La evaluación de los aprendizajes en el debate didáctico contemporáneo

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

PAIDÓS
Buenos Aires
Barcelona
México

la. Edición, 1998

ISBN 950-12-2129-6

SUMARIO

Presentación **9**

La evaluación: campo de controversias y paradojas o un nuevo lugar para la buena enseñanza **11**
Edith Litwin

¿Es posible mejorar la evaluación y transformarla en herramienta de conocimiento? **35**
Susana Celman

La calidad de los programas de evaluación y de los instrurnentos que los integran **67**
Alicia R. W. de Camilloni

La evaluación de las prácticas docentes y la autoevaluación **93**
María del Carmen Palou de Maté

Sistemas de calificación y regímenes de promoción **133**
Alicia R. W. de Camilloni

LA CALIDAD DE LOS PROGRAMAS DE EVALUACIÓN Y DE LOS INSTRUMENTOS QUE LOS INTEGRAN*

Alicia R. W. de Camilloni

INTRODUCCIÓN

La evaluación de los aprendizajes se realiza sobre la base de un programa que, al servicio de la enseñanza y del aprendizaje, está constituido por un conjunto de instrumentos de evaluación. Dado que cada tipo de instrumento permite evaluar diferentes aspectos de los aprendizajes de los alumnos, es menester garantizar la pertinencia y calidad técnica del programa considerado integralmente como una estructura, así como la de cada uno de sus componentes.

Un principio general que rige el diseño de un programa de evaluación es el de la economía de tiempo. E1 tiempo destinado a situaciones de evaluación puntual, diferenciadas de las situaciones de enseñanza, debe ser el más breve posible, ya que el factor más crítico de la educación formal es la carencia de tiempo suficiente para enseñar con profundidad toda la información y para desarrollar todas las competencias que los alumnos deben alcanzar.

Las condiciones en que se desenvuelve actualmente la educación escolar en todos los niveles implican, precisamente, la necesidad de sortear algunas dificultades para la evaluación porque las clases suelen ser numerosas como producto de los muy significativos y deseados procesos de democratización de la educación. E1 número de alumnos ha aumentado en los niveles de educación básica, secundaria y superior, planteando el requerimiento de encontrar las técnicas adecuadas para la enseñanza de grupos numerosos, así como las técnicas de evaluación que puedan acompañar adecuadamente a una enseñanza que promueva aprendizajes significativos en todos los alumnos, que evite el fracaso escolar, la repitencia y la deserción, y que permita hacer efectivo el propósito de igualar las oportunidades educativas reconociendo los derechos a la igualdad así como a la diversidad de los rasgos personales de los alumnos.

Por ello, la calidad de un programa de evaluación está sujeta, desde un punto de vista pedagógico y por esto, también, ético y político, a su capacidad para evaluar justa y equitativamente a grupos numerosos integrados por alumnos diferentes.

No es posible hablar de la evaluación de los aprendizajes al margen de los procesos de enseñanza y aprendizaje que los han generado. Una condición fundamental para el diseño de programas de evaluación, en cualquiera de los niveles que se deseen o se deban evaluar en la institución escolar, es que tienen que ser consistentes con los proyectos de enseñanza y aprendizaje de la institución ya que sólo alcanzan su propósito de servir como dispositivos para el perfeccionamiento de los resultados de la educación cuando se convierten en juicios de autoevaluación, tanto para los alumnos como para los docentes y las autoridades de la escuela y del sistema.

Para lograr una buena configuración de los procesos de enseñanza, el docente debe tomar algunas decisiones adecuadas. Las decisiones de los docentes están fundadas sobre las concepciones que ellos tienen, entre otras, acerca de qué es enseñar, qué es aprender, cuál es la naturaleza de los conocimientos que los alumnos deben adquirir y qué funciones tiene que cumplir la escuela. Hay, pues, un conjunto de teorías o ideas más o menos explícitas, o a veces, implícitas, basadas en el sentido común o en teorías científicas, según los casos, a través de las cuales cada docente procura construir explicaciones de los procesos y las funciones relacionadas con la intervención pedagógica. Esas explicaciones le sirven de fundamento cuando diseña su proyecto de enseñanza.

Los docentes también ponen en juego la información que poseen acerca de lo que está ocurriendo con sus alumnos y las características que suponen que éstos tienen como individuos y como grupo de clase, ya que ellos son los destinatarios de la enseñanza. En esta tarea, la referencia se centra en la necesidad de prever cuáles habrán de ser los impactos más duraderos y los efectos más inmediatos que posiblemente tengan sus propias acciones de enseñanza sobre los alumnos. Para efectuar esas previsiones, el docente debe poseer información de distintos tipos. Ella puede ser recogida por medio de la observación del desarrollo de los procesos de enseñanza o puede surgir de las respuestas de los alumnos colocados en situaciones creadas especialmente por los docentes con ese fin.

Si se parte de estas premisas, se advierte que la importancia del papel que desempeñan las teorías explicativas de cómo ocurren estos procesos, qué modalidades adoptan en su presentación y cuáles son las

* 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).

variables que inciden en ellas, obliga al docente a efectuar el análisis y la reconstrucción de dichas explicaciones teóricas, de modo de trabajar con aquellas a las que estima más apropiado adjudicar un mayor valor científico por su actualidad, seriedad y rigor.

Pero, además, en lo que se refiere a la información relativa a los individuos concretos o empíricos que conforman el grupo de clase o de la institución, el docente necesita contar, también, con la mejor información. Para obtenerla puede recogerla de las situaciones naturales tal como ocurren en la sala de clase o fuera de ella, o bien puede generar intencionalmente situaciones específicas. La evaluación tiene un estado inicial que puede construirse a partir de dos tipos de operación: una modalidad se constituye sobre la base de recoger datos a través de la observación sistemática o asistemática o de algún otro medio no diferenciado del proceso de enseñanza volcados en la elaboración de sus respectivos registros y, la otra, con características bien diferenciadas de la anterior, consiste en obtener datos en situaciones especialmente diseñadas para recoger una información que se considera relevante como indicadora del estado de situación de los aprendizajes que se quieren evaluar.

Además de esta operación inicial, en alguno de sus dos tipos básicos, para que haya una evaluación con sentido es indispensable que existan criterios que permitan construir juicios de valor acerca de lo que la información recogida significa en términos de aprendizaje de los alumnos. La primera operación, por lo tanto, consiste en recoger información; y la segunda, en analizar e interpretar la información recogida sin que, necesariamente, el orden enunciado implique la diferenciación temporal o el establecimiento de una secuencia de estas dos operaciones. La interpretación permite completar el proceso de evaluación y conduce a la construcción de juicios de valor que son característicos de la evaluación como función didáctica. No hay evaluación sin juicios de valor. Debemos mencionar aquí, sin embargo, que algunos autores discrepan con esta afirmación. Jean Cardinet (1989), por ejemplo, advierte, como veremos más adelante, acerca de los riesgos que implica el uso de juicios de valor en la evaluación.

