



Investigación Educativa
Vol. 17 N° 31 93 - 108
Enero-Junio 2013
ISBN N° 1728-5852

EVALUACIÓN PSICOMÉTRICA Y EVALUACIÓN EDUMÉTRICA ¹

PSYCHOMETRIC EVALUATION AND EDUMETRIC EVALUATION

Rubén Mesía Maraví ²
Augusto Frisancho León ³

RESUMEN

El artículo aborda un asunto muy importante que involucra a todos los niveles educativos, cual es el proceso de evaluación del aprendizaje. Se pretende realizar una reflexión argumentada e informada acerca de ese aspecto trascendente que se lleva a cabo en forma cotidiana en el proceso educativo, pero sobre el cual no se reflexiona ni se discute lo suficiente. Es paradójico: en el sistema educativo se hace uso intensivo de la evaluación, pero los cambios que se producen son muy pocos. ¿Se está evaluando con eficacia?

Se trata de un asunto neurálgico y complejo que es necesario abordar con diligencia y seriedad si se quiere innovar el proceso evaluativo. Por ello se presenta una síntesis desde diversos puntos de vista, que pueda servir a la vez tanto de información actualizada como de base para la reflexión y discusión en el cotidiano quehacer educativo.

Palabras clave

Evaluación, medición, calificación, normotipo, evaluación normativa, evaluación criterial.

ABSTRACT

The article discusses a very important issue that involves all levels of education,

1 Artículo presentado el 3/01/13 y aceptado el 11/04/13.

2 Doctor en Educación. Profesor Principal de la Facultad de Educación, UNMSM. Director del Instituto de Investigaciones Educativas.

3 Magister en Medición y Evaluación de la Calidad Educativa. Profesor de la Facultad de Educación, UNMSM.

which is the process of assessment of learning. The aim is to make a reasoned and informed reflection about the transcendent aspect takes place on a daily basis in the educational process, but which do not reflect or discussed enough. It is paradoxical: in the educational system makes extensive use of the evaluation, but the changes that occur are very few. Does being evaluated effectively?

This is a complex issue and nerve to be addressed diligently and seriously if you want to innovate the evaluation process. Therefore presents a synthesis from various points of view, which can serve both as much updated information as the basis for reflection and discussion in the daily educational work.

Keywords

Evaluation, measurement, qualification, normotipo, normative evaluation, criterial evaluation.

EL SENTIDO DE LA EVALUACIÓN

A la antigua o tradicional concepción de evaluación, como un medio o instrumento de clasificación de los estudiantes, se viene oponiendo la idea de poner de manifiesto o *develar las posibilidades* de cada estudiante para que pueda desarrollar su potencialidad al máximo y al mismo tiempo descubrir sus carencias y limitaciones de modo que pueda remediarlas o eludirlas.

Por otra parte, se tiene que, entre sus múltiples dimensiones, el proceso educativo se presenta como un medio de selección y promoción social. En esta perspectiva la evaluación es una vía de "clasificación" con una dimensión social y selectiva y no propiamente educativa. Porque si nos referimos a una evaluación que sea específicamente educativa, ésta debe tener un *rol orientador* en el desarrollo de las *posibilidades* de estudiante y en la superación de sus *limitaciones*.

Aquellos profesores que entienden la evaluación como un medio de detectar y promover las aptitudes de los estudiantes, así como de satisfacer sus necesidades, se enfrentan con la embestida del "movimiento de los estándares". Ve que todos reclaman la aplicación de "estándares universales" y pruebas estrictas para demostrar cuales son los estudiantes que están a la altura de esos estándares y cuáles son los que no lo están. Sin embargo, parecen no percatarse que la responsabilidad de la institución educativa y la función docente también son evaluadas y se equiparan con el desempeño de los estudiantes en las pruebas estandarizadas.

Los buenos estándares, como otros muchos buenos instrumentos, constituyen un *recurso útil* para el docente. No obstante eso, muchos de esos instrumentos ya establecidos, y la retórica que los rodea, reducen la enseñanza a la transmisión

mecánica de datos y la evaluación a la actividad más mecánica todavía de marcar círculos en una prueba estandarizada de opciones múltiples.

