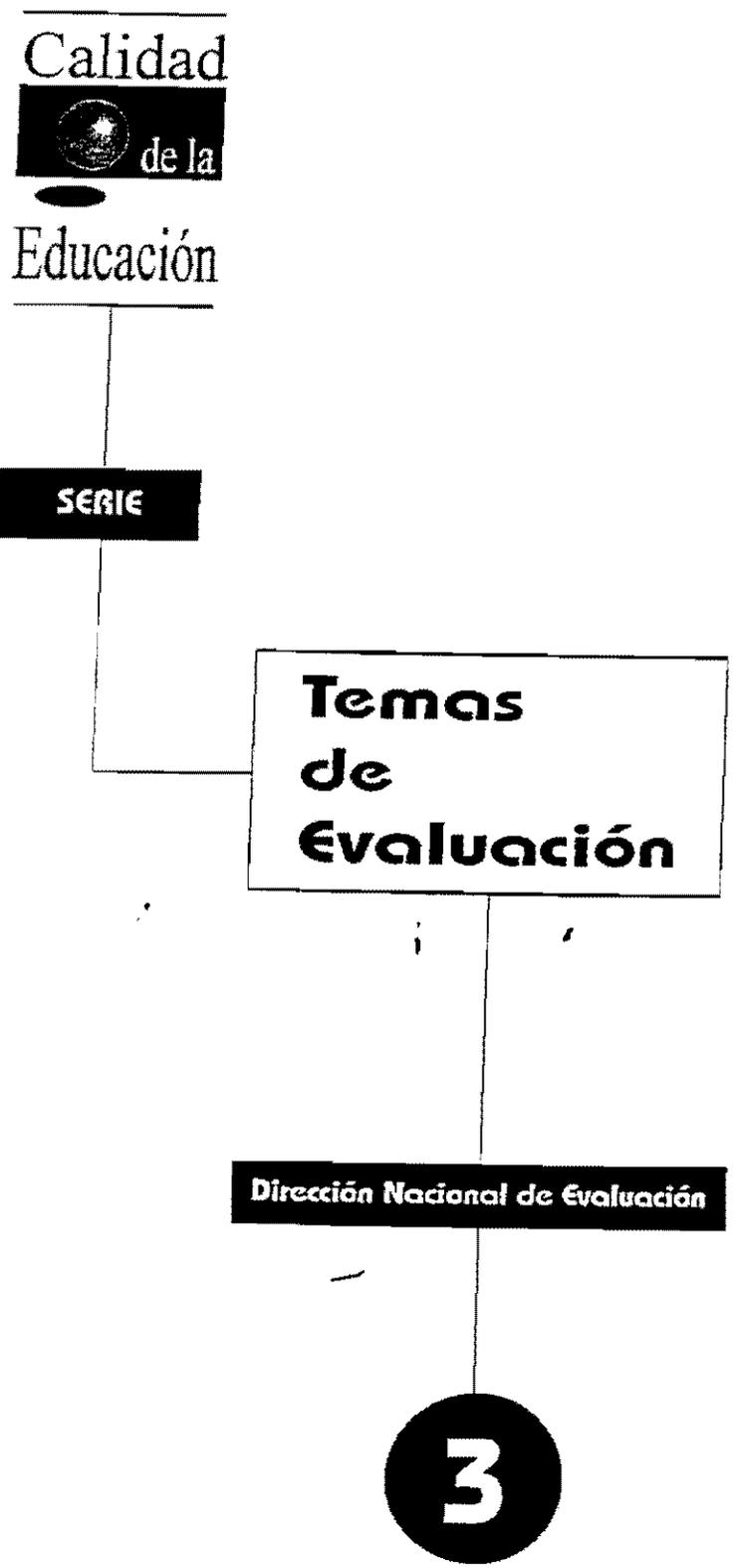


21554



INV	021554
SIG	FOLL 371.26
LIB	5

PRESIDENTE DE LA NACIÓN  
DR. CARLOS SAÚL MENEM

MINISTRA DE EDUCACIÓN DE LA NACIÓN  
LIC. SUSANA BEATRIZ DECIBE

SECRETARIO DE PROGRAMACIÓN Y EVALUACIÓN EDUCATIVA  
DR. MANUEL GARCÍA SOLÁ

SUBSECRETARIA DE EVALUACIÓN DE LA CALIDAD EDUCATIVA  
PROF. HILDA MARÍA LANZA

SUBSECRETARIA DE PROGRAMACIÓN EDUCATIVA  
LIC. INÉS AGUERRONDO

SUBSECRETARIO DE GESTIÓN EDUCATIVA  
PROF. SERGIO ESPAÑA

DIRECTORA NACIONAL DE EVALUACIÓN  
LIC. MARÍA LUCRECIA TULIC

EQUIPO DE COORDINACIÓN

LIC. ANA DIAMANT  
LIC. GABRIEL PETRUCCI  
ING. CARLOS RONDÓN CARDOSO

EQUIPO DE EDICIÓN

ROSANA MASINI  
HECTOR O. MARTÍN  
JAVIER MAIZA

Calidad



de la

Educación

**Parte 1**

**La enseñanza y  
la evaluación**

Adaptación  
de la conferencia  
dictada en  
**CONSUDEC**  
Julio, 1995

**Nydia Elola**

Abordaremos la temática de la enseñanza en relación con la evaluación.

Habitualmente, cuando consideramos la transformación educativa lo hacemos pensando sobre todo en el sistema, en los contenidos dentro del sistema, en los ciclos, y estamos hablando poco del aula. Ese es el lugar donde se ubicará este trabajo: en el proceso de enseñanza, en eso que ocurre cuando se juntan alumnos y docentes para tratar de que pase algo.

Durante su formación, el docente aprendió, en términos actuales, *procedimientos* para manejarse con grupos chicos, para manejarse en situación interpersonal, que implica una multiplicidad de factores que están simultáneamente presentes en toda la situación de enseñanza. ¿Qué significa esto?, que están presentes los chicos, que cada uno es distinto, que están los docentes, también cada uno con sus características, está presente la sociedad, está presente el Ministerio o quien sea la jurisdicción que ha dado ciertos mandatos sobre lo que se tiene que enseñar. Todas estas cosas están presentes permanentemente, y esto hace que la situación de enseñanza sea una situación muy dinámica, muy impredecible, que exige un ajuste permanente a lo que está pasando en el aula. Este ajuste permanente es una manera de responder a las situaciones. Cuando estamos atentos a qué le está pasando a los alumnos y también a qué se quiere enseñar, resulta que todo lo que va ocurriendo es novedoso, no sabemos qué va a pasar en el momento siguiente.