E1 juicio de valor puede ser construido de dos maneras. Según una de ellas, puede responder a estados afectivos del docente, relacionados con actitudes de aceptación o rechazo, de agrado o desagrado ante determinada conducta del alumno, observada directamente, inferida de conductas observadas o, también, sólo supuesta más o menos indirectamente a partir de la interpretación de variadas situaciones. El juicio de valor puede surgir, igualmente, como producto de una elaboración seria y rigurosa de la información recogida sistemáticamente, a partir ya no de emociones, sentimientos o actitudes, sino de una base de conocimientos que permite fundamentar el juicio de valor de una manera que pretende ser "objetiva". Ante estas dos maneras diferentes de construcción es posible que, por la naturaleza misma de los actos pedagógicos, en el ejercicio integral del papel del docente se requiera poner en juego las dos modalidades de construcción de los juicios de valor, buscando el modo de articularlas adecuadamente, entendiendo que el docente es una persona y que la relación educativa es una relación entre personas.

La evaluación tiene que contar con criterios que permitan analizar e interpretar la información recogida con base en una teoría orientada a responder a la serie de cuestiones que antes mencionamos. Entre ellas incluimos la teoría referida a la evaluación como función didáctica. Si el docente tiene que contar con teorías explicativas acerca de cómo aprenden los alumnos y sobre qué es enseñar, también deberá trabajar con teorías acerca de la evaluación, su relación con los procesos de enseñanza y aprendizaje, el lugar que ocupa la evaluación en la administración de un proyecto de enseñanza, qué alcances y qué límites debe tener.

Es siempre importante insistir en la necesidad de que todas esas teorías sean consistentes entre sí y conformen, en su conjunto, una base de percepción, pensamiento, sentimiento y conocimiento que permita fundar y justificar la acción docente.

Estos principios no suponen, sin embargo, que exista una única concepción científica acerca de las cuestiones que hemos mencionado, incluyendo entre ellas a la propia evaluación. Así como hay diversas concepciones acerca del universo, las hay también sobre la enseñanza y el aprendizaje.

Cuando se adopta una concepción acerca de la enseñanza, de lo que la enseñanza debe ser y no sólo acerca de lo que ella es, esta concepción normativa conduce a una idea normativa de la evaluación, de lo que la evaluación debe ser. Así pues, para una teoría normativa de la enseñanza, no todas las concepciones sobre la evaluación son válidas; algunas son aceptables, otras no lo son.

La evaluación es un recurso indispensable para el perfeccionamiento de los dos procesos que nos interesan, la enseñanza y el aprendizaje. Encarar una acción docente significa actuar para mejorar las probabilidades de que se produzcan procesos deseables de aprendizaje. Pero es menester señalar que, aunque por lo general cuando hay enseñanza se produce algún aprendizaje, éste no siempre coincide con los propósitos del docente. La actitud del docente frente a la evaluación deberá ser abierta, coincidiendo con su

teoría de la enseñanza. Si se está de acuerdo con la idea de que, al enseñar, el docente no debe desarrollar una intervención caracterizada por su unidireccionalidad en la que la única voz a escuchar es la del propio docente sino que hay que dar lugar a la voz del alumno, esto es, a la manifestación de su capacidad para pensar y construir significados, del mismo modo en el proceso de evaluación debe encontrar el alumno un lugar para expresar los significados desde su propia perspectiva.

Cuando se opta, entonces, por concepciones de la enseñanza en las que se manifiesta nuestro respeto por la capacidad de producción personal del alumno, los instrumentos de evaluación que utilicemos deberán ser coherentes con esa postura.

En la actualidad disponemos de una cantidad importante de instrumentos diversos, creados muchos de ellos a lo largo del siglo XX. Consisten, por un lado, en las llamadas "pruebas objetivas", diseñadas a partir del programa de investigación científica del conductismo, con una serie de especificaciones técnicas importantes en cuanto a la construcción de ítemes y al análisis e interpretación de resultados. También contamos con un conjunto de normas técnicas para la utilización de las denominadas "pruebas subjetivas"- a veces llamadas "tradicionales"- en las que, a diferencia de las pruebas objetivas que sólo admiten una única respuesta correcta, se aceptan diferentes niveles de calidad de respuesta. En este último campo, las normas técnicas desarrolladas han apuntado a mejorar la construcción, la administración y la interpretación de los resultados, con el propósito de mejorar también la confiabilidad y la validez de las pruebas subjetivas.

En este siglo, el trabajo realizado en el campo de la evaluación didáctica ha sido muy profuso, ya que se ha considerado que se trataba de un tema prioritario en las teorías de la enseñanza. No es extraño, por ejemplo, que para las diversas teorías conductistas del aprendizaje, la evaluación se convirtiera en un aspecto central de su desarrollo teórico y experimental y de su propuesta didáctica, hasta el punto de sostener que se debía limitar lo que se enseñaba a aquello que era posible evaluar en los términos que la teoría aceptaba con carácter exclusivo, es decir, cuantificando los rasgos de las conductas observables. Dada la significativa importancia que se otorgaba a la evaluación de los aprendizajes y a la simplicidad técnica del problema planteado, se comprende la gran fertilidad de las teorías conductistas en el dominio teórico y práctico de la evaluación.

Consideramos que se trata de una función fácil de resolver técnicamente en el marco de las teorías del conductismo porque se limita a evaluar lo que se puede observar. Las corrientes cognitivas, en cambio, se proponen evaluar los procesos mentales. Su problema es más difícil de resolver que el que se planteaban los conductistas y sus respuestas técnicas son más dificultosas de construir. La cuestión de la evaluación, en cambio, es menos importante para las didácticas humanistas que plantean una relación diferente entre docente y alumno en la que los aprendizajes son aún más difíciles de evaluar que en las propuestas didácticas cognitivas.