Los estándares son necesarios para *discriminar* con claridad realizaciones aceptables de las no aceptables. Se utilizan diferentes nombres como sinónimos: *referentes, normas, estándares o normotipos*. Tradicionalmente se señalan tres tipos de estándares, en función de los cuales se habla de tres tipos de evaluación, aunque los autores no se ponen de acuerdo en la terminología utilizada. Se tiene así la evaluación relativa o normativa, la evaluación criterial o absoluta o basada en la autoridad y la evaluación personalizada o idiográfica o idiosincrásica o de progreso.

NORMOTIPOS DIVERSOS

Es necesario tener siempre en cuenta que evaluar es *comparar* y que la comparación es inherente al proceso evaluativo. Por otra parte, cuando se evalúa se realiza tanto una medición como una valoración, basadas, precisamente, en dicha comparación. Ésta para poder llevarse a cabo, va requerir de la existencia de un punto de referencia, de un *referente*, con el que se comparan los datos recogidos en la medición. Este punto de referencia para la medición es lo que se ha dado en llamar *normotipo*.

Se presentan *tres modalidades* básicas de normotipo con sus diferentes matices: el normotipo estadístico, el normotipo de criterio y el normotipo individualizado. Cada uno de ellos con sus respectivos rasgos distintivos:

- a. El normotipo *estadístico* hace la comparación con respecto a un grupo de referencia, a lo que es o se considera "normal" en dicho grupo. Se le emplea para la construcción de percentiles y baremos y con poca frecuencia en la evaluación didáctica.
- b. El normotipo *de criterio*, que establece la comparación con respecto a un modelo general o a un marco de referencia externo que especifica las características que ha de reunir el objeto evaluado para ser valorado positivamente. Este marco externo es el "criterio". En la evaluación didáctica el criterio principal que se utiliza son los objetivos, cuya especificación configura el producto que se pretende lograr. Lo que se hace es comparar los resultados logrados con los objetivos previstos.
- c. El normotipo *individualizado*. Aquí lo importante son los datos de la "medición" de un resultado y sus valores desde la perspectiva de la evolución del sujeto. En la enseñanza se comparan los resultados de un

estudiante con lo que era a comienzo del curso y se valora su calidad en función de la mejora, de las ganancias o pérdidas que ha manifestado a lo largo del periodo a que se refiere la evaluación.

Cada clase de normotipo presenta virtualidades y riesgos, la habilidad docente decidirá la oportunidad del empleo de uno u otro. Así, se puede utilizar el normotipo de criterio cuando hay un marco de referencia claro y consensuado que se puede tomar como criterio. De lo contrario, sería mejor emplear el normotipo estadístico. No obstante, aun existiendo esos marcos-criterio (Programas oficiales), cuando la situación específica de los estudiantes no coincide con las presunciones oficiales, como en las zonas deprimidas o en los grupos marginales, comparar en estos casos los resultados logrados con los criterios oficiales supondría tener que desaprobado a todos, lo cual es en sí mismo injusto y contraproducente.

LAS DOS TENDENCIAS BÁSICAS

En la nueva concepción educacional se dan dos tendencias predominantes para enfocar la medición y la apreciación de los resultados del aprendizaje de los estudiantes, lo cual se exterioriza en dos aspectos: a) La manera de elaborar los instrumentos de medición y b) La forma de interpretar los resultados del proceso de enseñanza aprendizaje.

Es en este sentido que la interpretación de los resultados de las pruebas de evaluación, o sea, el juicio sobre la situación del estudiante puede hacerse respecto a la norma o respecto al criterio, lo cual señala y da lugar a las dos tendencias o enfoques de la evaluación: la evaluación referida a la norma o evaluación *psicométrica* y la evaluación referida al criterio o evaluación *edumétrica*.

Desde otro punto de vista, es posible categorizar las diversas modalidades de la evaluación en función del referente con respecto al cual se compararán los datos obtenidos en la primera fase del proceso. Dichos referentes pueden ser, básicamente, a) El rendimiento previo del estudiante, b) El rendimiento medio de un grupo de referencia y c) Un criterio externo previamente decidido. El primero y el tercer referentes dan lugar a la evaluación *criterial*, en la cual el rendimiento del estudiante es valorado de manera absoluta, mientras que el segundo va a dar lugar a la evaluación *normativa*, la cual es relativa y de carácter estadístico.

