calificación promedio

Emplea una de las medidas estadísticas de tendencia central como base para la determinación de la promoción. La medida más utilizada con este fin es la media aritmética. Ésta se obtiene sumando primero todas las calificaciones obtenidas por el alumno y dividiendo después la suma por el número de calificaciones sumadas:

$$\text{Media aritmética} = \frac{\text{Suma de } x}{n}$$

en donde x son las calificaciones numéricas obtenidas por el alumno y n el número de calificaciones independientes sumadas.

Las calificaciones de bajo puntaje son equilibradas por calificaciones de puntaje más alto. En el régimen de promoción se determina el promedio mínimo que debe alcanzar el alumno para obtener esa promoción.

Este régimen de promoción tiene algunas características que deben tenerse en cuenta cuando se elige un sistema de promoción o cuando se aplica como resultado de la selección efectuada por la institución o el sistema escolar.

Para que las calificaciones promedio sean confiables debe cumplirse la condición de que la escala empleada sea una escala de intervalos o una escala de razones. Si los intervalos entre pares

de grados consecutivos no son iguales no tiene sentido la pretensión de realizar operaciones aritméticas que pretendan dar cuenta exacta de resultados de aprendizaje traducidos en números que expresan tanto logros como diferencias entre alumnos con una precisión no justificada. Si en una escala, por ejemplo, la distancia entre dos grados consecutivos (punto 5 y punto 6 de la escala) no es igual a la distancia que existe entre otros dos (punto 6 y punto 7), el promedio de las calificaciones 5 y 7 no puede ser 6 puntos ya que la distancia entre 5 y 6 no puede ser compensada por la distancia 6 a 7 puntos. Ahora bien, en la práctica, en las escalas con calificación promedio mínima exigida para la aprobación es común que la distancia entre esa calificación y la inmediata inferior sea mayor que la que existe entre el promedio mínimo exigido y la calificación inmediata superior a él. El docente suele exigir mayor rendimiento relativo para pasar de la calificación inmediata inferior al promedio mínimo que se exige para promover al alumno, por ejemplo, el que el mismo docente requiere para asignar la calificación inmediata superior a éste. Por lo tanto, la escala utilizada en la práctica, a pesar de su apariencia, no es una escala de intervalos o de razones, porque se constituye en ella una zona límite, de extensión variable y definición ambigua, en torno a la calificación promedio y en la que, con frecuencia, puede no haber diferencia efectiva entre los alumnos promovidos y los no promovidos.

Compárense las calificaciones siguientes que corresponden a diferentes alumnos en un régimen que exige una calificación promedio mínima de 6 puntos para la aprobación:

- A: 4; 6; 6; 7 calificación promedio: 5,75
- B: 7; 6; 7; 4 calificación promedio: 6
- C: 5; 7; 6; 6 calificación promedio: 6
- D: 7; 6; 5; 7 calificación promedio: 6,25

Las calificaciones promedio, en este caso, muestran cómo se construye esa zona límite. Para salvar el concepto de zona, en reemplazo del punto límite para la promoción (que haría que el alumno A no fuera promovido y sí los tres restantes), es común que se aproximen la calificaciones que se encuentran entre $n,50$ y $n,99$ al entero más cercano. Pero este esfuerzo de corrección de la escala no

es suficiente para evitar y justificar la inexactitud real de la calificación. También habría que demostrar que existe real diversidad entre el n,49 y el n,50. Por otra parte el sistema tampoco permite que se manifiesten con claridad en los promedios las diferencias de rendimiento entre los cuatro alumnos en las diferentes instancias de evaluación.

Como vemos, la calificación promedio no es suficientemente descriptiva del nivel de aprendizaje de los alumnos. Esto es producto de la conjunción de diversos factores. Entre ellos podemos mencionar: 1) la variación de las distancias entre grados consecutivos en las escalas de calificación que se utilizan habitualmente en la escuela, y 2) la combinación de este sistema de promoción con escalas que poseen un número relativamente grande de grados. Si se emplea, por ejemplo, una escala de 11 grados (0 a 10) con el sistema de promoción por promedio calculando hasta los centésimos, la escala resultante tendrá 1001 grados. Si la escala de base tuviera 5 grados, la escala para los promedios tendría 401 grados.

Como ya vimos, esta excesiva discriminación cuantitativa entre valores asignados a conjuntos de aprendizajes no es válidamente descriptiva de éstos ni de las diferencias existentes entre los aprendizajes de distintos alumnos. Sería difícil demostrar que la diferencia entre 574 y 575 puntos tiene alguna correspondencia significativa cuando se interpretan logros y fracasos en esos aprendizajes. Del mismo modo no puede considerarse relevante una diferencia de 25 puntos en una escala de 1001 grados.

El sistema de promoción por promedio presenta otras serias desventajas:

1. Un alumno que ha hecho un trabajo excelente y uno muy malo obtendrá la misma calificación promedio que otro que ha hecho dos trabajos pobres. Las calificaciones de uno y otro, por ser idénticas, aunque se refieren a dos situaciones sustancialmente distintas, no son útiles porque no pueden servir de base ni para una tarea de diagnóstico ni de pronóstico.

2. Por este sistema de promoción adquieren el mismo valor, respecto de la calificación promedio, trabajos de naturaleza heterogénea: orales, escritos, prácticos, teóricos, breves, extensos, menos importantes, muy importantes, etcétera.

3. Marca un tope de rendimiento para los alumnos. La calificación promedio mínima exigida para la promoción se convierte en el objetivo perseguido y en el símbolo del nivel de aspiración de una gran cantidad de alumnos, incluso de muchos estudiantes que podrían alcanzar mejores logros.