La mayoría de las situaciones de enseñanza tienen semejanzas entre sí. Encontramos un qué se quiere aprender, hay un para qué se está en esto, hay un cómo representado por ciertas formas propuestas y se buscan algunas formas de investigar el valor de lo que está ocurriendo -de evaluación-. La situación de enseñanza tiene como identificatoria la intención de que ocurra un proceso de aprendizaje, de enseñar para aprender. Esta intención es lo que hace que, por diferentes que sean las situaciones, podamos clasificarlas dentro de los quehaceres que identificamos como enseñanza.

Hay otro conjunto de cosas, aquellas que se pueden prever antes de la clase, antes de que ocurra el encuentro entre docentes y alumnos. Para tener cierta garantía de poder guiar una situación de clase, es necesaria una previsión, una planificación -si se quiere expresar en términos técnicos- de algunos elementos, tomando en cuenta el juego de variables que podemos predecir. A partir de esa hipótesis se organizan las actividades que son consideradas imprescindibles para que ocurra el proceso de enseñar y de aprender. Habrá que tener en cuenta que hay muchas otras circunstancias que van a exigir decisiones inmediatas en el aula, diferentes a las planificadas.

Peero para poder estar atento a lo que pase, para poder estar atento a los alumnos y responder a sus demandas, es necesaria la libertad que da tener previstas ciertas cosas. Con esto estamos defendiendo la necesidad de alguna forma de planificación, la tan maltratada planificación. La planificación es nada más, ni nada menos que esto: prever lo previsible para poder responder mejor a necesidades y objetivos.

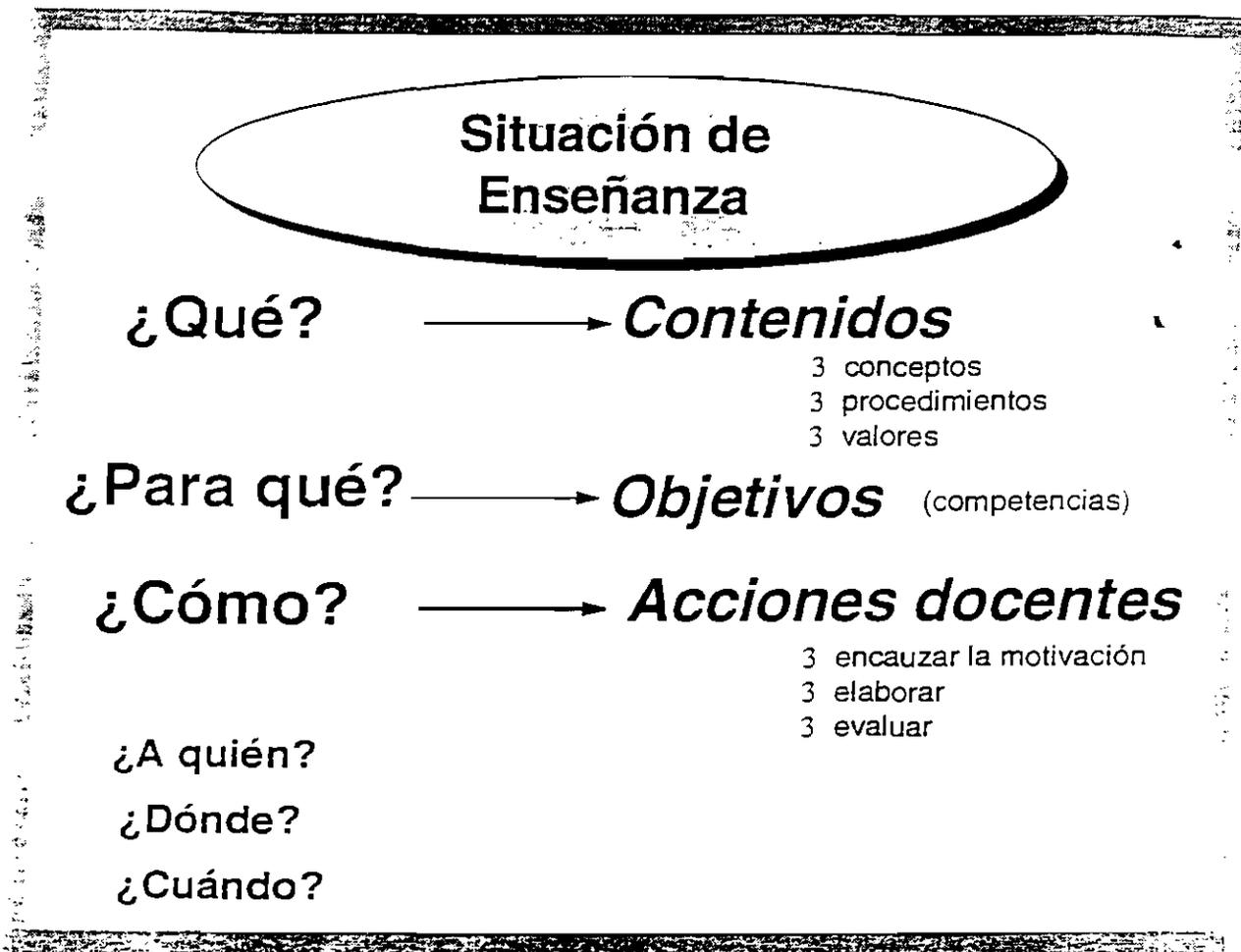


Figura 1

Se pueden tomar distintos tipos de esas decisiones previas que dan libertad en cada momento. En una situación de enseñanza hay un primer punto: está la idea del *qué*. Lo primero que generalmente se prevé es qué es lo que se quiere enseñar: es la idea de contenidos. Cuando decimos qué se quiere enseñar, estamos hablando de aquellos objetos -objetos en sentido muy amplio- que se pretende que el alumno, de alguna manera, haga suyos, es decir se está esperando que el alumno incluya en su estructura de conocimiento una cierta clase de cosas. Esa clase de cosas son las que se llaman **contenidos**. Importa mucho definir ahora el concepto de contenido, porque después habrá que diferenciarlo de otros. En este momento, se está hablando básicamente de tres tipos de contenidos: contenidos conceptuales, procedimentales y valorativos. Cuando pretendemos que el alumno haga suyo una clase de contenidos, estamos pretendiendo que haga algo con ellos: ésta es la noción de objetivos. Vamos a intentar que los alumnos se apropien, hagan suyos esos contenidos, para algo, para que se modifique algo en ellos, para que sean capaces de hacer ciertas cosas. Aparece así el concepto de **competencia**.