Sin embargo, se da un tercer caso en el cual la comparación se hace en función de un referente interno, propio. De este modo dicha comparación se establece

con el propio objeto evaluado (el estudiante). Se compara, entonces, la situación inicial con la situación actual del objeto de evaluación, lo cual da lugar a la llamada evaluación *idiográfica* o personalizada.

La Evaluación Psicométrica

Este tipo de evaluación, llamada también *normativa*, es habitual en las prácticaspsicométricas, que descansan en el supuesto de una *distribución normal*, estadísticamente hablando, de las diversas aptitudes y demás constructos psicológicos en la población general. Trata de captar las diferencias inter individuales comparando las puntuaciones de cada sujeto *con las de su grupo* de pertenencia, al tiempo que obtiene una idea global de la ejecución de cada individuo. Es, por ejemplo, la forma habitual de proceder en la medición de la capacidad intelectual. Ésta orientación está ligada a la Teoría de la Medición clásica yes importante para la elaboración de tests estandarizados orientados a obtener normas (curvas, baremos) locales o nacionales.

Hay muchas concepciones o "modelos" de normalidad: jurídica, psicológica, estadística, sociológica, pedagógica, etc. Sin embargo, siempre se suele entender algo por "normal" y que cuando no se cumple preocupa e inquieta. Esta concepción evaluativa suele ser utilizada también en el ámbito del proceso educativo pero ha generado muchas *críticas* y *confusiones* en cuanto a su naturaleza y empleo eficaz para evaluar el aprendizaje de los estudiantes. Así, entonces, aplicada a la función evaluadora de la enseñanza, su ejecución se corresponde con el criterio estadístico, en el que se considera que el estándar de rendimiento es relativo y está representado *por la norma de una curva de distribución*, de carácter general.

En el proceso educativo este movimiento de medición educacional se ha basado en el desarrollo y logros de la teoría psicométrica y en la observación de que las calificaciones escolares no eran confiables y carecían de objetividad. En consecuencia, los instrumentos desarrollados por los expertos en medición dieron las bases conceptuales para la evaluación. Dichos instrumentos proporcionan *puntajes* y otros *índices* diversos que pueden ser manejados matemática y estadísticamente, lo cual permite manejar masas de datos, comparar resultados y establecer normas.

Pero se da el hecho de que con este enfoque instrumentalista la evaluación se convierte en una *técnica de construcción de instrumentos* y de interpretación de resultados. Evaluar llega a ser sinónimo de *construir* y *aplicar instrumentos* y el evaluador se limita a evaluar aquellas variables que puede medir. En consecuencia, aquellas variables para las que no tiene instrumentos las llama

"intangibles" y que no tienen importancia, precisamente, porque no se pueden medir. Entonces, la falta de instrumentos pone límite a la evaluación y eso se considera que son sus límites naturales o reales. El instrumento aparece así como una herramienta y por eso libre de juicios de valor y cultural y sociológicamente antiséptico. Por eso es que se le podría cuestionar, por ejemplo, su precio o su apariencia pero no sus resultados.

Se trata de obtener resultados que se ajusten a la curva normal, lo que vendría a ser la expresión "natural" de un rendimiento educacional. Pero se cuestiona esta dependencia de la curva normal, porque ella *no es infalible*. Describe el producto de un suceso fortuito y la educación, en cambio, es una actividad intencionada en la que se persigue que los estudiantes aprendan lo que se les enseña, por lo tanto, la curva representativa del rendimiento debería ser *muy diferente* a la curva normal. Más bien, se podría decir inclusive que los esfuerzos educativos no han sido exitosos en la medida en que el rendimiento de los estudiantes ha venido siendo distribuido normalmente (con referencia a la curva normal).

Sucede en la práctica docente cotidiana que la estructura cognitiva de muchos profesores se identifica con una *producción final cuantitativa* de la evaluación. Esto ocurre porque el modelo tradicional que vienen utilizando conlleva un afán de querer convertir o traducir los hechos educativos en cuantificables o mensurables. El modelo estadístico de normalidad define lo normal como todo aquello que se ajusta a la medida. Por consiguiente, y en este sentido puramente cuantitativo, anormal sería todo aquello que se aparta de dicho término medio, de lo ordinario y frecuente. Desde esta perspectiva se trata de situar al individuo en posiciones de "sigma" en relación a una *curva normal* o gaussiana (Curva o campana de Gauss como expresión de la normalidad).