4. Crea en el docente una tendencia a cuantificar todos los comportamientos de los alumnos en clase, estereotipando las categorías de observación.

5. Se presta a las especulaciones de los alumnos con respecto a la calidad de los trabajos que presentan, funcionalizándolos en relación con su necesidad de lograr sólo el puntaje necesario para obtener el promedio requerido para la promoción.

6. El régimen de promoción por promedio facilita que la ignorancia en un tema o el mal desempeño en una tarea sea compensado por el conocimiento de otro tema o la realización de otra tarea diferente.

Algunas de estas características no deseadas son propias de la naturaleza misma del sistema, otras, en cambio, nacen de su mala utilización cuando no se adoptan los recaudos debidos. En los dos casos afectan seriamente el aprendizaje de los alumnos. No se puede negar, sin embargo, que su amplia aplicación se debe a su practicidad, porque facilita el manejo de gran cantidad de calificaciones y de decisiones de promoción o no promoción de los alumnos.

Se puede intentar reemplazar la promoción por promedio (media aritmética) por algún otro sistema que emplee también una medida de tendencia central diferente, la mediana o el modo.

La mediana presenta la desventaja de no tomar en cuenta los puntajes extremos. Si un alumno tuviera las siguientes calificaciones: 5; 6; 7; 8; 9, la mediana sería 7 (puntaje debajo del cual cae el 50 % de las calificaciones). Pero seguiría siendo 7 si el 5 fuera reemplazado por un 1 y el 9 por un 10. No es, por lo tanto, más representativa del nivel de aprendizaje del alumno que la media aritmética. En cuanto al modo, que es el puntaje que se presenta con mayor frecuencia en una serie de calificaciones de un alumno, sólo sería realmente representativo si el estudiante tuviera un número grande de calificaciones, lo cual no es muy habitual. Es corriente, por otra parte, que si la escala tiene un número de grados superior a cuatro

o cinco, del procesamiento de los puntajes de un alumno se deriven dos o más modos. Podemos afirmar, sin embargo, que si se considera adecuado utilizar una medida de tendencia central, se podría interpretar que el modo es más descriptivo del nivel de aprendizaje típico alcanzado por el alumno que la media aritmética o la mediana.

El problema parece ser que es necesario encontrar un sistema que permita tomar en cuenta todas las calificaciones del alumno, incluyendo los que se encuentran alejados del punto medio de la distribución, ya que ellos reflejan, también con claridad, los déficit y las posibilidades del estudiante.

Un régimen que permite paliar algunas de las deficiencias del promedio (media aritmética) y que conserva la practicidad de éste, permitiendo, a su vez, tomar en cuenta todos los puntajes obtenidos, es la media ponderada. Para calcularla se establece previamente una categorización de las producciones del alumno, según su carácter, su extensión y su importancia, y luego, una ponderación numérica para cada tipo de trabajo o instancia de evaluación. Esta categorización y su ponderación deben ser comunicadas a los alumnos. A la hora de calcular el promedio (media aritmética), cada calificación se multiplica por la ponderación correspondiente. Se suman todos estos resultados y a continuación se divide la suma total por la suma de las ponderaciones empleadas. Si, por ejemplo, los trabajos han sido clasificados por su importancia en tres tipos: los muy importantes, los medianamente importantes y los importantes (ya que suponemos que si no fueran de algún grado efectivo de importancia no serían tomados en cuenta) y se hubieran asignado respectivamente las siguientes ponderaciones: 3; 2 y 1, entonces, el cálculo de la media aritmética ponderada se efectuaría como sigue:

Calificaciones obtenidas por un alumno X

Categoría "muy importantes":	9 (3); 8 (3); 9 (3)
Categoría "medianamente importantes":	5 (2); 8 (2)
Categoría "importantes":	4 (1); 6 (1)

La media ponderada se calcula del siguiente modo:

$$9 (3) + 8 (3) + 9 (3) + 5 (2) + 8 (2) + 4 (1) + 6 (1)$$

$$= \frac{3 + 3 + 3 + 2 + 2 + 1 + 1}{27 + 24 + 27 + 10 + 16 + 4 + 6}$$

$$= \frac{114}{15}$$

$$= 7,60 \text{ puntos}$$

Si se calculara una media aritmética no ponderada, la calificación promedio sería de 7 puntos. Como se ve en este caso, el mayor peso de las instancias "muy importantes" mejora la calificación final.

Se trata en los casos que hemos presentado hasta aquí, de calificaciones puntuales, obtenidas en instancias separadas una de la otra, pero se puede adoptar otro punto de vista respecto de la evaluación final de un curso. De Ketele (1984) atribuye a la evaluación, entre otras, la función de certificar que la persona evaluada posee los conocimientos suficientes para pasar al próximo año, al curso siguiente o a la vida profesional, según proceda. Pero, en esos casos, señala la importancia de que se tomen en cuenta dos aspectos: 1) "este tipo de evaluación se refiere a los objetivos terminales, es decir, a los macroobjetivos que integran un número relativamente importante de objetivos intermedios o microobjetivos, así pues, la evaluación terminal no se refiere a estos últimos. Nos atreveríamos a decir —afirma De Ketele— que en cierta manera pueden olvidarse como tales, es su integración lo que nos interesa"; y 2) "la certificación es una decisión dicotómica: el alumno tiene o no la competencia mínima; interesa, pues, definir bien los criterios" (pág. 30). Así, de modo consistente, en el modelo de la Pedagogía de la Integración que propone De Ketele (1993), la evaluación certificativa final toma como objeto para evaluar el objetivo terminal de integración, en tanto los objetivos intermedios son evaluados por evaluaciones formativas o diagnósticas (pág. 70). En atención a estas observacio-

nes, por ende, conviene emplear otro régimen de promoción, diferente de la calificación promedio. Pero, conviene señalar, que si éste último fuera el régimen establecido en la institución o en el sistema escolar, una adaptación que lo mejoraría consistiría en la introducción del concepto de una calificación promedio ponderada (media ponderada) en la que la ponderación se estableciera diferenciando el peso de las calificaciones que corresponden a los que De Ketele denomina microobjetivos, con un coeficiente de valor 1, por ejemplo, y macroobjetivos con un coeficiente de valor 2 o 3.