Cuando se tienen claros los contenidos, cuando se han propuesto los objetivos y competencias, se debe pensar cómo se va a llegar a que esos contenidos se hagan propios en el alumno; ésta es la idea de acciones docentes, el *cómo* es todo lo que hace el docente para que esos contenidos se internalicen en los alumnos. No estamos planteando conceptos novedosos, lo que estamos

buscando es subrayar aspectos que después se van a contraponer. Las acciones docentes implican una guía del aprendizaje, y esta guía se puede sintetizar en tres clases de acciones: las que tienden a encauzar motivaciones y conocimientos, las que llevan a elaborar y las que tienen por finalidad evaluar. No es una secuencia, son tres clases de acciones que apuntan a distintos aspectos del aprendizaje.

Se pueden prever muchos otros aspectos: quién o quiénes, dónde, cuándo. Trabajaremos sólo con el qué y el cómo.

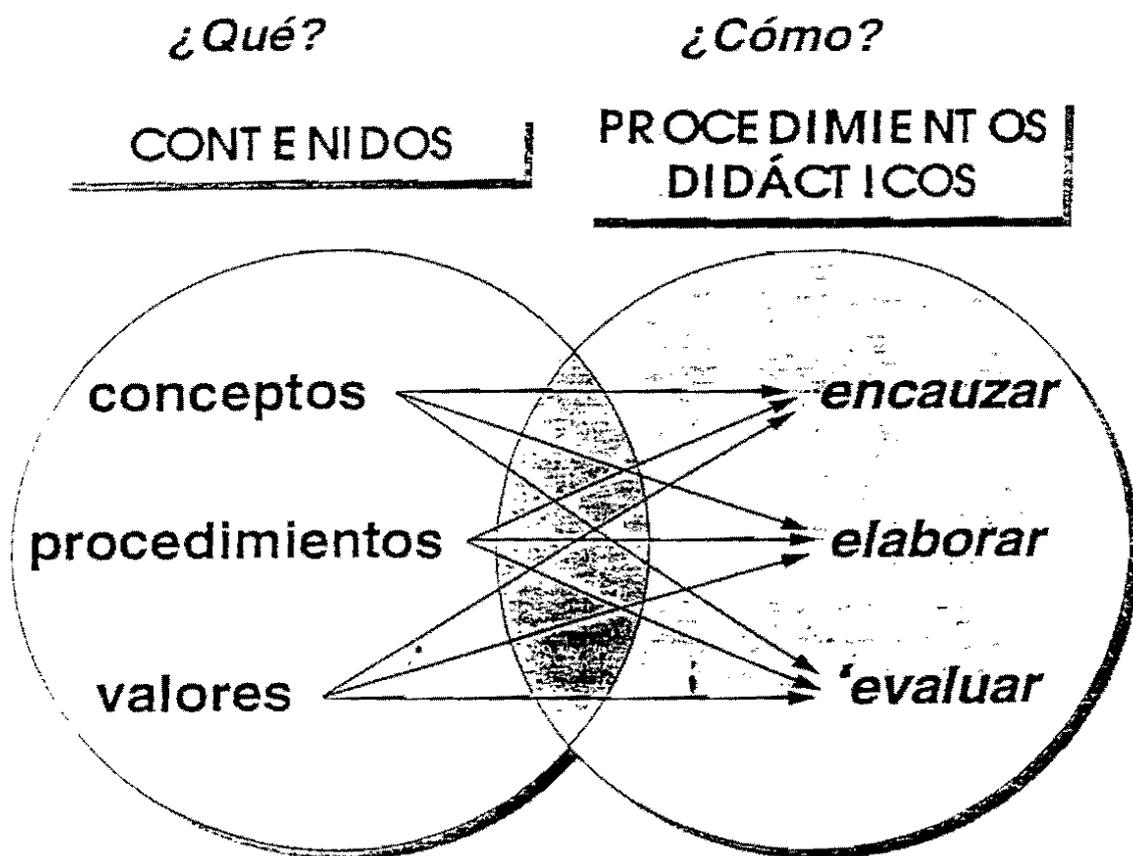


Figura 2

Hay un conjunto de qué, los conceptos, los procedimientos, los valores. No son jerárquicos, aunque se presenten en un orden. No son más importantes los conceptos que los procedimientos y que los valores, las tres cosas son objeto de la enseñanza.

Siempre se han enseñado conceptos, procedimientos y valores, ahora se está tratando de analizarlos, así como se están analizando una serie de factores de la enseñanza. ¿Por qué?. Porque estamos pensando en la enseñanza como un proceso de gestión. Lo que estamos intentando es que la enseñanza use tecnologías gestionales. La enseñanza no es una tecnología gestional, usa tecnologías gestionales. ¿Qué quiere decir esto? que hay un margen de libertad, de creatividad y de artístico en la enseñanza que tiene que ver con todo eso impredecible que se señalaba antes pero que, cuando lo preveo, estoy previendo un manejo gestional.

La intención de diferenciarlos es, por un lado, para no olvidar de desarrollar estas distintas cosas, que no son siempre parejas, ni van juntas, no siempre que se enseña un concepto tendríamos que enseñar los procedimientos adecuados para trabajar ese concepto o viceversa, no siempre que se enseña un procedimiento es posible o conveniente enseñar el concepto o los conceptos subyacentes a ese procedimiento. En seguida se va a ver ejemplos de esto. Una razón por la cual conviene separarlos es para llamar la atención sobre ellos y tomarlos en cuenta. Pero además hay aquí un punto importante de las teorías de la enseñanza y del aprendizaje sobre el cual queremos hacer hincapié. No se enseña ni se aprende de la misma manera cada uno de los tipos de contenidos. Por eso la imagen (fig 2) hace corresponder a cada concepto a enseñar, un procedimiento didáctico para encauzar esa enseñanza, un procedimiento didáctico para elaborar conceptos, y un procedimiento didáctico para evaluar conceptos. Lo mismo ocurre con los procedimientos, no se puede encauzar el aprendizaje de procedimientos de la misma manera que se encauza el aprendizaje de conceptos o de valores. Es decir cada uno de estos elementos está forzando, está llevando a que se elija una manera de enseñarlo para que se aprenda diferenciadamente, porque son clases de cosas distintas.