La distribución normal es una curva simétrica en forma de campana, cuya mitad izquierda es reflejo en espejo de la mitad derecha. En ella la media, mediana y moda están localizadas en un mismo sitio, en el centro exacto de la curva. La mayoría de los valores están agrupados alrededor de estas medidas de tendencia central. Casi todos los valores están dentro de tres unidades de desviación típica a cualquiera de los lados de la media. En esta curva normal más de los dos tercios de los valores están dentro de una desviación típica de la media, menos del 5% de las puntuaciones están a más de dos desviaciones típicas de la media y un mínimo porcentaje está a una distancia mayor de tres desviaciones.

En tal sentido, un profesor poco avisado que encuentra esa curva después de corregir un examen no tendría preocupación alguna, puesto que *su grupo responde al criterio de normalidad*. Ha ocurrido, para él, lo que tenía que ocurrir, "lo normal", es decir, pocos han obtenido la calificación de notable o

sobresaliente, otros pocos la de deficiente o desaprobado y la mayor parte la de suficiente o aprobado porque se sitúan alrededor del punto medio (la media del grupo). Las preguntas de su examen han "discriminado" correctamente a los que estudiaron de aquellos que no se han esforzado y, claro, a aquellos "que se han defendido", que son la mayoría, "el montón". Queda así contento, sin ningún motivo de preocupación. No se ha percatado que a su evaluación le ha dado un carácter eminentemente selectivo y discriminatorio, no el necesario carácter educativo.

Sin embargo, debe tenerse en cuenta que la curva normal es absolutamente relativa para la evaluación educacional. Al respecto, en palabras de Bloom, se considera que "La curva normal de calificaciones ha sido utilizada por tanto tiempo que los profesores terminaron por creer en ella. Las mediciones del rendimiento tratan de detectar diferencias aún triviales entre los alumnos, de manera que sea posible asignar calificaciones según la norma. Se espera que un 10% de estudiantes obtengan calificación alta y se está listo a desaprobado una proporción similar. Así, los fracasos de los estudiantes están determinados por su posición jerárquica dentro del grupo más que por su incapacidad para las ideas centrales de la materia en estudio".

En suma, este enfoque pretende indagar y establecer diferencias de aprendizaje entre los estudiantes, para así determinar la *ubicación relativa de desempeño* que cada estudiante tiene dentro del grupo del cual forma parte, en un momento y circunstancia dados. Se toma como "norma de desempeño" la ejecución promedio observada en el grupo a su cargo y el juicio valorativo "el mejor o peor estudiante del curso" no indica lo que el estudiante sabe o no sabe de la materia ni garantiza si la situación se mantendrá si se compara al estudiante con un grupo diferente. Sólo indica si es el mejor o el peor del grupo, nada más.

Esta forma evaluativa puede *caracterizarse* por lo siguiente:

- a. Está ligada a la teoría de la medición clásica y todos sus aspectos están muy centrados en la programación y en los contenidos.
- b. Mide las *diferencias interindividuales* con respecto a una característica o rasgo y lo hace comparando las puntuaciones de prueba de cada sujeto con su grupo de pertenencia. Determina así la posición relativa del alumno con respecto al grupo. Por eso es que el desempeño grupal promedio y su variabilidad constituyen la norma o marco de referencia para analizar y valorar la calidad del rendimiento.
- c. El criterio para evaluar es *una norma* determinada establecida de antemano que consiste en una *curva de distribución* que representa el estándar de

rendimiento. La norma está dada por los baremos interpretativos sacados de anteriores aplicaciones de tests. Por tal motivo se le emplea, precisamente, en la elaboración de tests estandarizados para obtener normas (curvas, baremos) locales o nacionales.