Debemos señalar, empero, que el programa conductista caracterizado por su gran capacidad de producción de instrumentos técnicos de evaluación durante varias décadas del siglo XX no fue el único programa que realizó trabajos en este tema; no fue ni siquiera el que ejerció una hegemonía completa en la evaluación. En el campo teórico y práctico, simultáneamente con el trabajo que en este ámbito fuera realizado por los psicólogos conductistas, también se trabajó con otros instrumentos de evaluación que respondían a perspectivas diferentes. Algunas corrientes europeas que desarrollaron trabajos importantes en este terreno lo hicieron desde enfoques diferentes, alternativos a las producciones conductistas. Entre ellos, por ejemplo, se destacan los que entregan al alumno la responsabilidad de seleccionar el material, organizarlo y encontrar formas adecuadas de expresar su producción personal, cuestiones que no pueden ser evaluadas por medio de la administración de pruebas objetivas. De este modo, y aún en el ámbito de mayor influencia de las teorías conductistas, se emplearon siempre y se enseñó a los docentes a construir, analizar e interpretar los resultados de las llamadas "pruebas subjetivas" o "tradicionales".

Es importante hacer notar que la gran disponibilidad de instrumentos de diferente carácter, alcance y función, es un factor de enriquecimiento del abanico de posibilidades que se abren para el diseño de programas de evaluación. E1 problema que se ha de resolver es el de su adecuada selección y organización de modo de alcanzar consistencia con el proyecto docente.

Uno de los productos más importantes del trabajo de los diversos programas de investigación sobre la evaluación en este siglo ha sido, precisamente, el mejoramiento de la normativa de construcción de instrumentos y análisis de resultados, lo cual nos permite abordar el problema de la evaluación contando con una amplia variedad de instrumentos que incluyen pruebas escritas y orales y pruebas de observación de procesos y de producciones, utilizables de diversas maneras según las necesidades de la enseñanza. E1 aporte del conocimiento desarrollado en las últimas décadas sobre los procesos cognitivos da un nuevo sustento a las teorías y técnicas evaluativas y permite confirmar o modificar las prácticas instaladas en la

educación formal desde hace largo tiempo en algunos casos, y en otros recientemente, como resultado de la introducción de nuevas propuestas que es necesario reanalizar a la luz de la investigación actual.

LOS PROGRAMAS DE EVALUACIÓN

Si nos preocupamos hoy por determinar cuáles son los ejes en torno de los cuales gira esta cuestión, tendríamos que comprender, en primer lugar, que el propósito principal de la evaluación no se puede lograr si la evaluación no se convierte en autoevaluación tanto para el docente cuanto para el alumno. La evaluación, en segundo lugar, debe ser consistente con las concepciones de la enseñanza y del aprendizaje. En tercer lugar, la resolución técnica que se le dé debe permitir evaluar todos los aspectos que están comprometidos con los procesos de aprendizaje, lo cual supone desarrollar programas complejos de evaluación en los que se empleen una cantidad de instrumentos diversos y donde cada técnica sea adecuada para evaluar diferentes aspectos.

Los programas de evaluación deben ser diseñados por los docentes, como parte de la programación didáctica de sus cursos. Para ello, el docente debe conocer en profundidad la o las teorías de la evaluación y la variedad de instrumentos que existen. Estos conocimientos le permitirán diseñar programas coherentes con la programación de la enseñanza combinando instrumentos diversos, así como crear nuevos instrumentos de evaluación acordes con la multiplicidad y heterogeneidad de los propósitos de su enseñanza. Es también indispensable, si quiere desarrollar un buen programa de evaluación, que el docente conozca las normas técnicas para la construcción, administración, análisis e interpretación de resultados.

Queremos señalar que, en la medida en que un docente es responsable de la elaboración de su programación didáctica, deberá tener la misma libertad para diseñar y administrar su programa de evaluación. Los grados de libertad en la toma de decisiones deben ser equivalentes. Si así no ocurriera y si el docente tuviera libertad para decidir algunos aspectos de su proyecto de enseñanza y no la tuviera para decidir acerca de su programa de evaluación porque ésta le es impuesta desde el exterior, la evaluación no cumpliría otra función que la de servir de control, resultando minimizados sus posibles efectos positivos de mejoramiento de los procesos y de los resultados de esos procesos de enseñanza y de aprendizaje. Hay que recordar siempre que la libertad de quien enseña debe tener su correspondencia en la libertad de quien evalúa y en la libertad de quien aprende. Y, respecto de este último es importante recordar, como lo hace E. R. House (1994) citando a John Rawls, que ". . . respetar al otro como persona moral supone tratar de comprender sus aspiraciones e intereses desde su punto de vista y presentarle consideraciones que le hagan posible aceptar las limitaciones de su conducta" (pág. 129).

LOS DETERMINANTES DE LA CALIDAD DE LA EVALUACIÓN

Un programa de evaluación y cada uno de los instrumentos de evaluación que lo integran deben cumplir con ciertas condiciones. Mencionaremos aquí algunas de ellas: debe permitir evaluar los aprendizajes que la institución y el o los docentes esperan que los alumnos desarrollen. Las consignas de trabajo que integran el instrumento deben ser seleccionadas, en consecuencia, de modo de permitir que tanto la presencia de una respuesta como su ausencia sean significativas en relación con los aprendizajes que se ha buscado promover. En términos generales, las consignas no deben producir respuestas que resulten irrelevantes para la evaluación de los aprendizajes previstos. Esto no significa, sin embargo, que se descarten las respuestas irrelevantes de los alumnos cuando aparecen, ya que a través de esos resultados en apariencia disfuncionales, a veces incomprensibles, el docente puede advertir errores que sí pueden ser altamente significativos y que requieren especial atención. El instrumento, aunque siempre haga una lectura parcial de lo aprendido por los alumnos, debe proponerse registrar no sólo los éxitos y fracasos sino también el origen de esos fracasos. E1 instrumento debe presentar el grado de organización suficiente para que la apreciación que efectúa del aprendizaje permita desprender algunas conclusiones acerca del desempeño presente y futuro del alumno, en cuestiones específicas pero también con visión integral.

Cada uno de los instrumentos de evaluación resuelve estos problemas de diversas maneras. Actúa como un reflector que en el escenario ilumina a algunos personajes y deja en la penumbra a otros, que sin embargo están allí, presentes, pero con una clase diferente de presencia. Por esta razón, la elección de los instrumentos de evaluación adecuados a la hora de diseñar el programa de evaluación de un curso, por ejemplo, constituye una de las decisiones más importantes para garantizar el valor didáctico de la evaluación en relación con los procesos de enseñanza y de aprendizaje.

Existe una variedad de modelos de instrumentos de evaluación. Cada uno de ellos, tomado aisladamente, es insuficiente para obtener información sobre el aprendizaje de los alumnos. E1 diseño del programa exige, pues, la combinación de instrumentos diversos para obtener una cobertura adecuada. La eficacia de la evaluación depende, entonces, de la pertinencia de la combinación de diferentes instrumentos, de la oportunidad en que se administran y de la inteligencia y propiedad del análisis e interpretación de sus resultados.