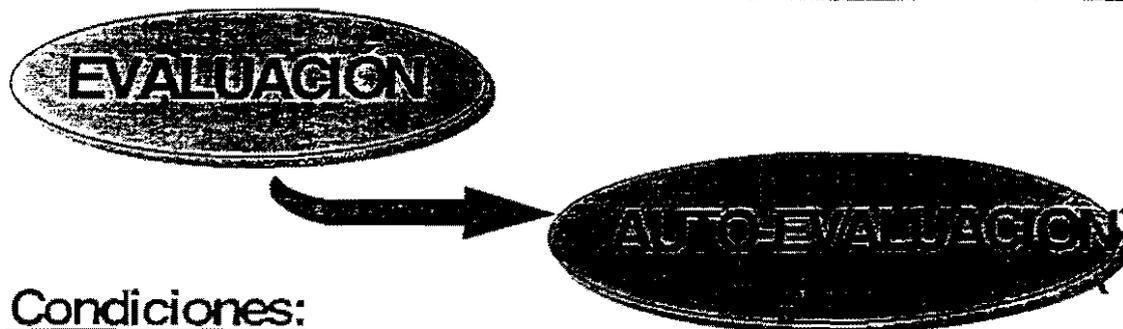
Estas clases de cosas a veces no son distintas en sí mismas, sino que la diferencia está en función de la enseñanza. Por ejemplo, en los C.B.C. del área de Tecnología, se plantea un procedimiento esencial: el análisis de los productos. Se habla de todos los elementos que implica ese análisis de un producto, y de los distintos tipos de análisis: un análisis morfológico, un análisis estructural, de funcionamiento, de función, una larga lista. Ese procedimiento se puede usar, por ejemplo con respecto al termo, al lápiz, etc. Los alumnos pueden analizar ejemplares de esos objetos como producto tecnológico. Por supuesto, de ese proceso surge el concepto de "termo" o de "lápiz", pero no interesa el concepto como tal, lo que importa es destacar "qué clases de procesos se pueden seguir para diferenciar un producto tecnológico", eso es lo que podemos orientar elaborar y evaluar.

En tanto los conceptos tendrían por detrás una red, un mapa, una estructura conceptual, los procedimientos tendrían por detrás un diagrama de flujo. Habría una secuencia de pasos para aprender un procedimiento, una secuencia que puede ser con algunos pasos "y/o", es decir, si pasa tal cosa entonces... esto, o tal otra cosa, o tal otra cosa... Esta gran diferencia ya está exigiendo encauzarlo de una manera distinta y elaborarlo de una manera distinta. Los

procedimientos se dan en el tiempo, los conceptos son una unidad, tienen un significado único que aparece sincrónicamente. En cambio, cuando al hablar de contenidos valorativos, es necesario pensar que los valores aparecen adheridos a diferentes conceptos y procedimientos, entonces también hay que aprender y enseñar con otro esquema. Muchas veces la forma en que se plantea o presenta un concepto o un procedimiento está implicando la valoración con la que se lo está cargando. Este es el punto clave, es imprescindible que estos tres procedimientos didácticos se diferencien en función del tipo de contenido que se quiere plantear. No se puede, en ausencia de ciertas condiciones, enseñar procedimientos. El procedimiento lleva a un saber hacer, para saber hacer hay que hacer. En un gran auditorio, quien sólo escucha lo único que puede "hacer" es seguir un tipo de pensamiento en el orden en el que se está presentando, pero no se está rescatando un procedimiento. Más adelante se presenta una propuesta de proceso de evaluación. El lector puede seguir un proceso de evaluación, una línea de acción para evaluar, pero no puede aprender a hacerlo. Es posible aprender el proceso de evaluación como un concepto, pero no será una línea de acción. No es que los objetos sean en sí mismos procesos o conceptos sino que se los están enseñando y aprendiendo como tales. Este también es un punto importante a tener en cuenta.

Además volviendo a la imagen, se está usando en dos círculos la misma palabra: **procedimientos**. Procedimientos como contenido y procedimientos didácticos en el cómo, o sea hay maneras de proceder para enseñar cada una de estas cosas, pero no son lo mismo que el contenido procedimientos, salvo en la formación de docentes. En la formación de docentes esos *procedimientos didácticos* son un contenido, pero en la escuela primaria o en el nivel inicial, el procedimiento didáctico no es para que el alumno lo aprenda, el procedimiento didáctico debe ser el más adecuado para que aprenda cualquiera de los contenidos.<sup>1</sup> Esto es importante tenerlo en cuenta cuando se leen los Contenidos Básicos Comunes. En algunos casos pareciera que la diferencia entre el contenido básico y la actividad elegida para la enseñanza no está del todo clara y eso puede perturbar, llegar a mal interpretar que hay que aprender la actividad, en lugar de aprender el procedimiento. Hay un procedimiento didáctico que es el más adecuado para cada uno de estos momentos y para cada uno de estos contenidos, pero el procedimiento didáctico en sí no es el que hay que aprender, no es lo que se quiere proponer como objeto de aprendizaje, usamos muchos procedimientos didácticos para acceder a un mismo objeto.

Evaluar, en la figura, está puesto en último término pero esto no es un orden. *Evaluar* puede ocurrir en cualquier momento. *Evaluar* es cuando se lo considera necesario. Eso puede suceder en momentos diferentes: puede considerarse necesario antes de encauzar algo, puede ser necesario para saber cuál es el momento actual de aprendizaje, entonces se evalúa. Podemos investigar qué motivaciones tienen los alumnos, qué conocimientos previos traen consigo, en qué están en relación a un determinado contenido. Esa es una circunstancia para evaluación. Cuando estamos en plena elaboración, hay una crisis de aprendizaje, en esas crisis es necesario evaluar. Esas evaluaciones pueden aparecer en cualquier momento. Requieren diferentes procedimientos y los procedimientos están en función del sentido y el significado de la *evaluación*.



**Condiciones:**

- 3 manifiesta
- 3 explícita
- 3 con resultados compartidos



**INFORMACIÓN**  
**CALIFICACIÓN**  
**PROMOCIÓN**

Figura 5

Evaluar -dice un autor<sup>2</sup>- es ayudar a otro con menor capacidad, con menor formación en lo que se está haciendo, a que vea lo que de otra manera permanecería oculto. Si estamos intentando aprender, por ejemplo, a usar la computadora -el drama actual de los adultos, no de los chicos- ¿a quién buscamos para que nos ayude a manejar la computadora?. Al que sabe más, en todo caso nos juntamos unos cuantos que sabemos poco, llamamos a alguien que ayude. Esa persona es la que, con su evaluación, va a ayudar a ver qué hicimos, dónde erramos, por qué se "pinchó" la computadora. Ahí es cuando ayuda a ver lo que de otra manera permanecería oculto". Indudablemente, si se usa este concepto de evaluación, se está hablando de un algo muy distinto de la calificación y la promoción, se trata de otro proceso, que está involucrado esencialmente en los de enseñar y de aprender. Y este concepto lleva necesariamente a la *autoevaluación*. Cuando, a través de la evaluación, nos damos cuenta de dónde estamos, qué está pasando, qué momento del aprendizaje atravesamos, estamos elaborando una autoevaluación. Y esto es fundamental, porque no siempre tenemos quien nos enseñe. Habitualmente, en los comienzos tenemos alguien que enseña, pero después seguimos solos. Esto es, los docentes son un "accidente" en nuestra vida. El docente que está enseñando, en algún momento deja de estar al lado nuestro; está cumpliendo su función docente cuando ayuda a que armemos nuestras propias formas de evaluarnos. Si la evaluación no se va convirtiendo en una autoevaluación no es formativa, no va ayudando a cambiar y a crecer.