- d. Es muy importante el *tamaño de la muestra normativa*, cuanto mayor sea, mejor. De lo contrario no es significativa. De allí la dificultad para aplicarla en la evaluación del aprendizaje en el aula, porque, a menos que una clase sea extremadamente numerosa, no se da una distribución perfecta y normal.
- e. Para llevarla a cabo es necesario contar con una *descripción completa* de la norma del grupo o muestra normativa. Ésta está constituida por los que ya rindieron la prueba y con quienes se compara el desempeño del que la está rindiendo en ese momento.
- f. Los resultados de prueba se interpretan comparando el rendimiento de un estudiante *con el de los demás* del grupo y los juicios evaluativos son *relativos* pues dependen de las características del grupo del que forma parte el evaluado. Además, esos resultados son válidos, útiles y generalizables únicamente dentro del grupo evaluado.
- g. Los resultados dan escasa información sobre el *grado de eficiencia* que tiene el estudiante respecto a las capacidades que se están evaluando. Sólo señala si el estudiante tiene mejor o peor, mayor o menor capacidad que sus compañeros para enfrentar la tarea. Esto puede ser útil sólo para seleccionar a los que están en mejores condiciones o son más aptos para ejecutar una actividad determinada.
- h. Cuanto mayor sea la *variación* o *dispersión* de los resultados de prueba dentro del grupo, más útil y significativa es la comparación entre la puntuación del estudiante y el desempeño del grupo.

Esta forma evaluativa ha sufrido *muchas críticas* y desde diversos aspectos, de modo que se pone en tela de juicio su aplicación en la evaluación sistemática y cotidiana del aprendizaje en el aula. Las críticas más saltantes que se le han formulado son:

- a. La distribución normal de los individuos respecto a determinadas aptitudes sólo se produce cuando el número de casos es *muy alto*. De allí que dentro del proceso educativo es casi seguro que no se da esta distribución si se considera que en salón de clases existe un promedio de 30 o 40 estudiantes.
- b. Los tests empleados miden, más bien, capacidades generales, como la

inteligencia por ejemplo, y casi *no lo hacen con los conocimientos o habilidades* específicas propias de una asignatura o materia.

- c. Los tests normativos tienen un defecto fundamental respecto a su relevancia: están contruidos para que los resultados presenten una distribución normal, por lo que se orientan a *discriminar* entre los alumnos. Así, se tiene que un ítem que es contestado por todos los alumnos no es un buen ítem y se elimina. Cómo es fácil que todos los estudiantes respondan a cuestiones de interés real, porque el profesor ha insistido en ellas, al cabo de varias selecciones o cribas sólo se preguntarán cosas sin importancia. Tampoco son buenos ítems, y se eliminan, los que no son respondidos por ningún estudiante. Entonces el profesor no va a tener datos acerca de que algo importante no ha sido comprendido.
- d. Si se considera que un interés real del profesor es averiguar en qué tienen dificultades los estudiantes, tanto con fines de calificación como con fines de orientación se tiene que esta forma evaluativa no ofrece esa posibilidad. Es que no es posible con ella averiguar *las dificultades ni las fallas* de los estudiantes e, igualmente, tampoco es útil para evaluar el impacto de algún sistema o método de enseñanza.
- e. El estudiante es evaluado respecto a *un grupo normativo ideal*, que sólo existe en la concepción y mente del profesor y el resultado medio con el que realiza las comparaciones se proyecta en un "alumno medio" que es *una ficción* porque en la realidad se tienen estudiantes concretos y singulares.

La Evaluación Edumétrica

Esta forma evaluativa proviene de una concepción tyleriana de la evaluación, la cual "busca esencialmente en qué medida los objetivos son conseguidos y cómo éstos apuntan a la producción de ciertos cambios deseables en el estudiante; la evaluación viene a ser el proceso que determina el grado en que se están verificando esos cambios de conducta". Esta conceptualización tiene el mérito de incorporar la evaluación al proceso educativo, ya no como un agregado sino como *consustancial al proceso* de enseñanza aprendizaje. Sin embargo, los efectos prácticos de esta concepción no han llegado a satisfacer las expectativas creadas por sus propuestas y sus posibilidades.

Ella surge como contraposición a la otra tendencia y ofrece un giro importante en relación con la anterior, tradicional o convencional y muy centrada en la programación y los contenidos. Sin embargo, aunque tiene otra visión evaluativa, sigue estando sometida a la consecución de los objetivos como

criterio de eficiencia y en la valoración final siguen primando los objetivos terminales y el producto alcanzado. El paso de una situación de aprendizaje inicial hacia otra situación final va a requerir una *observación* que sea lo más objetiva posible por parte del profesor y constituyen objeto de observación las *operaciones externas* en todas sus dimensiones mensurables y cuantificables, pero no se preocupa ni nada se sabe de los procesos cognitivos internos por los cuales el estudiante llega al conocimiento. Este modelo es de claro corte conductista y propone el planteamiento de la enseñanza a partir de unos objetivos claros y concretos, de modo que sea posible para el evaluador observar con claridad si los cambios en la conducta del estudiante se han producido efectivamente. Pero no sólo consiste en una observación clara y precisa de dichos cambios, sino que tiene en cuenta también en qué cantidad y condiciones esos cambios se han llegado a producir, de modo que puedan, incluso, ser medidos y que el profesor pueda *objetivar* y *cuantificar el logro* obtenido. Ya la referencia no va a ser el grupo sino *el propio conocimiento* del estudiante en relación con la materia y con los objetivos que le correspondía alcanzar en un momento determinado.