Régimen de promoción por logros mínimos exigidos

Este régimen se establece sobre la base de la exigencia de un cierto rendimiento mínimo preestablecido en todas las competencias determinadas como significativas para la promoción. A diferencia del régimen de promoción por promedios, la decisión de promoción no resulta del cálculo de una sumatoria ni de una síntesis que globaliza rendimientos heterogéneos sino del análisis del cumplimiento de cada tarea considerada por sí, en tanto indicadora del logro de la o las competencias correspondientes, las cuales no pueden ser sustituidas por otras competencias distintas. El propósito es garantizar que el estudiante haya alcanzado todos los saberes que se consideren necesarios. Los niveles mínimos de calidad de la tarea para su aprobación son fijados previamente, por lo general, por el docente o por la escuela. Este régimen tiene algunas ventajas sobre el de promoción por promedios ya que asegura un mejor trabajo del alumno a lo largo de todo el curso y hace inútiles las especulaciones con la calificación que son típicas de ese régimen. Pero presenta, también, algunos riesgos que es imprescindible tener en cuenta para evitar su efecto perjudicial. Entre ellos, tiende, al igual que la promoción por promedio, a fijar un tope muy bajo de rendimiento, al nivel del que se ha fijado como rendimiento mínimo aceptable. Si en una escala 0-10, el mínimo de aprobación se estableciera en 5 puntos, éste se convertiría en el objetivo de aprendizaje de muchos alumnos. Los grados superiores de la escala podrían ser percibidos como superávit no necesarios por algunos estudiantes.

Existe siempre, por otra parte, el riesgo de que un alumno que hubiera fracasado en una de las tareas iniciales del curso, en el

entendimiento de que la promoción para él sería imposible o, al menos, muy difícil de alcanzar, perdiera interés y abandonara ese curso sin hacer el esfuerzo necesario para remontar los resultados negativos del comienzo.

Estos problemas, en un régimen que tiene ventajas, pueden ser obviados hasta cierto punto por un buen manejo del docente. Se debería mostrar la importancia de los aprendizajes profundos frente a los superficiales que el régimen podría estar estimulando y ofrecer oportunidades reales de recuperación en relación con tareas en las que el alumno hubiera mostrado tener dificultades, eliminando el carácter definitivo de las desaprobaciones iniciales.

La determinación de los estándares mínimos de aprendizaje para la aprobación es una labor de gran importancia y complejidad. En primer lugar, los estándares deben estar definidos con la mayor claridad posible. Además, no deben ser demasiado bajos porque anularían la eficacia del sistema. Pero tampoco excesivamente altos porque se convertirían en exigencias imposibles de alcanzar. En ambos casos, de excesiva facilidad o excesiva dificultad, los niveles de aspiración de los alumnos sufrirían un descenso, generando un nivel correlativo menor de calidad del aprendizaje. Por lo demás, y desde un punto de vista práctico, la implementación del régimen por logros mínimos no presenta mayores dificultades siempre que se compensen los peligros mencionados con estímulo adecuado para que cada alumno alcance su mejor rendimiento de acuerdo con sus posibilidades y se evite la fragmentación de los aprendizajes en objetivos parciales.

Régimen de promoción por examen final

En este régimen se requiere que, para obtener su promoción, el estudiante apruebe una instancia de evaluación escrita, oral o de actuación, o una combinación de dos o de todas ellas. El examen debe estar diseñado de manera tal que permita apreciar en síntesis el aprendizaje logrado por el alumno a lo largo de todo el curso.

No puede tratarse de una simple prueba de actuación o ejecución de una tarea, sino que debe tener un alto grado de validez de contenido y de validez predictiva. Debe representar, en consecuencia, una muestra significativa del universo de contenido de la

asignatura, a la vez que debe predecir eficazmente el desempeño posterior del alumno, sea dentro o fuera del ambiente escolar.

Puede utilizarse de diferentes maneras:

- Como criterio único para decidir la promoción de los alumnos, en cuyo caso el trabajo a lo largo del curso tiene como un objetivo principal preparar al estudiante para la aprobación de la instancia de evaluación final. Las calificaciones parciales, si existen, sirven de orientación al aprendizaje y a la enseñanza y, a veces, para seleccionar a los alumnos a los que se permite rendir el examen final con probabilidad de éxito.
- Combinado con otras formas de promoción, el examen final provee calificaciones que se computan en la calificación que decide la promoción. Un ejemplo es el cómputo del puntaje obtenido en este examen como parte de una calificación promedio, ponderada o no.
- Como supletorio de otras formas de promoción cuando el alumno no ha obtenido ésta por medio del sistema regular establecido en primer término.

El examen final o instancia evaluativa equivalente, presenta ventajas y también desventajas tanto cuando se emplea como régimen único como cuando sólo tiene una participación parcial en la decisión de promoción.