Para que esto ocurra, es necesario que se den ciertas condiciones: la evaluación tiene que ser manifiesta. En una verdadera evaluación *enseñante* no alcanza con que sólo el docente perciba como se está desarrollando el proceso de aprendizaje; es necesario que sea manifiesta y que sea explícita y explicitada, que haga evidente en qué estamos cada uno de los participantes de la situación. Esto implica que podamos trabajar con resultados compartidos; es decir que la información que obtenemos a través de la evaluación sea compartida, compartida con todos los que tienen que ver en el proceso.

Hasta ahora estamos hablando de algo que no es lo que habitualmente sugiere la palabra evaluación. La palabra evaluación comúnmente refiere, hace pensar en *calificación*, hace pensar en *promoción* y también en *boletines* y *certificados* (la información de esa evaluación). Trataremos de diferenciar cada uno de estos conceptos. El proceso de evaluación es inherente al proceso de enseñar y de aprender, la evaluación representa la retroalimentación necesaria para esos procesos.

La calificación es una manera de ponderar y traducir los resultados de la evaluación. Se traduce a una escala convencional, a una escala comparativa que todos puedan entender, que supuestamente tiene un significado común -supuestamente porque a veces no lo tiene tanto- pero esta calificación no es imprescindible para el proceso de evaluación. Lo que es imprescindible es compartir resultados, pero no necesariamente a través de la calificación. Cuanto más analítico sea el resultado mejor va a informar sobre lo que está oculto. Con una nota o un concepto global no hay elementos suficientes para reconstruir el proceso de aprender a partir del momento presente. La calificación es sintética, informa poco, en todo caso pone en un ranking. Para reelaborar es imprescindible una información más exhaustiva que ayude a ver la situación presente. La calificación es necesaria en una institución, en un sistema, pero no siempre hace falta en el aula. En el aula se pueden hacer acciones concretas, específicas y manifiestas de evaluación sin que se traduzcan en una calificación, pero sí que estén informadas, que se compartan los resultados, pero no necesariamente que sean calificados.

Y mucho menos exige -el aula- la promoción. La promoción es un planteo de la institución. La institución escolar promueve. ¿Por qué promueve?. Para pasar etapas, para pasar dentro del sistema, se trata de un problema de la organización. En un grupo de taller, por ejemplo, fuera de la escuela, hay enseñanza, aprendizaje y también evaluación. Podemos encontrar todos esos procesos, ¿pero, hay promoción?. No, no es necesaria. Se van superando logros en el aprendizaje sin necesidad de decir "pasa a otro grupo o curso". La escuela si lo exige. Pero el problema está en que esto que debería ser un resultado final de toda la labor del aula, está presente en todo momento. La escala de calificación no es una escala neutra, según el sistema de promoción que se tenga, tanto el 4 como el 7 (en una escala de 10) son señales luminosas. La distancia entre un 3 y un 4 y entre un 4 y un 5 no es idéntica. No vale lo mismo el punto de diferencia entre 3 y 4 o entre 4 y 5. Actuamos de una manera distinta cuando tenemos que calificar sintetizando el resultado de una evaluación si sabemos que esa nota se va a ponderar, se va a promediar, se va a usar para la promoción. O sea que la promoción está distorsionando -en cualquier sentido-, está perturbando el informe de los resultados. Por lo tanto, en ese momento la evaluación no ayuda a ver lo que de otra manera permanecería oculto, porque se la está tiñendo con factores que son externos al aprendizaje en el aula. Este es un punto importante para reflexionar; no se trata de obviar la promoción ni la calificación, pero sí de diferenciarla.

Por otra parte en todos los niveles, se informa acerca de la evaluación. La información es un momento importante de la evaluación, pero no es lo mismo. La información es imprescindible para cerrar un momento y abrir otro en juego dinámico, pero debe ser ponderada, no hay por qué informar todo lo que se evalúa. No se trata de ocultamiento, sino de que la información tiene que ser una información que sirva, que informe a quién tiene que informar. Entonces hay que usar la información necesaria y suficiente para cada receptor: alumnos, padres, el mismo docente, directivos, colegas, otras instituciones, etcétera.

La escuela puede y debe evaluar a través y a partir del aprendizaje muchos otros procesos internos de la organización, de la información que se obtenga, todo lo que el chico entienda, hay que compartirlo con el chico, y todo lo que al padre le sea significativo también hay que compartirlo. No ocultar, pero sí ponderar cuál es la información que se da.

Par obtener esas informaciones, estamos proponiendo el análisis del proceso de evaluación que se muestra en la figura 4.