Esta es una forma de evaluación basada en objetivos y se llama también evaluación referida a criterios. Lo que pretende medir son los *cambios intraindividuales* relativamente estables y determina si el estudiante logra el dominio de ciertos objetivos específicos o criterios (objetivos suficientemente operativizados y jerarquizados que el estudiante debe alcanzar). Aquí se confronta al estudiante con unas metas previamente establecidas y se procura que los juicios valorativos sobre el aprendizaje de los estudiantes expresen el grado en que cada uno de ellos ha logrado los objetivos propuestos. Los resultados de prueba se comparan con *un criterio absoluto* que es lo que a juicio del profesor debe saber hacer el estudiante. A este "saber hacer" se conoce como "dominio" o "competencia", lo cual implica tanto el tipo de tarea que ha de ejecutarse (conducta) como el contenido o materia involucrada en la ejecución (saberes).

El test basado en criterios va a determinar la posición de un individuo con respecto a un *dominio de conducta* perfectamente definido, que no es sino un conjunto de habilidades o disposiciones que los examinados ponen en juego en una situación de examen. Por eso los tests necesitan que los objetivos estén perfectamente definidos operativamente y que se establezca un nivel mínimo de realización que se considere aceptable. El aspecto más importante es *situar al estudiante* en el punto en que se encuentra respecto al dominio de cada tarea, así los puntajes de la prueba van a indicar si el estudiante puede o no realizar una tarea o exhibir una conducta adquirida como efecto del proceso de enseñanza aprendizaje.

El dominio sirve como *marco de referencia* para comparar y determinar lo que puede hacer el estudiante y si su ejecución es igual o inferior al aprendizaje deseado. No se toma en cuenta para efectos de comparación los resultados del resto del grupo evaluado. Al definir el dominio que sirve de base para elaborar la prueba y que evaluará la calidad de la ejecución respecto a un objetivo, se define también el "criterio de éxito" esperado. Estas pruebas se llaman pruebas edumétricas o pruebas referidas a criterios. En la prueba cada pregunta debe relacionarse íntimamente con los objetivos del dominio, garantizando que existe *congruencia* entre los objetivos y los ítems elaborados para su comprobación.

El criterio de éxito o *nivel mínimo de aceptación* debe definirse y establecerse antes de la administración de la prueba. Dicho nivel mínimo debe ser resultado de lo que el docente determina como *mínimo razonable* para el logro de objetivos de su asignatura, basado en su experiencia, en las características socioculturales de los estudiantes, en las estrategias de enseñanza, en la infraestructura administrativa y material con que cuenta, etc. Ello determinará que no siempre la expresión cuantitativa del logro mínimo sea del 100 % del dominio definido.

En suma, sus *características básicas* son:

- a. Es una forma de evaluación basada en objetivos y mide *cambios intraindividuales*, determinando si el estudiante logra el dominio de objetivos específicos.
- b. Se ajusta a la medición del nivel de *aprendizaje para el dominio* (mastery learning) que establece mínimos esenciales. Se fija con anterioridad el criterio o estándar absoluto representado por los objetivos, lo cual permite confrontar al evaluado con las metas previamente establecidas.
- c. Tiene como referente a *las condiciones del propio sujeto*, a un criterio formulado previamente, conforme a lo que pueda esperarse de él.
- d. Los *aspectos clave* que tiene en cuenta son: la definición del criterio utilizado por el elaborador de la prueba, la relevancia del criterio para los objetivos de la evaluación y la evidencia que se tiene y sirva para apoyar el uso de la prueba con ese propósito.
- e. Toma en cuenta la interacción de todos los elementos del proceso: estrategias docentes; pre requisitos; análisis de necesidades; análisis lógico y psicológico de la materia; técnicas de redacción de objetivos; diferenciación entre objetivos generales, específicos y operativos; revisión de la metodología; pruebas parciales; etc.