Veamos, entonces, los criterios o requisitos que deben tenerse en cuenta en el momento de tomar decisiones acerca del programa de evaluación de un curso, por ejemplo, y sobre cada uno de los instrumentos de evaluación que lo integran. Las características generales que los instrumentos de evaluación deben reunir, aunque con distinto grado de relevancia, son cuatro: validez, confiabilidad, practicidad y utilidad.

VALIDEZ

Se dice que un instrumento de evaluación es válido cuando evalúa lo que se pretende evaluar con él. Como un instrumento es utilizado para apreciar ciertos logros de aprendizaje de un cierto grupo de alumnos en una cierta circunstancia y en un cierto momento de su proceso de aprendizaje (al inicio, en el transcurso o al final), la validez de un instrumento no puede ser determinada de manera absoluta, sino siempre en relación con su adecuación a los propósitos y situación específica de su aplicación. Como lo señala Henry E. Garrett (1966), "un test es válido para un propósito particular o en una situación especial; no es universalmente válido". Cuando se quiere determinar si un instrumento es válido se requiere, entonces, información acerca de los criterios que han presidido su construcción y su administración. Los criterios son, entonces, externos a la evaluación misma. Esta cuestión sería relativamente simple de resolver si todos los propósitos de la enseñanza, en todos los niveles de decisión y para todos los actores, fueran coincidentes. Pero rara vez ocurre esto. A veces los objetivos del docente coinciden plenamente con los de la escuela y los de ésta con los de la jurisdicción y los de esta última con los de la nación. Pero la plena armonía de todos esos niveles parece más improbable que posible de alcanzar y, más aún, ni siquiera se puede afirmar que, en lo pormenorizado de los aprendizajes escolares, sea absoluta y efectivamente deseable. Las diferencias en la educación surgen de la acción y la resolución de cada docente y de cada alumno. Se podrán encontrar, en consecuencia, criterios generales, pero también criterios diferentes. Y no tiene sentido hablar de la validez de un instrumento si no se ha validado el criterio. Para lo cual se requiere, a su vez, una validación mediante criterios más amplios. Según Robert L. Ebel (1963), "esto requiere un criterio para validar el otro criterio y así sucesivamente *ad infinitum* .

En la Didáctica, ese proceso de validación al infinito podría ser reemplazado por una teoría de bases firmes, tanto desde un punto de vista científico como filosófico. De otra manera, y ante la obligación ineludible de evaluar a sus alumnos, los docentes se encontrarían en la situación de ser incorporados como nuevos personajes, junto a Aquiles, la tortuga y la flecha, de las aporías de Zenón de Elea.

No se trata, en consecuencia, de validar instrumentos y criterios como en un juego de permanente retroceso hacia los principios, sino de apoyarse en principios didácticos fuertes, con cierto grado de generalidad y que permitan fundamentar racionalmente las decisiones de diseño del programa y de los instrumentos. Veámoslo en algunos ejemplos.

Si para un docente el propósito principal de su enseñanza ha sido procurar que sus alumnos aprendan a resolver un cierto tipo de problemas, el instrumento de evaluación válido será aquel que introduzca problemas al alumno para que éste los resuelva. Deberá tener presente, sin embargo, que no obtendrá la misma información acerca de lo que sus alumnos saben cuando en el instrumento se plantean problemas que corresponden sólo al tipo de problema cuya solución los alumnos deben hallar, que si también incluyen otros problemas de tipos diferentes, de modo que los estudiantes deban reconocer a qué tipo de problema corresponde cada uno de

los planteados por el docente. Se podría afirmar que, en tanto el propósito didáctico no habría de ser que el alumno resuelva sólo una clase de problemas, la capacidad de discriminar clases de problemas debería estar presente en la formulación de los propósitos de un programa de enseñanza, por lo que esa segunda alternativa, que incluye diversos tipos de problemas en el instrumento, es más válida que la primera, que presenta solamente un único tipo de problema. Así, pues, para lograr que la construcción del programa o de un instrumento de evaluación sean válidos, es menester poner en juego no sólo una lectura fragmentada de los propósitos de la enseñanza sino principios didácticos más generales.

En otro sentido y refiriéndonos a otro ejemplo, veamos el papel que tiene el tipo de instrumento: si se trata de construir un examen oral o un coloquio válidos, hay que definir previamente preguntas centrales. Éstas deben referirse a cuestiones fundamentales del dominio de conocimiento que el alumno debe aprender. Pero la validez no llega a estar asegurada de este modo ya que en estos tipos de instrumentos de evaluación, el examen oral y el coloquio, la característica que les es propia y que hay que desarrollar adecuadamente es la elaboración de preguntas o comentarios del docente a partir de las respuestas del alumno. Si ellas no le permiten al alumno corregir sus respuestas anteriores, por ejemplo, o no lo animan a profundizar respuestas que inicialmente eran superficiales, o si no lo inducen a la elaboración de asociaciones y conclusiones reuniendo respuestas fragmentarias, la técnica del examen oral o del coloquio están mal empleadas y podrían ser reemplazadas por un instrumento de respuesta escrita. Como se ve, la validez de los instrumentos también depende de su apropiada construcción en relación con las características específicas que los definen.

Se puede observar, a través de lo expuesto hasta aquí, que la validez nunca es absoluta. En primer lugar, porque siempre debe ser referida a un criterio externo. En segundo lugar, porque sólo se puede hablar de validez mayor o menor, ya que un instrumento nunca es "completamente" válido pues sólo permite hacer inferencias acerca de los saberes de los alumnos. Y finalmente, porque la validez está condicionada a un ajuste que nunca puede ser perfecto entre el saber aprendido que se quiere evaluar y los medios técnicos, esto es, los instrumentos disponibles para evaluarlo. Con frecuencia se obtiene mayor validez a través de la combinación feliz de distintos tipos de instrumentos.

Se diferencian diversas clases de validez. Desarrollaremos sintéticamente sus principales modalidades.

Validez de contenido

Se dice que un instrumento tiene validez de contenido cuando representa una muestra significativa del universo de contenido cubierto por el curso o la unidad didáctica, la clase o una serie de clases. La validez de contenido, por lo tanto, requiere que la selección de contenidos que se ha efectuado para construir el instrumento sea equivalente a la de cualquier otro conjunto de contenidos que se hubiera podido seleccionar con ese fin. Se trata, por supuesto, en todos los casos de contenidos importantes por su significación en el aprendizaje de los alumnos.