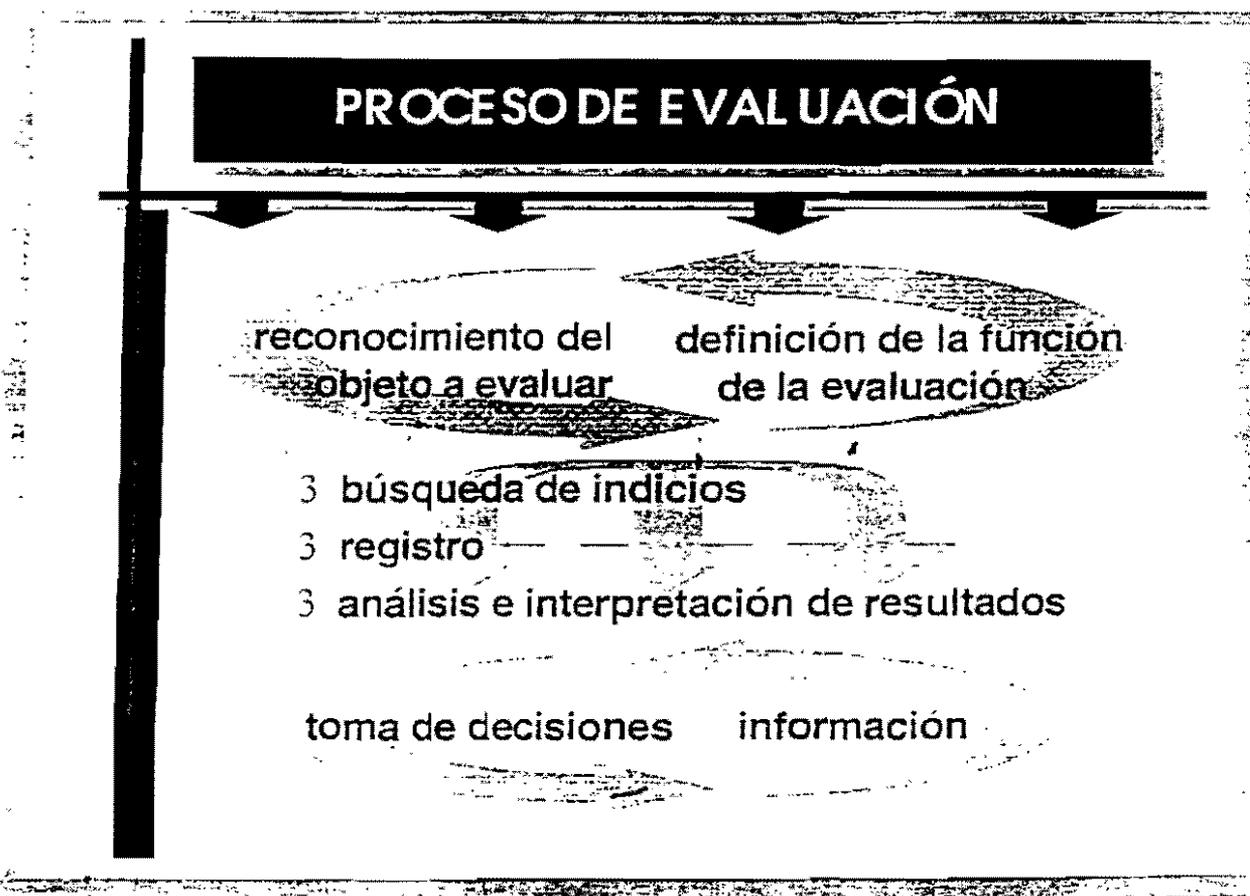


Figura 4

Por una parte, es necesario saber qué se quiere evaluar, reconocer claramente el objeto que se quiere evaluar, y por otro lado, definir para qué se lo quiere evaluar, que función o funciones espero que cumpla esa evaluación. Ambas decisiones con cierta simultaneidad, el para qué va a modificar un poco el qué quiero evaluar y esto se va dando como un círculo espiralado, donde es imperioso revisar las dos cosas juntas para después poder pensar qué indicios<sup>3</sup> buscar de aquello a evaluar, o sea, en qué

se va a ver todo lo que queremos evaluar. Registrar esos indicios- esto apunta al procedimiento de evaluar-. Es el momento clásico de recolección de los datos. Buscamos los indicios, los registramos para permitir el análisis e interpretación de los resultados (no de los resultados del aprendizaje<sup>4</sup> sino del resultado de nuestra búsqueda). Después de analizar e interpretar es el momento de decidir -la evaluación siempre es fundamento para tomar decisiones- cuál es el curso de acción apropiado en función de esos resultados e informar. Otra vez, estos dos procesos no tienen un orden sucesivo, aparecen juntos (ver figura 4). Tomar una decisión de cómo seguir con la enseñanza, por dónde seguir, o qué rever, y cómo informar y a quién se puede informar, a los chicos, a los padres, a la escuela, al sistema escolar. La forma en que resolvamos uno influye en como resolver el otro. Estas secuencias y simultaneidades se dan siempre, aunque a veces no seamos tan conscientes y no identifiquemos estos pasos. En este proceso también se sitúa el momento de la calificación. Cuando analizamos e interpretamos los resultados podemos traducirlos a una calificación, es posible informar a través de una calificación. Y también, a partir de este análisis e interpretación de los resultados es que se puede decir "este chico está en condiciones de ser promovido", se toman decisiones por ejemplo para la promoción. El análisis e interpretación de los resultados es el punto clave para hacer de la evaluación un proceso educativo.

En el momento del registro se utilizan distintos instrumentos, por ejemplo, guías de observaciones descriptivas, pruebas, producciones de los alumnos, cuestionarios, etc. La elección del instrumento va a depender de "el objeto a evaluar y la función de la evaluación", los procesos y contenidos que se quieran evaluar y de para qué los quiera evaluar<sup>5</sup>.

Los *para qué* aparecen en la otra figura.

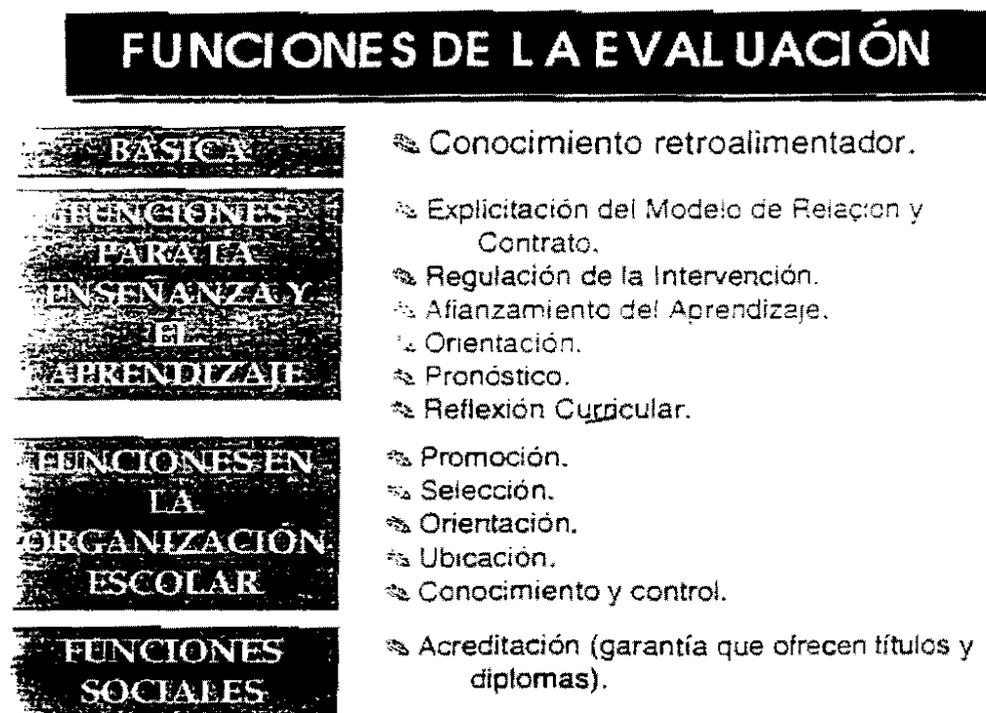


Figura 5

Vamos a enunciarlos rápidamente. Con esta visión de conjunto apenas nos asomamos a la complejidad de funciones que están adscriptas al concepto de evaluación y muchas veces las controversias acerca de ella son esencialmente un juego de atribuciones de énfasis diferentes a algunas de las funciones.