- f. Se le emplea en la evaluación educativa para determinar si los estudiantes *han dominado habilidades académicas básicas* y también cuando se debe demostrar el dominio de un material antes de pasar a otro más avanzado.
- g. La adecuación de las preguntas, la validez de contenido o la congruencia ítem-objetivo que es la propuesta considerada como muestra representativa de los aprendizajes evaluados, se decide mediante acuerdo entre los profesores de las asignaturas respectivas.
- h. Las pruebas son más fáciles de diseñar y elaborar, también de interpretar los resultados, de manera que permiten sugerir hipótesis acerca de las causas de las eficiencias y deficiencias de los individuos y los grupos.

Esta modalidad evaluativa también tiene *algunas críticas* referidas a su concepción y a la elaboración y validez de las pruebas. Pueden señalarse las siguientes:

- a. Supone un grado de *complejidad técnica considerable*, pues es necesario determinar estándares con validez experimental de contenido, que dependen, a su vez, de la tarea a realizar y del dominio de la misma. Por eso mismo, *no es fácil que cualquier profesor* lleve a la práctica esta evaluación, porque el aparato técnico necesario no está a su alcance.
- b. En ella es posible observar las actividades y conductas externas mensurables y cuantificables, pero nada se sabe de los *procesos cognitivos internos*.
- c. Puede tener valor en la evaluación del dominio de conocimientos o habilidades básicas, pero *ninguna en el continuo superior* del conocimiento. Es decir, lo es en el dominio de aprendizaje de la lectura, la escritura y la matemática básica, pero no, por ejemplo, en el progreso de un estudiante de doctorado.

ANÁLISIS COMPARATIVO

En el análisis de ambas formas evaluativas es necesario dar a cada una de ellas su lugar, puesto que contraponerlas no reviste ningún interés. La evaluación normativa, por ejemplo, tiene un papel importante en la valoración de las capacidades generales y en la selección profesional, mientras que se puede utilizar la evaluación de criterio cuando hay un marco de referencia claro y consensuado que se puede usar como criterio.

El proceso de enseñanza aprendizaje puede evaluarse, con reservas y con conocimiento de causa, desde *ambas perspectivas* y cada una de ellas implica la

utilización de instrumentos diferentes para estructurar, desarrollar y analizar las situaciones de evaluación. Si los propósitos del proceso educativo son de índole *formativa y remedial*, es conveniente la evaluación edumétrica, pero si tiene un propósito *selectivo* lo es la evaluación psicométrica.

Por otra parte, existen situaciones donde se pueda *integrar* las dos formas de medición, como sucede, por ejemplo, en las investigaciones evaluativas sobre métodos, programas y materiales o en la investigación del impacto de un sistema de enseñanza y su efectividad para determinados logros, etc.

Si se tiene en cuenta que el progreso del estudiante depende de su edad, de su habilidad para aprender, de la efectividad de la enseñanza y de la efectividad de los ítems del test, dada la complejidad y el número de factores presentes, no siempre va ser posible hacer juicios muy precisos acerca de los niveles esperados de rendimiento. Por ello, además de considerar los tests criteriales, es necesario *complementar la interpretación de los resultados* con la dimensión normativa, estableciendo comparaciones entre los sujetos del mismo grupo.

UNA ALTERNATIVA VALIOSA

Existe *otro modo* de interpretar los resultados de la evaluación: esta es no atender a la norma o al criterio externo, sino a *la persona como norma de sí misma*. Esta es la llamada *evaluación personalizada* o *idiográfica* o *idiosincrática* o *de progreso*, en la que la comparación se establece con el propio objeto evaluado y se compara *la situación inicial con la situación actual* del sujeto. Su *referente* es *interno*, es decir, se toma al propio sujeto como referencia; sus logros anteriores son los que se toman en cuenta para evaluar los logros actuales. Es así, entonces, como un rendimiento suficiente para la evaluación normativa o la criterial puede resultar insatisfactorio cuando se tiene en cuenta las posibilidades mismas de la persona. Este juicio podría no tener valor administrativo pero sí educativo porque diferencia dos tipos de juicio: suficiente-insuficiente y satisfactorio-insatisfactorio.