La dificultad que la construcción de un instrumento debe superar es la necesaria distancia que existe entre la gran cantidad y variedad de los contenidos de toda enseñanza y la brevedad exigida también para los tiempos destinados a una evaluación de los aprendizajes, si no se quiere fatigar a los alumnos, lo cual conduciría a desvirtuar el significado cognitivo de sus respuestas.

Esa distancia obliga al docente a elegir los temas, informaciones, competencias, habilidades, destrezas, procedimientos u otros que incluirá en el instrumento de evaluación. Cuanto mayor es la diferencia entre tiempos destinados a la enseñanza y al estudio y tiempos destinados a la evaluación, más difícil es lograr una validez de contenido aceptable, es decir, que permita asegurar que los resultados obtenidos por el alumno en una instancia de evaluación serían semejantes a los que obtendría en otra distinta. La justicia de la evaluación se pone en juego, entonces, cuando se efectúa el muestreo de los contenidos que se incluirán en el instrumento a través de las consignas de trabajo (por ejemplo, las preguntas) que éste presenta.

La validez de contenido es también llamada "validez curricular" porque el criterio externo que se emplea para garantizar esta clase de validez es el diseño curricular, ya que en éste se expresan los propósitos y los niveles de aprendizaje, las secuencias y las sugerencias sobre estrategias e incluso, en ocasiones, sobre actividades de aprendizaje y modalidades de evaluación. Con la libertad que debe caracterizar al manejo del diseño curricular por parte del docente, es indudable que este diseño es un recurso fundamental sobre el que se apoyará la estimación acerca de la validez de contenido de un instrumento de evaluación, así como la fuente en la que el docente encontrará la principal inspiración para realizar una selección adecuada y representativa de los contenidos de su curso. Es preciso recordar que cuando se habla de "contenido" se hace referencia a un concepto amplio que comprende todo lo que se quiere enseñar. Se incluyen, por tanto, informaciones, conceptos, principios, procedimientos, valores, actitudes, hábitos, destrezas motrices y todo aquello que se tiene el propósito de que los alumnos aprendan.

En virtud de que lo que se enseña varía de clase en clase, de docente a docente y de región a región, la validez de los instrumentos también es limitada en relación con sus alcances.

Validez predictiva

Esta clase de validez se refiere a la correlación existente entre los resultados obtenidos en una o varias pruebas combinadas y el desempeño posterior del alumno en aquellos aspectos que corresponden al área evaluada por esas pruebas. Los instrumentos de evaluación que se emplean en la escuela deben tener validez predictiva escolar y extraescolar y deben contribuir a pronosticar los éxitos y dificultades de los sujetos del aprendizaje tanto en los aprendizajes posteriores de las asignaturas escolares cuanto en sus desempeños ocupacionales o en su actuación social en general.

Un nivel significativo en la validez predictiva es una condición para un buen instrumento de evaluación. Y esto es así por cuanto no hay aprendizaje escolar cuyo único fin consista en procurar una única ejecución exitosa en un espacio escolar y un tiempo muy limitados. E1 mismo fin de la educación se define por la voluntad expresa de quien enseña de que los aprendizajes trasciendan el espacio y el tiempo escolares. Es el destino mismo de la educación buscar la perdurabilidad de los aprendizajes. De ahí que, aunque haya algunos instrumentos cuyo propósito específico es la predicción -como es el caso de las pruebas de admisión-, todo instrumento de evaluación empleado con fines pedagógicos tenga, entre otras, la función de predecir. Como antes dijimos, la validez de un instrumento no es absoluta, siempre es limitada. Pero el seguimiento de los alumnos para comprobar la capacidad del instrumento para efectuar predicciones adecuadas es una operación de medición relativamente sencilla, al alcance de todos los docentes que deben encararla como parte constituyente de sus programas de evaluación. Comparar los resultados obtenidos por cada alumno en pruebas diferentes, separadas en el tiempo y en las que el docente asume que hay relación entre los contenidos, es una tarea interesante a la hora de evaluar la enseñanza impartida. Si en una prueba se han detectado errores en las repuestas y se ha brindado atención especial a los alumnos que tenían esas dificultades específicas, el seguimiento de los resultados en una prueba posterior demostrará, a la vez, la validez de la prueba anterior y la adecuación del tratamiento didáctico de los errores que se ha encarado en la acción de enseñanza.

Debemos señalar, sin embargo -aunque esto no exima al docente de hacer el esfuerzo de evaluar la validez predictiva de los instrumentos que utiliza- que los aprendizajes que la escuela promueve tienen plazos más o menos prolongados para su logro pleno. Se han de diferenciar aquellos logros que el alumno puede demostrar en un plazo muy corto, casi contemporáneo con el período de su enseñanza en la escuela, con los aprendizajes que requieren un período algo más prolongado de procesamiento, trabajo y elaboración para una apropiación profunda y aquellos otros aprendizajes que, por el alcance temporal que tiene la maduración personal que requieren, no podrían ser evaluados durante la permanencia del alumno en la institución educativa y ni siquiera en un lapso próximo a su egreso. En estos últimos casos la apreciación de la validez predictiva de los instrumentos se ve limitada por razones técnicas difíciles de superar por el docente y por la institución. Queda en pie, empero, la necesidad de poner atención constante en la capacidad de predicción que tienen los instrumentos de evaluación que se emplean en la educación formal.

Validez de construcción

E1 programa de evaluación y cada uno de los instrumentos que lo integran deben estar construidos de acuerdo con los principios de la o las teorías didácticas que sostienen el proyecto pedagógico. Debe haber coherencia entre las teorías y las estrategias de enseñanza que se han de emplear o que se han utilizado, las teorías generales y específicas que describen y explican los procesos de aprendizaje que el programa de enseñanza se propone que los alumnos desarrollen y las teorías que fundamentan las elecciones de los tipos de instrumento, la secuencia empleada, la articulación de la enseñanza y de las instancias de evaluación que conforman el programa de evaluación diseñado. Lo mismo ocurre con las normas técnicas que deben respetarse cuando se construye cada tipo de instrumento de evaluación. De la coherencia de las teorías y el buen uso de las normas técnicas depende la validez de construcción del programa y de los instrumentos de e valuación.

Es éste un aspecto que debe ser particularmente cuidado por los docentes porque si la evaluación no tiene un fundamento teórico válido, las decisiones que se tomen a partir de la información recogida por esa vía y de las interpretaciones que se hagan de ella carecerán de legitimidad.

Puede comprenderse, por lo tanto, la importancia de esta clase de validez ya que ella condiciona las otras que antes hemos mencionado.