Debemos identificar siempre un objeto a evaluar y un para qué, una función. Normalmente la evaluación puede cumplir muchas funciones, la primera y básica es la que aparece en primer lugar: en tanto ayuda a ver lo que permanecería oculto, es un conocimiento retroalimentador, retroalimenta el proceso para poder seguir adelante. Esto ocurre siempre, sea buscado o no. ¿Por qué enfatizar esto?. Nos exige cuidado en tanto el cómo se evalúa retroalimenta siempre el proceso, y habría que ver si lo hace en la dirección pretendida. Si no cuidamos todo el proceso probablemente no ofrecemos ni gestionamos la retroalimentación que quisiéramos dar.

La evaluación presta *funciones para la enseñanza y para el aprendizaje*, en razón de que *explicita un modelo de relaciones, de contrato*<sup>6</sup>. Otro punto clave, según como evalúe es el mensaje que se está enviando acerca del proyecto educativo, qué es lo que realmente queremos que ocurra en este proceso, qué estamos dispuestos a dar y a recibir en esta situación.

Si se evalúa repetición de conocimientos -sólo procesos reproductivos- si se evalúan aspectos creativos, si en la evaluación se permite que los chicos muestren sus pensamientos, sus ideas, o solamente se pide "lo que nosotros queremos". Hay un precioso chiste de Mafalda, en el que cuando la maestra le presenta una prueba ella le dice: "Señorita, ¿usted qué quiere que yo conteste?, lo que yo sé, lo que yo pienso, lo que usted cree que yo debería saber o pensar, o lo que quiere la escuela y la sociedad que yo diga..." Es decir todo esto es el modelo de relación y contrato. Se puede decir "chicos, me importa que piensen" pero en la prueba pedir que repitan, "¿qué es la fuerza?", "la fuerza es...", "repitan", el chico va a responder al "repitan", y no al piensen. También para la enseñanza y el aprendizaje otra función es la de *regular la intervención*, cuando en un momento de crisis se evalúa para revisar y encauzar lo que está ocurriendo, es decir cómo se va adelante, cómo se sigue. Esta es la idea central, la evaluación permite regular la intervención.

Siempre la evaluación representa un afianzamiento del aprendizaje, es una orientación, sirve para pronóstico y es una reflexión curricular. En estos momentos que se está hablando del desarrollo curricular, el desarrollo curricular debe ir reajustándose permanentemente. La evaluación es la herramienta para el ajuste.

Se trata ahora de revisar las *funciones en la organización*. El aula es también la organización, no puede ser pensada separada. Acá está la idea de promoción. Aquí la de selección. Con selección se quiere decir por ejemplo: "con este grupo se van a trabajar tales cosas, con este grupo tales otras... porque eso es lo que están necesitando estos chicos", se agrupa. Seleccionar no quiere decir buscar a los mejores, quiere decir *diferenciar para distintas actividades*. Orientación, ubicación, conocimiento y control. Son acciones encadenadas. Obviamente los resultados de la evaluación son conocimiento y control de la organización acerca de la calidad de lo que se está haciendo. Esto es indudable, ocurre en la escuela, ocurre en el aula y ocurre en el sistema.

Además la evaluación del aula implica una **acreditación**, o sea qué garantía ofrecen los títulos y los diplomas dentro de la sociedad. La evaluación que se está haciendo en el aula y que se comunique como resultado de la enseñanza y el aprendizaje tiene una trascendencia social, la tiene siempre...

Cada una de estas funciones requiere un tratamiento exhaustivo, que no vamos a realizar en este trabajo. Como todo análisis este también parece desarmar el objeto de análisis. En la realidad las funciones no aparecen tan diferenciadas, pero pensarlas así ayuda ponderar y a reflexionar sobre nuestras acciones.

<sup>1</sup> Esto no significa que el procedimiento didáctico tenga un valor neutro para el aprendizaje, la selección que hagamos del procedimiento tiene valor de mensaje educativo.

<sup>2</sup> Eisner, Elliot W., *The Art of Educational Evaluation*, The Falmer Press, 1985.

<sup>3</sup> **Indicio**, según el diccionario, acción o señal que da a conocer lo oculto o desconocido.

<sup>4</sup> Podemos estar evaluando *resultados o procesos de aprendizaje*.

<sup>5</sup> Trabajar en profundidad todos los elementos y momentos del proceso de evaluación descrito exige un trabajo de taller. Este enunciado sólo pretende ayudar a la percepción de la cantidad de decisiones que involucra.

<sup>6</sup> **Contrato**, según definición de diccionario, es acuerdo de dos o más voluntades dirigido a crear una obligación de dar o hacer (...)

**parte 2**

**InstrUmeNtos dE  
evAluaciÓN**

**TomAdo De:**

**EvaLuaciÓN dEl**

**TRabajo EscOlAR**

**UNESCO, Quito,**

**1981.**

**Nydia Elola**

**Dirección Nacional de Evaluación**

## INSTRUMENTOS DE LA EVALUACIÓN

Mediante la evaluación se analizan e interpretan todas las respuestas de los alumnos en situación de aprendizaje.

Cualquier clase de conducta del ser humano puede ser objeto de evaluación (puesto que los objetivos de la enseñanza pueden apuntar a todo tipo de conductas), pero no es posible apreciarlas todas con la misma facilidad y con los mismos recursos. Las manifestaciones de una persona en una situación pueden tener distintos significados. No es siempre fácil desentrañar cuál es el cambio real que ha ocurrido y que puede explicar la conducta presente. Estas consideraciones nos llevan a revisar el campo de los procedimientos de evaluación en busca de las posibilidades y limitaciones que ofrece.

En términos generales, cuando se produce un cambio por aprendizaje, éste es **observable** en la conducta. La observación es el procedimiento general de la evaluación. Si observamos cómo actúa un sujeto antes y después de un determinado aprendizaje podemos inferir que los cambios que aparecen son producto de ese aprendizaje. Pero frente a la necesidad de observación surgen inmediatamente por lo menos dos problemas: 1) cómo sabemos que lo observado, responde exactamente al cambio que nosotros buscamos y no a otro. (Cómo saber, por ejemplo, cuando un alumno da una definición de ángulo, si tiene realmente el concepto de ángulo y no una simple sucesión de palabras 2) no todos los cambios que nos proponemos pueden ser observados en la escuela, y menos aún, en situaciones espontáneas. (Por ejemplo, el interés por la lectura es algo que va a aparecer necesariamente fuera de la escuela).

De estos problemas se desprenden dos necesidades: reelaborar los objetivos para estrechar la relación entre objetivos y evaluación y, por otro lado, ampliar la gama de procedimientos para abarcar todo tipo de situaciones.