Algunas desventajas se originan en el uso del examen como técnica de evaluación y otras en la adopción del examen como criterio definitivo para la promoción. En lo que respecta a la validez predictiva de los exámenes, ésta requiere que permita un pronóstico de las competencias de los alumnos en un período extenso y en situaciones escolares y extraescolares. Las pautas utilizadas como criterio de evaluación deben permitir la elaboración de inferencias de alto grado de probabilidad acerca de su desempeño futuro. "Como regla -decía Kurt Lewin- cuanto más macroscópica sea la situación que tenga que ser descrita, más largo será el período que habrá que observar para determinar la dirección y la velocidad de la conducta en un tiempo dado."*

* Citado por Ernest R. Hilgard: *Teorías del aprendizaje*, México, FCE, 1961, pág. 161.

La permanencia y desarrollo de las competencias que deben ser pronosticadas en los exámenes escolares se extienden de modo muy prolongado en el tiempo futuro ya que la escuela tiene fines de largo alcance. El examen, por tanto, de una extensión temporal acotada, no puede ser utilizado como único criterio ya que describe sólo la actuación del alumno en una situación de breve duración. Las hipótesis acerca de las ejecuciones futuras del alumno están aquí en dependencia de su rendimiento en un período de duración muy limitada y están viciadas por la emergencia de factores circunstanciales. Por esta razón no debiera utilizarse de modo exclusivo para decidir acerca de cuestiones fundamentales relativas a la selección o promoción de los alumnos. Su validez y confiabilidad son mayores cuando se emplea complementariamente a otros instrumentos de evaluación o de manera supletoria, como antes se dijo.

El sistema de promoción que incluye una instancia de evaluación final tiene un valor, es menester señalarlo, que lo hace irremplazable en un programa de evaluación, ya que permite evaluar de manera completa, no parcelada, el dominio alcanzado por el alumno sobre el universo de los contenidos de la asignatura, el área o un conjunto de asignaturas. En los sistemas de promoción por promedio y de logros mínimos exigidos se emplean evaluaciones parciales, por lo que en ellos queda siempre el riesgo de que falte la apreciación global del aprendizaje del universo de contenidos que se debe tomar como unidad. Y, como se ha señalado más arriba, el problema principal no es la cuestión de que, en el marco de la evaluación considerada aisladamente, se reduzca el grado de validez y confiabilidad de las calificaciones asignadas, sino la influencia que ello tiene sobre el aprendizaje que, consecuentemente, hacen y harán los alumnos. El efecto retroactivo de la modalidad de la evaluación elegida es negativo en este caso, porque induce a los alumnos a un estudio que puede ser muy serio y riguroso en un enfoque analítico pero pobre desde la perspectiva del establecimiento de relaciones entre las partes y del desarrollo de la capacidad de construir una visión integral de los contenidos estudiados.

Régimen de "promoción sin examen"

La modalidad de régimen de promoción que suele denominarse "promoción sin examen" se define por su diferenciación respecto del empleo de un examen final. No es un régimen del que se pueda ofrecer una definición unívoca. Los programas de evaluación en los que se expresa pueden asumir diversas formas, tratándose, por lo general, de la aplicación de un régimen de evaluaciones parciales cuyas calificaciones se procesan según el régimen de promoción por calificación promedio o de logros mínimos exigidos. En muchos casos se incluye una instancia de evaluación final integradora que ocupa el lugar de una evaluación parcial más, como las otras anteriores y, a la vez, de examen final por el papel que desempeña en el proceso de aprendizaje del alumno en la fase de preparación para enfrentar esta instancia que hace las veces de un examen final complementario.

Una modalidad peculiar que puede adoptar es la que corresponde a un programa de evaluación continua basado sobre el análisis, interpretación y apreciación de los registros sistemáticos de las observaciones efectuadas por el docente respecto de las conductas, los procesos y las producciones realizadas por los alumnos en situaciones no puntuales de evaluación. Estas últimas son situaciones de trabajo no diferenciadas de la tarea de aprendizaje. Orientación y guía del aprendizaje y evaluación del aprendizaje son tareas que el docente realiza de manera simultánea. La evaluación es continua y no se diferencian las instancias o momentos destinados a la evaluación de aquellos en los que el docente "enseña". Se trata de un trabajo de carácter clínico, muy individualizado, que se puede llevar a cabo cuando se trabaja con grupos pequeños y en el que la autoevaluación del alumno juega un papel de gran importancia, ya que no sólo es un instrumento útil de evaluación sino una meta de la enseñanza por su valor intrínseco en la formación del alumno.

Régimen de promoción por puntajes derivados

Los puntajes derivados son normas resultantes del procesamiento estadístico de los puntajes brutos obtenidos efectivamente por los alumnos del grupo que se toma como referencia. Los grupos son,

usualmente, el grupo de clase a la que pertenece el alumno, su cohorte, su grupo de edad o el grado o año de escolaridad que corresponde. El procesamiento de los puntajes obtenidos por los estudiantes cuando les son administradas las pruebas se realiza mediante la aplicación de reglas fijas. Las normas no representan niveles deseables de rendimiento definidos apriorísticamente, sino construcciones de los evaluadores que indican los niveles que se ha demostrado que los alumnos pueden alcanzar de manera efectiva. Se procura corregir, de este modo, el nivel de logro establecido en los objetivos de aprendizaje propuestos a priori por el docente, la institución o el sistema escolar para adecuarlos a las posibilidades reales de aprendizaje que tuvieron los alumnos y que pudieron demostrar en las pruebas de evaluación como resultado de la enseñanza que recibieron. De este modo, los puntajes derivados se diferencian de los que responden a la aplicación de la llamada "evaluación referida a criterios" que emplea, en cambio, escalas que tienen un valor absoluto y que corresponden a los niveles de logro determinados en objetivos previamente establecidos definidos en términos de aprendizajes deseables. Ambos enfoques, si bien diferentes uno del otro, responden por igual a la búsqueda de la construcción de procedimientos algorítmicos para la asignación de calificaciones a los alumnos, de acuerdo con los principios que han sido desarrollados por el programa de investigación científica de la psicología conductista y sus derivaciones didácticas. W. James Popham (1983) señala la importancia y la autoría de esta distinción:

[...] el mérito de haber creado una estrategia de solución le corresponde a Robert Glaser, que en 1963 publicó un ensayo que ya se considera clásico en el cual introdujo las expresiones evaluación basada en pautas de normalidad y evaluación basada en criterios. Glaser empleó el concepto de evaluación basada en normas estadísticas para describir los tradicionales tests de rendimiento en los cuales se pone el acento en la determinación de la posición relativa de un sujeto dentro de un grupo. Por otra parte, aplicó el concepto de evaluación basada en criterios a los tests que identificaban el dominio (o ausencia de dominio) de "conductas específicas" por parte del sujeto examinado (pág. 31).

Igualmente Norman E. Gronlund (1973) define a estas pruebas indicando que el criterio de conducta específica provee un estándar

absoluto con respecto del cual se compara el logro de aprendizaje de un individuo. Mencionando también a Robert Glaser como fuente de esta propuesta, Norman Gronlund indica que en virtud de que es este estándar de criterio más que la posición relativa en un grupo de norma el que es empleado para describir el rendimiento en una prueba, estas interpretaciones son denominadas referidas a criterios (pág. 2). La precisión requerida para la definición de criterios en términos de conductas específicas condujo a postular que los objetivos de aprendizaje debían ser formulados de manera operacional, postura que tuvo cierta dominancia durante algún tiempo en la práctica de la programación didáctica. Sin embargo, y a pesar de esta clara diferenciación que se establece entre los dos tipos de evaluación, Perrenoud (1990) los asocia, entendiendo que ambos están presentes en la evaluación formal:

Las prácticas observables participan de una combinación intuitiva de la evaluación de referencia normativa y de la evaluación de referencia criterial, ambas muy artesanales, cada una de las cuales neutraliza en parte los posibles excesos de la otra (pág. 132).

En el enfoque de la evaluación referida a normas en la que se emplean los puntajes derivados, los tipos de normas de uso más frecuente son los rangos percentiles, las normas de edad y de grado y las normas de puntaje estándar.

1. *Rangos percentiles*: el percentil representa el porcentaje de todos los puntajes brutos que se hallan por debajo de un puntaje bruto dado. La escala de rangos percentiles se confecciona ordenando en forma decreciente los puntajes brutos obtenidos por los alumnos al ser administrada la prueba de evaluación. Se construye una tabla de distribución de frecuencias, anotando junto a cada intervalo de puntaje el número de alumnos que ha obtenido dicho puntaje. Este número se denomina frecuencia (f). Junto a la columna f se ubica la columna de las frecuencias acumuladas que se determina sumando de abajo hacia arriba las frecuencias de los intervalos de puntaje.

El cálculo de los percentiles requiere la aplicación de una fórmula que permite tener en cuenta cuál es el puesto de cada alumno en el conjunto total de la clase. Si se encuentra entre los percentiles 90 a 99,

un estudiante demuestra tener un puntaje superior al del 90 a 99 % de los compañeros que se han tomado como referencia para la construcción de la escala de percentiles. Si se encuentra en el percentil 50, esto significa que la mitad de sus compañeros tienen un puntaje inferior al de ese alumno, lo cual equivale a la mediana. Si se ubica en el percentil 20, su puntaje supera sólo al del 20 % de sus compañeros. La escala de percentiles está construida, pues, sobre la base de la comparación entre los aprendizajes efectivamente logrados por los estudiantes, y se emplea válidamente para establecer comparaciones, pero no permite medir aprendizajes en términos de niveles deseables de logro.

2. *Normas de edad*: para determinar estas normas se administran pruebas a grupos de alumnos, representativos de varios niveles sucesivos de edad. Se computan luego las medidas de tendencia central (media aritmética, mediana y modo) correspondientes a cada uno de los niveles de edad. De esta manera se establece cuál es el puntaje bruto típico para cada grupo de edad en las pruebas que se han utilizado. Se debe tener presente, precisamente, que esos puntajes sólo se refieren a las pruebas utilizadas y que la extrapolación de los resultados a otras pruebas debe efectuarse con mucha precaución para garantizar que se trata de pruebas razonablemente equivalentes entre sí.

3. *Normas de grado*: tienen características similares en cuanto a su proceso de construcción a las que se mencionaron en el apartado anterior. La diferencia estriba en que aquí se adopta como referencia para la comparación el grupo de grado, año o nivel de escolaridad. Las normas de grado están basadas sobre los puntajes brutos obtenidos al administrar una prueba a grupos representativos de alumnos de grados sucesivos. Se calcula luego la mediana (percentil 50) de los puntajes brutos obtenidos por los alumnos de cada grado o nivel de escolarización. Se construyen así tablas de puntajes que permiten comparar el puntaje de cualquier alumno de ese grado o nivel con el de los alumnos considerados típicos de éste.

4. *Normas de puntaje estándar*: en sus varias expresiones, estas normas se construyen sobre la base de la concepción de que el

rendimiento de los alumnos de un curso o cohorte se distribuye en la forma de una curva normal (curva de Gauss). De esta manera se expresa la idea de que en un grupo hay siempre un pequeño número de alumnos de muy buen rendimiento, un número igualmente pequeño de alumnos de mal rendimiento y un número importante de alumnos que tienen un rendimiento mediano (algo superior, igual o algo inferior a la media del rendimiento de ese grupo). Para determinar cuáles han de ser las normas del puntaje estándar es necesario trabajar a partir de los puntajes brutos calculando la media de los puntajes y el desvío estándar.