Es ésta una orientación más abierta, donde las consecuciones de los alumnos no están tan definidas. Aquí ocurre que propuestas una serie de actividades con valor educativo en sí mismas, cada estudiante puede llegar a conseguir *metas diferentes* y a producir *realizaciones distintas*. La idea que subyace es la de educación como inducción al conocimiento. La evaluación aquí es también un juicio que realiza el profesor, pero no lo hace comparando el logro del estudiante con algo ya previsto y "perfecto" sino considerando la comprensión de la naturaleza de su materia y cualificando el trabajo de cada estudiante, que será distinto al de los demás.

Existe en la actualidad la tendencia a alejarse de los estándares de eficiencia externamente determinados de manera común para todos y se propone una evaluación basada en la *consecución de rendimientos mínimos determinados* tanto en el Proyecto del Centro como en la Programación de Aula, llamado *rendimiento suficiente*, pero sobre todo de manera idiosincrática, valorando el progreso de cada estudiante con respecto a su rendimiento previo o *rendimiento satisfactorio*, lo que constituye una referencia personalizada. Se tienen en cuenta las condiciones personales de cada estudiante y se tiene muy claro lo que se le puede y lo que se le debe exigir. Para juzgar al estudiante no se toma como marco de referencia o punto de comparación una norma preestablecida sino las posibilidades del propio sujeto.

Esta evaluación personalizada requiere saber previamente cual es el rendimiento adecuado a las condiciones personales de cada estudiante, o sea, el rendimiento *que se espera de él*. Si se atiende a la evaluación del rendimiento respecto a un conjunto (comparando el nivel logrado por un estudiante con el nivel establecido previamente) se puede dar un "rendimiento suficiente" (evaluación positiva) o con un "rendimiento insuficiente" (evaluación negativa). Si la evaluación se hace pensando en la persona que realiza el trabajo, se observa un "rendimiento satisfactorio" si es adecuado a la capacidad del estudiante o un "rendimiento insatisfactorio" si es inferior al previsto. Se presentan *cuatro situaciones* en las que puede hallarse un estudiante atendiendo a la relación entre evaluación y rendimiento:

- a. *Suficiente y satisfactorio*: no se observa problema alguno y se debe animar al estudiante a seguir desarrollando su trabajo con el mismo tesón.
- b. *Suficiente e insatisfactorio*: el rendimiento es aceptable pero debe alentarse al estudiante a trabajar más y mejor porque puede hacerlo.
- c. *Insuficiente y satisfactorio*: el estudiante responde a las expectativas creadas y merece elogio, pero debe esforzarse para llegar a alcanzar un rendimiento suficiente.
- d. *Insuficiente e insatisfactorio*: plantea la situación más delicada porque es altamente probable que el estudiante precise de algún tratamiento especial.

La evaluación *personalizada* plantea el problema de la *predicción del rendimiento*. Una evaluación satisfactoria o insatisfactoria requiere establecer el nivel que se estima que puede llegar a alcanzar un estudiante concreto. Hay que hacer el diagnóstico de cada estudiante para, desde él, formular el pronóstico o predicción de su rendimiento. La llamada "predicción extrapolativa técnica" es la más adecuada a la naturaleza de los fenómenos educativos y permite la

creación de tablas de predicción que son muy útiles. En función de la predicción se van a programar las actividades y su temporalización, de manera que cada estudiante pueda cumplir su plan de actividades.

BIBLIOGRAFÍA

- Brown, S. y Glasner, A. (2003). *Evaluación en la universidad*. Madrid: Narcea.
- Castillo, S. (2002). *Compromisos de la evaluación educativa*. Madrid: Prentice Hall.
- López, B. (2005). *Evaluación del aprendizaje*. México: Trillas.
- Lukas, J. F. y Santiago, K. (2004). *Evaluación educativa*. Madrid: Alianza Editorial.
- Mateo, J. (2000). *La evaluación educativa, su práctica y otras metáforas*. Barcelona: ICE-Horsori.
- Salinas, D. (2002). *¡Mañana examen! La evaluación: entre la teoría y la realidad*. Barcelona: Graó.
- Santibáñez, J. (2001). *Evaluación del aprendizaje estudiantil*. México: Trillas.
- Tenbrink, T. (2002). *Evaluación. Guía práctica para profesores*. Madrid: Narcea.