Validez de convergencia

Esta modalidad de la validez concierne a la relación que existe entre un programa de evaluación o un instrumento y otros programas o instrumentos de validez ya conocida. Cuando un programa o una prueba de evaluación son reemplazados por otros, por ejemplo, debe demostrarse que son, por lo menos, tan buenos o mejores que los empleados anteriormente. La validez de convergencia puede tener gran importancia, pues, cuando se modifican prácticas establecidas en la evaluación para mejorar su validez, su confiabilidad o su practicidad, o cuando se busca mejorarla ampliando el campo de los aspectos evaluados.

Tomemos, por ejemplo, una situación en la que es necesario encontrar una nueva modalidad de evaluación que permita trabajar adecuadamente con un número muy superior de alumnos al que se acostumbraba atender en la institución. En ese caso, se trata de encontrar un programa e instrumentos más prácticos pero igualmente válidos que los que se empleaban con anterioridad. Los resultados obtenidos con los nuevos programas deberán ser contrastados, en consecuencia, con los obtenidos por medio de la aplicación de los usados antes. La validez de convergencia se apreciará analizando las coincidencias entre ambas series de resultados de la evaluación. Pero si, en cambio, se procura obtener mayor información acerca de un aspecto significativo del aprendizaje, profundizando el examen de los logros obtenidos por los alumnos en competencias o ejecuciones específicas que hasta ese momento se evaluaban de manera más general, en situaciones diferentes o en tipos de expresión distintos, la elección de los nuevos instrumentos deberá ser sometida a un análisis de convergencia para establecer si los resultados de unas y otras series son coincidentes o contradictorios. Si ocurriera esto último, habría que reexaminar tanto los instrumentos empleados previamente como los nuevos, antes de adoptarlos y dar estabilidad al nuevo programa de evaluación.

Validez manifiesta

Esta modalidad de la validez no se refiere a la validez intrínseca del programa o del conjunto de las pruebas o de cada prueba, sino al modo en que ellos se presentan, aparecen, ante el público.

Su significatividad deriva de que los programas y los instrumentos de evaluación deben ser percibidos por los alumnos, sus padres y el público como instancias adecuadas para la evaluación. No pueden ser vistos como absurdas carreras de obstáculos ni como simples o complejos filtros que tienen el único fin de producir una selección, más o menos azarosa y más o menos injusta, entre los alumnos. La razonabilidad de los instrumentos debe ser visible y explicable. Los instrumentos deben ser vistos como formas capaces de dar cuenta de los aprendizajes necesarios de los alumnos, tanto en lo que respecta a la forma como a los contenidos evaluados.

La validez manifiesta exige, entonces, que los instrumentos evalúen de manera verosímil los aprendizajes. Sin esto la evaluación pierde transparencia y no puede convertirse en autoevaluación. Al perderse la confianza en ella, pierde de este modo su sentido pedagógico y se requiere efectuar las correcciones y operaciones que resulten necesarias para lograr restablecerlo.

Validez de significado

Dietel, Herman y Knuth (1991) añaden a los distintos tipos de validez que se describen habitualmente, la validez de significado, que se mide en respuesta a la siguiente pregunta: ¿las tareas de evaluación tienen significado para los estudiantes y los motivan para alcanzar su mejor rendimiento?

Semejante a la modalidad de validez manifiesta que hemos descrito, la validez de significado se centra, en particular, en la relación que se establece desde la perspectiva de los estudiantes, entre el programa y los instrumentos de evaluación, por un lado, y los procesos de enseñanza y de aprendizaje, por el otro. La evaluación tiene diversos efectos sobre el aprendizaje. Entre los efectos deseados se encuentra el de proveer información al alumno sobre los progresos y los obstáculos que encuentra en el proceso de su aprendizaje y sobre los errores que debe superar, dándole indicios claros acerca de cuál puede ser el origen de sus dificultades. Pero la evaluación también tiene el efecto de incentivar la motivación del alumno por aprender. La superación de dificultades constituye un aliciente para el aprendizaje, así como la reiteración de los fracasos disminuye el interés, si se perciben las dificultades como insuperables. En ese espacio de dimensiones variables para cada alumno, y que se establece entre los problemas demasiado difíciles para ser resueltos y los problemas demasiado fáciles en los que no se puede fallar, se encuentra el punto adecuado en el que debe ubicarse la dificultad que presenta la prueba de evaluación para que tenga significado para el estudiante y constituya un desafío en el que éste ponga su mejor esfuerzo de producción.

Validez de retroacción

La evaluación tiene un efecto de retroacción sobre la enseñanza. Un caso clásico es el que se presenta en las evaluaciones externas, que ejercen una influencia normativa sobre lo que se enseña. Conocidos los contenidos y las técnicas empleadas en las evaluaciones, ellas se convierten para los docentes en los modelos de lo que se debe enseñar y aprender. Lo mismo ocurre, con frecuencia, con los aprendizajes encarados por los alumnos, que se conforman a las pautas establecidas por los docentes en las evaluaciones que administran. Es un efecto, en este caso no deseado, de la evaluación, por cuanto, en vez de ocupar el lugar de un apoyo para el mejoramiento del aprendizaje y de la enseñanza, la evaluación se convierte en la verdadera reguladora de esos procesos. Cuando esto ocurre, el peligro es grande, ya que, como dijimos antes, el dominio de los aprendizajes en la educación es mucho más amplio que el que se puede cubrir mediante la evaluación.

La validez de retroacción, que debe ser medida al igual que los otros tipos de validez, impone la realización de acciones de corrección con el fin de restablecer el verdadero sentido didáctico de los procesos de evaluación.

CONFIABILIDAD

Se dice que un instrumento de evaluación es confiable cuando une exactitud en la medición y sensibilidad para la apreciación de la presencia y las diferencias de magnitud de los rasgos que mide.

Para que los resultados obtenidos mediante la administración de un instrumento de evaluación puedan ser considerados dignos de confianza, deben ser estables, permaneciendo semejantes en todas las ocasiones en que se administre ese instrumento u otro similar. Si, por ejemplo, se administra una prueba o se emplea algún otro tipo de instrumento de evaluación y se obtienen ciertos resultados, el instrumento es confiable si, unas semanas después, administrado nuevamente se obtienen resultados similares. Esa constancia de los resultados demostraría, primero, que el papel del azar es muy pequeño y no distorsiona de modo significativo los resultados obtenidos. Y, además, que la administración del instrumento ha podido dejar de lado la influencia de factores transitorios que no deberían tener relevancia en la consideración de los resultados del aprendizaje que se quieren evaluar.