Así, es posible conocer el puntaje en torno del cual se concentran la mayoría de los puntajes brutos obtenidos por los alumnos en la prueba y, merced al conocimiento del desvío estándar, cuál es la amplitud de la curva.

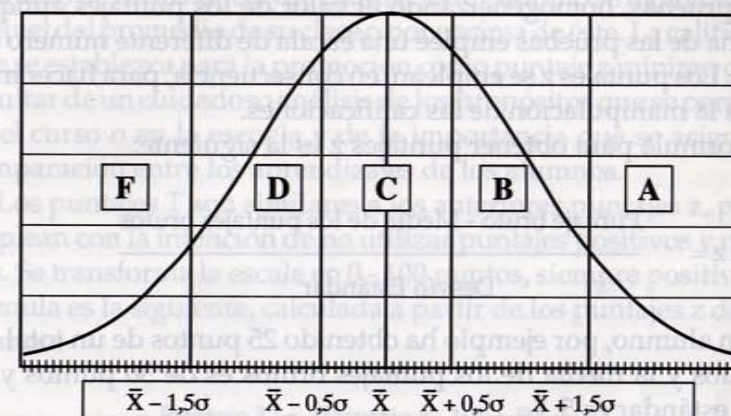
La combinación de estas dos medidas estadísticas permite construir los estándares que se emplearán para obtener los puntajes derivados. Hay tres modalidades que, con escalas diferentes, se emplean con este fin: la "graduación sobre la curva", los "puntajes estándar z " y los "puntajes T ".

a) La graduación sobre la curva trabaja habitualmente con una escala de 5 grados en orden de calidad decreciente de los aprendizajes:

A ; B ; C ; D ; F

En esta escala, los primeros cuatro grados (A, B, C y D) son calificaciones de aprobación, en tanto el quinto corresponde a desaprobación o fracaso (F).

Graduación sobre la curva



Como se observa en el gráfico, la banda que corresponde a los puntajes que habrán de recibir la calificación C se encuentra entre $+1/2$ desvío estándar y $-1/2$ desvío estándar. La calificación B se asigna a los que se encuentran entre $+1/2$ y $+1 1/2$ desvío estándar. Los puntajes que están por encima de $+1 1/2$ desvío estándar reciben la calificación A. Del mismo modo se procede, pero inversamente, para otorgar los puntajes inferiores a C: a los que se encuentran entre $-1/2$ y $-1 1/2$ desvío estándar se les otorga la calificación D y los que están por debajo de $-1 1/2$ desvío estándar reciben la calificación F de desaprobación.

En esta escala se emplea la medida estadística Desviación Estándar que es la medida de variabilidad más utilizada. La escala combina el empleo de la media como medida de tendencia central con el supuesto de que los datos correspondientes se aproximan a una curva normal. Por las características de esta curva, si los puntajes se distribuyeran de acuerdo con ella, el 34 % de los casos se encontrarían entre la media y un desvío estándar por encima de la media y el 34 % entre la media y un desvío estándar por debajo de ella. En una curva con estas características, dentro de la zona delimitada por tres desviaciones estándar por encima de la media y otras tres por debajo de ella, estarían incluidos todos los casos.

Los "puntajes z " se construyen sobre la base de las mismas medidas estadísticas que la "graduación sobre la curva". Estos

puntajes permiten obtener puntajes combinados provenientes de varias pruebas, homogeneizando el valor de los puntajes aunque cada una de las pruebas emplee una escala de diferente número de grados. Los puntajes z se emplean, en consecuencia, para hacer más sencilla la manipulación de las calificaciones.

La fórmula para obtener puntajes z es la siguiente:

$$\text{Puntaje } z = \frac{\text{Puntaje bruto} - \text{Media de los puntajes brutos}}{\text{Desvío Estándar}}$$

Si un alumno, por ejemplo ha obtenido 25 puntos de un total de 40 puntos y la media de los puntajes brutos es de 30 puntos y el desvío estándar es 5:

$$\text{Puntaje } z = \frac{25 - 30}{5}$$

$$\text{Puntaje } z = \frac{-5}{5}$$

$$\text{Puntaje } z = -1$$

Como se ve, si el puntaje del alumno es inferior a la media, su puntaje derivado z será negativo. Si, en cambio, es superior a la media, z será positivo. Cada puntaje z se podrá calcular sobre la base de pruebas cuyo puntaje total es distinto. Si en el caso anterior, el máximo puntaje era 40, en otra prueba podría ser de 25 o de 98. El puntaje z que se obtenga será siempre en unidades positivas o negativas, en enteros o con fracción decimal. La combinación de los puntajes z para definir, por ejemplo, la promoción de un alumno, resultará de la suma algebraica de los puntajes z .

$$\text{Alumno XC: Puntaje } z \text{ combinado} = +1 + 2 + (-3) + 2 + (-1) = +1$$

Se puede determinar que la promoción se obtendrá al conseguir

0 punto z o más. La calificación 0 punto z representa que en el conjunto de las instancias de evaluación el alumno se ha colocado en el nivel del promedio de su clase o por encima de éste. La calificación que se establezca para la promoción como puntaje z mínimo deberá resultar de un cuidadoso análisis de los propósitos que se persiguen en el curso o en la escuela y de la importancia que se asigna a la comparación entre los aprendizajes de los alumnos.

Los puntajes T son similares a los anteriores puntajes z , pero se emplean con la intención de no utilizar puntajes positivos y negativos. Se transforma la escala en 0 - 100 puntos, siempre positivos. La fórmula es la siguiente, calculada a partir de los puntajes z de cada alumno:

$$\text{Puntaje } T = (\text{Puntaje } z \cdot 10) + 50$$

$$\begin{aligned} \text{Por ejemplo: Alumno FK} &= (+2 \cdot 10) + 50 \\ &= +20 + 50 \\ &= 70 \text{ puntos } T \end{aligned}$$

$$\begin{aligned} \text{Alumno JM} &= (-3 \cdot 10) + 50 \\ &= -30 + 50 \\ &= 20 \text{ puntos } T \end{aligned}$$

La escala de puntajes asignados por "graduación sobre la curva" ha sido muy utilizada. También, pero con menor frecuencia, se utilizan las escalas de puntajes z y T .

Como hemos visto, el uso de pruebas estandarizadas se apoya, fundamentalmente, en la necesidad de permitir la comparabilidad de las calificaciones y garantizar su confiabilidad. Pero, como indica Sachs Adams (1970), "no se puede aplicar mecánicamente este procedimiento de calificación en todos los cursos, debido a las diferencias existentes entre los resultados de los diversos cursos, la labor desarrollada en cada uno de ellos, etcétera." Y agrega: "No se pueden evitar preguntas como la siguiente: un curso concreto ¿puede considerarse representativo de la población escolar general?" (pág. 676). Después de analizar diferentes efectos de la adopción de grupos más o menos numerosos o amplios, por un lado, y de grupos más específicos, por el otro, concluye que "1) no es lógico calificar a cada curso según una política de calificaciones uniforme para toda

la escuela; 2) habrá que modificar la política de calificación para cursos determinados sobre la base de las diferencias existentes entre los diferentes cursos con respecto a uno o varios tests correlacionados a las diferencias de competencia individual con respecto a la materia en cuestión" (pág. 680), ya que esas diferencias son aleatorias y si se ignoraran podrían influir de modo decisivo e injusto en las calificaciones de los alumnos menos dotados por comparación con su grupo de compañeros de aula o del resto de la escuela.

Aun cuando se han reconocido ventajas en el uso de pruebas estandarizadas para el cumplimiento de alguna de las funciones de la evaluación, han recibido fuertes críticas, además de la mencionada, en razón de que es innegable que contribuyen a acrecentar la competencia y el conflicto entre los estudiantes porque las altas calificaciones de los alumnos de mejor puntaje bruto incrementan el puntaje de la media y, por tanto, aumentan las posibilidades de desaprobación de un mayor número de estudiantes. Otra crítica de que han sido objeto proviene de la comprobación de que, en algunos casos, por acuerdo entre los alumnos se ha disminuido el puntaje bruto de la prueba intencionalmente para que, al ser inferior la media, los puntajes requeridos para la aprobación resultaran igualmente inferiores. Pero, en última instancia, es el supuesto básico sobre el que se construye esta escala el que merece las críticas más negativas porque es incorrecta la concepción de que los puntajes, y las personas según la calidad de sus aprendizajes o en relación con sus talentos, se distribuyen en la forma de una curva normal. En primer lugar, porque ésta no es una ley de la naturaleza. Los distintos aspectos que se evalúan pueden distribuirse y, por tanto, representarse con diversos tipos de curvas. Si un instrumento de evaluación se utiliza para evaluar los conocimientos básicos que todos los alumnos deben haber aprendido, la curva debería tener una amplitud cero. Y respecto de los talentos, en realidad se puede encontrar mucha gente altamente dotada y, de hecho, esto dependerá también del umbral que se considere punto de partida en cada caso. Las diferencias, de ese modo, pueden aparentar ser pequeñas o grandes, lo cual sería producto de una combinación del efecto umbral y de la escala que en cada caso se utilizara.

Evaluación de portafolio

Se trata de una evaluación que se realiza sobre las producciones de los alumnos. Al modo en que un profesional presenta su trabajo y se presenta a sí mismo en una situación de búsqueda de empleo o similar, el alumno selecciona sus mejores trabajos y los pone a disposición del análisis del docente. Para el docente y para el alumno, esta carpeta o portafolio de trabajos tiene la función de representar los aprendizajes que el alumno ha realizado a lo largo de su curso. Esta estrategia de evaluación se apoya sobre varias ideas: la evaluación de certificación o promoción no se efectúa sobre la base del desempeño del alumno en una instancia desarrollada en un lapso breve. Los trabajos son los que se han ido produciendo durante el transcurso de todo el curso. El alumno hace una evaluación de sus propios trabajos, debiendo analizarlos en sus logros, sus errores y sus dificultades. La forma de presentación de los trabajos puede ser variada, incluyéndose producciones escritas, gráficas, grabadas, vídeos, objetos, etcétera. El alumno puede mostrar la evolución de su desempeño, su mejoramiento progresivo, discutir los problemas y aspectos que no ha podido resolver todavía. Quienes utilizan esta modalidad de evaluación señalan la importancia pedagógica de que la evaluación en una situación no genera los niveles de ansiedad típicos de las situaciones de examen, el valor que se da a la responsabilidad del alumno, su papel en la puesta en juego de los mecanismos de autoevaluación que acompañan a todo el proceso de producción y la autoestima que contribuye a desarrollar en el alumno al reconocerlo como participante necesario y no como objeto del proceso de evaluación. El logro de estas características positivas depende de la actitud con la que el docente asuma su función evaluadora y de la relación que, en consecuencia, entable con sus alumnos. Como forma única de decidir la certificación de los aprendizajes o como complementaria de otras modalidades, la evaluación de portafolio es insustituible a la hora de evaluar la producción de los alumnos. A las otras ventajas que ya mencionamos debemos agregar que es un factor determinante en la programación de actividades de aprendizaje, ya que requiere que los alumnos cuenten con trabajos realizados por ellos, esto es, con una carpeta o portafolio de trabajos, lo cual estimula la adopción de un

modo de enseñar y de aprender centrados en tareas productivas, no rutinarias. Esto por sí mismo es bueno desde la perspectiva de la formación que la escuela debe brindar.