Un instrumento confiable permite aislar los aspectos que mide de otros que para el caso se consideren irrelevantes. Así, por ejemplo, una lámpara colgante puede servir para detectar un sismo cercano, pero no puede ser considerada un instrumento confiable pues, al no encontrarse aislada de otros factores, el viento o una mano, por ejemplo, podría ser accionada por otras fuerzas que falsearían la interpretación de su movimiento pendular. De la misma manera, un instrumento utilizado en la escuela con fines de evaluación de los aprendizajes alcanzados por los alumnos debe permitir la consideración de estos aprendizajes separándolos de otros factores que pueden incidir en las respuestas de los alumnos, tales como el azar, la fatiga o los estados de tensión nerviosa de los alumnos o de los docentes.

La confiabilidad depende, también, de la exactitud y precisión con que mide el instrumento. En primer lugar, debe detectar si el aprendizaje o el rasgo que es evaluado está presente o ausente (¿saben o no saben *x*?, ¿tienen *y* destreza o no la tienen?, ¿quieren *z* o no la quieren?, etcétera). Pero debe hacerlo con sensibilidad y sin ambigüedad ante las variaciones de intensidad, profundidad y calidad de las conductas observadas o inferidas.

La confiabilidad de un instrumento depende, pues, de su estabilidad, su exactitud y su sensibilidad. Agregamos a estas condiciones la objetividad, que supone que la evaluación es independiente de la persona que evalúa, esto es, que los resultados reciben una interpretación similar tanto de diferentes docentes como de un mismo docente en distintos momentos.

Entre los factores que afectan la confiabilidad se encuentran los siguientes: la longitud y duración de la prueba, ya que cuando es excesivamente breve su capacidad de cobertura de contenido es escasa y puede estar muy distorsionada por factores de azar. "La muestra limitada de conducta puede dar como resultado una puntuación inestable y poco confiable" (Thorndike y Hagen, 1989). Pero cuando la prueba es excesivamente larga, la duración prolongada puede proporcionar información afectada por la fatiga del alumno. Sin embargo, señala Gronlund (1973) que es preferible el riesgo de diseñar pruebas más largas para que tengan mayor validez de contenido, siempre que se añadan contenidos diferentes y que no se superpongan con los que ya presenta la prueba. Cuando no se trata de instrumentos para medir velocidad sino de pruebas de competencia o capacidad, el tiempo debe ser suficiente para la realización del trabajo, incluso para estudiantes lentos.

Otro factor que incide sobre la confiabilidad son las condiciones de administración del instrumento. En situaciones puntuales de evaluación formal, la tensión de los alumnos suele ser mayor que en las informales, y la manera en que son tratados por el examinador puede favorecer o perjudicar su desempeño. También tienen influencia sobre los rendimientos las condiciones materiales y físicas de administración de las evaluaciones tales como iluminación, silencio, mobiliario, equipamiento adecuado y comodidad en general de alumnos y docentes.

Alcances de la confiabilidad

La confiabilidad de un instrumento debe permitir cubrir apropiadamente las necesidades planteadas por el programa de evaluación. No todos los instrumentos tienen la misma exactitud y sensibilidad, ya que ella depende de la forma técnica que adoptan y de las características propias que les otorga su construcción específica. Así, en otro orden de mediciones, por ejemplo, un barómetro doméstico puede ser aceptablemente confiable porque las lecturas que permite efectuar brindan la información que, con márgenes de error adecuados a ese uso doméstico, le permiten cumplir con la función a la que está destinado. En una estación meteorológica, en cambio, esas mediciones serían consideradas insuficientemente confiables, pues la sensibilidad y exactitud requeridas para cumplir su función son mucho mayores. Los instrumentos de medición de precisión son los que tienen mayor estabilidad, exactitud, sensibilidad y objetividad. Pero se debe recordar siempre que todas las mediciones no son absolutas, sino que son relativas a los instrumentos de medición que se emplean.

En cada caso, en consecuencia, habrá de determinarse cuáles son los grados de confiabilidad que resultan necesarios para cumplir con los propósitos que el programa de evaluación asigna a cada instrumento.

No se puede confundir confiabilidad con validez. Es frecuente que los autores de libros sobre este tema afirmen que la confiabilidad es condición necesaria, aunque no suficiente, de la validez (Karmel, 1986; Gronlund, 1973; Popham, 1983; Ebel, 1963; Sachs Adams, 1970; Pidgeon y Yates, 1979). Veamos algunas de estas afirmaciones.

Popham (1983): "Un test carente de (con)fiabilidad no puede ser válido para fin alguno". Ebel (1963): "Para que una prueba sea válida debe ser confiable". Pidgeon y Yates (1979): "Un examen no puede ser válido a menos que sea (con)fiable; pero puede ser (con)fiable sin ser válido". Karmel (1986): "Recuerde que ningún test será válido a menos que sea confiable. Por otra parte, un test puede ser confiable y carecer de validez". Esta exigencia de confiabilidad como condición para hacer posible la validez ha tenido mucha influencia en la teoría y la práctica de la evaluación. Consideramos, sin embargo, que es un principio que debe ser examinado y modificado.

La validez y la confiabilidad de un instrumento de evaluación nunca son absolutas. Hemos visto que para la validez los criterios son los que sirven de pauta y, del mismo modo, que la confiabilidad presenta márgenes de aceptabilidad según el uso que se le quiera dar al instrumento.

Pero por otra parte, en los instrumentos de evaluación de los aprendizajes, cuando se exige confiabilidad como primera condición para lograr cumplir con la exigencia de validez del instrumento, se privilegia un cierto tipo de instrumentos en los que el peso de la subjetividad desaparece. Esos instrumentos de alta confiabilidad son aptos para evaluar un cierto tipo de aprendizajes o rasgos pero no para evaluar todos. Las llamadas "pruebas subjetivas" o "tradicionales" (exámenes o cuestionarios orales, pruebas escritas de ensayo o disertación o de respuesta extensa; pruebas de respuesta restringida o de cuestionario; monografías, tesinas y tesis; evaluación de portafolio; estudios de casos, entre otras) están destinadas a evaluar la producción del alumno sin pautas totalmente estructuradas ya que le dan libertad para escoger la información, organizarla, expresarla y presentarla. En estos casos la validez aumenta a expensas de la confiabilidad. Esta última no es, entonces, condición necesaria de la validez sino, por el contrario, una limitación de la validez en ciertas circunstancias, aquellas en las que se busca que la respuesta del alumno sea original y permita demostrar competencias que no pueden ser evaluadas por medio de instrumentos cuyas respuestas están previamente estructuradas por los evaluadores.

Conviene señalar, sin embargo, que según se trate de propósitos de alcance e impacto muy importante en las vidas de las personas (exámenes de admisión o de selección, por ejemplo) o de las instituciones, la transparencia de las situaciones de evaluación constituye una exigencia ineludible para la evaluación y que, para ello, se requiere una alta confiabilidad además de validez predictiva.