La promoción y los juicios de valor

Hemos desarrollado los principios básicos de seis modelos de régimen de promoción. Son los que presentan mayor frecuencia de administración en las instituciones escolares.

Un problema que ha dado lugar a controversias es el de la importancia otorgada a la cuestión de la comparabilidad. Diversos autores adoptan una postura crítica ante el establecimiento de sistemas de calificación y regímenes de promoción en los que la comparación entre alumnos juega un relevante papel. Entre ellos, Cardinet (1989) señala las diferencias esenciales que surgen de la contrastación de los tests psicométricos que sirvieron en su momento de modelo a las pruebas formales de evaluación, y estas últimas cuando se administran con sentido pedagógico. En los tests psicométricos se procura elaborar un diagnóstico exhaustivo pero con el mínimo número de datos necesarios (por aplicación del principio de parsimonia). Se buscan, también, características estables, estandarizadas, y las escalas cuantitativas que se emplean son relativas a los estándares "normales" para la población estudiada. En pedagogía, en cambio —sostiene Cardinet—, la búsqueda de la parsimonia es totalmente inexistente. Todos los contenidos son esenciales. Una habilidad ausente no puede compensarse con la presencia de otra habilidad. Todas las combinaciones posibles de contenidos, de procedimientos cognitivos y de situaciones de observación, constituyen objetivos independientes e irreductibles. "El deseo de obtener características estables —dice Cardinet— está muy ausente en la escuela" (pág. 45 y sig.). Las escalas cuantitativas son, además, más bien absolutas que relativas. El papel de la situación es, finalmente, de enorme importancia para que el evaluador pueda comprender el significado de la conducta del alumno. Acentúa Cardinet, por otra parte, la necesidad de incluir en el análisis la relación del evaluador y el evaluado como un aspecto determinante de esa situación. Ante las características que son propias del acto de

evaluar, según este autor, la evaluación pedagógica debe evitar la emisión de juicios que serían incompatibles con la ayuda que el docente debe dar a su alumno. Éste sí necesita información sobre su desempeño pero, es la conclusión de Cardinet, sin juicios de valor.

Queremos resaltar, sin embargo, que el juicio de valor que Cardinet rechaza es, básicamente, el que resulta de la comparación entre diferentes alumnos. Queremos pensar, en cambio, en una evaluación que resguarda todas las consideraciones que hemos ido mencionando hasta aquí, que conserva la conciencia de la precariedad de los juicios evaluativos, en la que el evaluador asume el modesto papel que está implicado en el acto de medir aspectos de la conducta humana, en donde una persona evalúa a otra y hay evaluaciones mutuas en el curso de la interacción. Esta evaluación lejos de perder su papel pedagógico lo asume plenamente y se estructura como componente esencial de los actos de enseñanza y de aprendizaje. Para ello es menester que el docente integre a su labor de enseñanza la tarea de la evaluación: "Un sistema de registro acumulativo racional es indispensable —afirma Sachs Adams (1970)— para permitir que la medición y la evaluación alcancen una eficacia máxima para facilitar la comprensión de los alumnos y de sus problemas" (pág. 682). El registro no es suficiente, sin embargo. Hay que analizar e interpretar las informaciones.

La promoción se resuelve siempre a partir de una decisión y ella lleva implícito un juicio de valor. El régimen de promoción que se establezca debe dar a este acto la mayor explicitación posible, de modo de garantizar la transparencia y la justicia de la decisión.

BIBLIOGRAFÍA

- Cardinet, Jean: "Evaluer sans juger", *Révue Française de Pédagogie*, n° 88, julio-agosto-setiembre de 1989.
- De Ketele, Jean-Marie: *Observer para educar*, Madrid, Aprendizaje-Visor, 1984.
- : "L'évaluation conjugée en paradigmes", *Révue Française de Pédagogie*, n° 103, abril-mayo-junio de 1993.

- Gronlund, Norman E.: *Preparing Criterion-Referenced Tests for Classroom Instruction*, The Macmillan Co., Collier-Macmillan Publishers, 1973.
- Hotyat, E.: *Los exámenes*, Buenos Aires, Kapelusz, 1965.
- House, E. R.: *Evaluación, ética y poder*, Madrid, Morata, 1994.
- Lavin, David E.: *The Prediction of Academic Performance*, Russell Sage Foundation, 1965.
- Perrenoud, P.: *La construcción del éxito y el fracaso escolar*, Madrid, Paideia-Morata, 1990.
- Piobetta, J. B.: *Exámenes y concursos*, Buenos Aires, Kapelusz, 1952.
- Popham, W. James: *Evaluación basada en criterios*, Madrid, Magisterio Español, 1983.
- Purkey, William W.: *Self Concept and School Achievement*, Prentice-Hall Inc., 1970.
- Sachs Adams, Georgia: *Medición y evaluación en educación, psicología y guidance*, Barcelona, Herder, 1970.
- Thorndike, Robert E. y Hagen, Elizabeth P.: *Medición y evaluación en psicología y en educación*, México, Trillas, 1991.
- Wiliam, Dylan: "Some technical issues in assessment: a user's guide", *British Journal for Curriculum and Assessment*, 2 (3), 1992.

Este libro se terminó de imprimir en el mes de febrero de 1998, en los Talleres Gráficos D'Aversa, Vicente López 318 Quilmes, Buenos Aires. Rep. Argentina.