La confiabilidad es, por lo tanto, una característica importante, difícil de aplicar adecuadamente y que exige un manejo técnico muy correcto de las técnicas de la evaluación puestas al servicio de los propósitos que ésta tiene asignados en los procesos de enseñanza.

PRACTICIDAD

Además de las características mencionadas -validez y confiabilidad- hay que considerar una tercera, de mucha importancia en los programas de evaluación.

La practicidad de un programa o un instrumento resulta de la conjunción de tres aspectos: su administrabilidad, la facilidad de análisis e interpretación de sus resultados y elaboración de conclusiones y, por último, de la evaluación de la economía de tiempo, esfuerzo y costo de su utilización.

Administrabilidad

La administrabilidad de un programa o un instrumento es un producto de varias características:

1. el tiempo de trabajo de los docentes que insume su diseño y construcción;
2. el tiempo que exige su puesta en práctica;
3. la claridad de la/las consigna/s, es decir, de las directivas que se dan a los estudiantes;
4. los materiales, equipos y lugares especiales necesarios para su administración;
5. la cantidad y preparación de las personas indispensables para la administración, cómputo, análisis e interpretación de los resultados y elaboración de las evaluaciones.

Cuanto menor sea el tiempo necesario para el diseño y la elaboración de las conclusiones y menos especializado el personal necesario para la puesta en práctica y el cómputo, mayor será la administrabilidad del programa o instrumento.

ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS Y ELABORACIÓN DE CONCLUSIONES

Se puede advertir que hay una correlación inversa entre el tiempo que demanda la construcción de un instrumento y el que exige el análisis e interpretación de sus resultados. Es necesario señalar, sin embargo, que el diseño de programas e instrumentos de evaluación siempre requiere mucho tiempo, también cuando se trata de pruebas "subjetivas". E1 tiempo de "corrección" de las pruebas o de registro de observaciones suele ser prolongado pero la tarea puede ser facilitada por la construcción de claves o grillas de corrección elaboradas previamente a la administración de la prueba, sea escrita, oral o de evaluación de procesos o productos a través de la observación. Ellas permiten disminuir el tiempo y los problemas que se les plantean a los docentes en el momento de la corrección, les dan mayor seguridad y contribuyen a mejorar la confiabilidad de los resultados de la evaluación. Se deben emplear con precaución para no destruir el espíritu de estas pruebas que es la libertad de presentación y organización que se le brinda al alumno. De otro modo sería preferible utilizar instrumentos de tipo objetivo, esto es, instrumentos que admiten una y sólo una respuesta correcta y no diferentes calidades de respuesta que son más difíciles de analizar.

Economía

Es importante, como hemos dicho antes, que el programa de evaluación sea económico en tiempo, trabajo y costo. Es conveniente examinar previamente estas características para determinar la conveniencia de su empleo.

INTEGRACIÓN DE LOS TRES ASPECTOS

Los tres aspectos relativos a la practicidad del programa o de cada instrumento deben considerarse conjuntamente para lograr una evaluación integral de este rasgo.

Cada aspecto presenta ventajas y desventajas que se compensan o equilibran cuando se analizan los otros dos. Un costo relativamente elevado, por ejemplo, podría ser compensado por un alto grado de administrabilidad, o un tiempo prolongado dedicado a la construcción de un instrumento puede ser equilibrado por un tiempo menor destinado a la corrección e interpretación de los resultados. Del mismo modo, un tiempo extenso de construcción puede compensarse por la facilidad de administración simultánea a un grupo muy numeroso de alumnos.

La practicidad no es, seguramente, la característica más importante de un programa o de un instrumento de evaluación. Pero sí adquiere relevancia cuando de ella depende la frecuencia de la evaluación. Se observa que, por lo general, cuando el programa es práctico o cuando el instrumento es de fácil administración e interpretación se evalúa más frecuentemente y ello contribuye a estructurar mejor los

procesos de enseñanza y los procesos de aprendizaje de los alumnos por parte del docente y de los mismos alumnos.

UTILIDAD

Íntimamente asociada a las otras tres características que mencionamos -validez, confiabilidad y practicidad-, la utilidad de una prueba resulta de su capacidad para satisfacer las necesidades específicas relacionadas con los procesos de enseñanza y aprendizaje. Si los resultados de la evaluación no son útiles para la orientación de los alumnos, del docente, de la escuela como institución, de los padres de los alumnos, de los sectores interesados en la calidad de la educación, poco importa que, en un marco restringido, el programa y cada instrumento sean confiables y prácticos. En lo que respecta a la validez, ya hemos dicho que es relativa a los criterios y, tratándose de públicos diferentes, dichos criterios pueden ser heterogéneos. Para que la evaluación sea útil, la validez para alguno o algunos de esos públicos es indispensable. La pregunta esencial en el diseño de un programa de evaluación es, por lo tanto, ¿qué usos se les dará a los resultados obtenidos? Son los implicados en los procesos de evaluación quienes deben dar las respuestas.

BIBLIOGRAFÍA

Cardinet, Jean: "Evaluer sans juger", *Révue Française de Pédagogie,* n° 88, julio-agosto-setiembre de 1989.
Dietel, R. J., Herman, J. L. y Knuth, R. A.: "What does Research say about Assessment?", NCREL, Oak Brook, 1991.
Ebel, Robert L.: "Must all Tests be Valid?", en *Human Learning in School*, ed. John De Cecco, Nueva York, Holt, Rinehart & Winston, 1963.
- *Fundamentos de la evaluación educacional*, Buenos Aires, Guadalupe, 1977.
Garrett, Henry E.: *Estadística en psicología y educación*, Buenos Aires, Paidós, 1966.
Gronlund, Norman E.: *Medición y evaluación en la enseñanza*, México, Pax, 1973.
House, E. R., *Evaluación, ética y poder*, Madrid, Morata, 1994.
Karmel, Louis J.: *Medición y evaluación escolar*, México, Trillas, 1986. Perrenoud, P.: La construcción del éxito y el fracaso escolar, Madrid, Paideia-Morata, 1990.
Pidgeon, Douglas y Yates, Alfred: *Evaluación y medida del rendimiento escolar*, Madrid, Anaya, 1979.
Popham, James W.: *Evaluación basada en criterios*, Madrid, Magisterio Español, 1983.
Sachs Adams, Georgia: *Medición y evaluación*, Barcelona, Herder,1970. Thorndike, Robert L. y Hagen, **Elizabeth P.:** *Medición y evaluación en psicología y educación*, México, Trillas, 1989.