### Tabla de especificaciones

Ya hemos hablado de la necesidad de definir los objetivos para evaluar. Pero además podemos apelar a un recurso eficaz para hacer más cercana la relación entre los objetivos y la evaluación. Se trata de la elaboración de una **tabla de especificaciones**. Esta tabla representa el modelo de resultados de aprendizaje deseado, el que se espera evaluar. Generalmente se confecciona como una tabla de doble entrada, donde en una dimensión aparecen los distintos tipos de conductas o de capacidades esperados y, en la otra, los contenidos a los cuales queremos que esas conductas se apliquen. Cada intersección es un objetivo que podemos evaluar. De acuerdo con el marco total de lo que deseamos evaluar y con la función que deseamos que cumpla esa evaluación, se pondera el peso de cada objetivo dentro de la tabla, de manera que refleje, lo más estrechamente posible, la relación entre la evaluación propuesta y los objetivos.

Por ejemplo, si en un curso, después de trabajar una unidad sobre Evaluación del Rendimiento deseamos evaluarla, podemos plantear una tabla de especificaciones semejante a éste:

	Conocimiento de terminología	Comprensión	Aplicación	Cálculo	
Concepto	5 % 3	10 % 6	10 % 6	-	25 % 15
Proceso	5 % 3	10 % 6	5 % 3	-	20 % 12
Instrumentos	5 % 3	10 % 6	10 % 6	-	25 % 15
Análisis de resultados	5 % 3	10 % 6	10 % 6	5 % 3	30 % 18
	20 %	40 %	35 %	5 %	100 % 60

Los porcentajes de los extremos nos indican los pesos relativos que deseamos que tengan en nuestra evaluación cada uno de los tipos de conducta y cada uno de los contenidos. Los porcentajes en cada casillero, los de cada objetivo, la especial combinación de contenido y conducta.

Si esta tabla fuera el esquema para una prueba de selección los porcentajes se podrían traducir en cantidad de ítems. Los números que aparecen en cada casillero representan la cantidad de ítems de ese objetivo en el caso de que la prueba tuviera 60 ítems.

Si tratamos de elaborar la prueba manteniendo esa distribución es probable que la prueba represente fielmente nuestros objetivos. Si elaboráramos la prueba sin ese marco, es posible que aparecieran más ítems de algunos contenidos, habituales aunque no relevantes, o más ítems de conocimiento, por ejemplo, que pueden ser más fáciles de elaborar. Es decir, aun cuando representarían a nuestros objetivos, los ítems no presentarían un muestreo adecuado.

#### ◆ Observación y pruebas

Vamos a trabajar sobre el problema de los procedimientos. La **observación en situaciones directas** es posible sólo para muy pocos objetivos que pueden aparecer espontáneamente en la vida de la escuela. Vale en general para aquellos objetivos que apuntan a acciones cotidianas: leer, hablar, escribir. Pero aún en esos casos, si nos interesa observar la actuación en algunos aspectos más elaborados, como la lectura con un vocabulario específico o la escritura de palabras no comunes, es probable que tengamos que provocar una situación determinada para poder observar ese rendimiento. Debemos poner a prueba la acción en cuestión, comprobar si aparece la conducta que estamos buscando.

Frente a estos dos tipos de situaciones debemos buscar los instrumentos que nos permitan una observación más confiable. En la situación de prueba, puede quedar un producto concreto: las respuestas a un cuestionario, un objeto elaborado, un registro grabado, que son susceptibles de análisis. En las situaciones espontáneas, es necesario recurrir a instrumentos que permitan sistematizar las observaciones, ajustarlas y registrarlas. La observación está viciada de alguna manera, por los factores distorsionantes propios del observador -actitudes, conocimientos, motivación- y de la situación. El instrumento disminuye la distorsión.

Llamamos pruebas, en general, al registro obtenido en el primer tipo de situaciones, las provocadas.

Para los otros casos empleamos registros sistemáticos, listas de control y escalas.

### ◆ Calidad de los instrumentos

Los valores de la evaluación, por supuesto, van a estar determinados, en gran medida, por la calidad de los instrumentos que se usen. El control de calidad del instrumento es una tarea que no siempre se realiza y que sin embargo puede ser básica para tomar decisiones a partir de los resultados. Nadie se atrevería a comprar una tela, sin estar seguro de que las medidas que tomó son exactas. Si duda de la exactitud de su metro, va a probar con otro para ver si las mediciones coinciden. En la enseñanza, muchas veces, no podemos repetir la medición pero, por lo menos, podemos "suspender el juicio", hasta asegurarnos de que el instrumento es bueno. Antes de aceptar los resultados debemos juzgar hasta que punto el instrumento posee objetividad, discriminación, confiabilidad y validez.

Algunas de estas características pueden averiguarse analizando el instrumento aun antes de aplicarlo, pero para otros es necesario revisar los resultados que se obtuvieron. La mayoría de estos análisis valen para cualquier tipo de prueba, pero son más fáciles de realizar y más seguros para las pruebas de selección. También son más significativos cuando se han aplicado a un grupo de sujetos; pero la información es igualmente útil, aunque no tan digna de confianza, cuando se trabajan con diez o quince alumnos.

**Objetividad:** una prueba es más objetiva cuanto mayor pueda ser el acuerdo que logran los expertos, al juzgar las respuestas. La objetividad no es intrínseca al tipo de prueba sino que depende del contenido al cual apunta. La respuesta, propuesta como correcta para un ítem de selección múltiple, podría no ser elegida por todos los especialistas; por lo tanto, podríamos decir que no tiene objetividad.

Ciertamente, la falta de objetividad es mayor cuanto más grande es la cantidad de elementos en juego en la respuesta. Es más fácil lograr objetividad frente a una respuesta de selección múltiple, respuesta corta, donde sólo interesa el significado, que lograrla frente a una respuesta elaborada (donde se puede juzgar no sólo qué dice, sino también la forma en que está dicho) o frente a una ejecución compleja (donde se debe observar simultáneamente aspectos muy diversos).

A veces la falta de objetividad puede provenir de un planteamiento inadecuado de la pregunta o la tarea: el autor de la prueba puede saber muy bien lo que quiere pedir, pero no expresarlo con exactitud. Suele ser una causa muy frecuente. Es fácil remediarlo. Conviene que el autor de una prueba lo haga leer, antes de aplicarla, por otras personas que entiendan el tema, para ver si ellos interpretan las preguntas o consignas en la misma forma que él las pensó. Estas personas no deben conocer de antemano la respuesta elegida, porque ese conocimiento, podría encausar la respuesta y llevarlos a pensar como

pensó el autor, sin percibir posibles distorsiones. La distorsión puede estar en la forma de expresar las ideas o en el vocabulario; pero a menudo, aparecen problemas porque no se delimita bien el campo de la pregunta, hay contextos demasiado amplios que dan lugar a ambigüedades o vaguedades.

Otras veces, la falta de objetividad proviene de discrepancias en la interpretación de los conceptos subyacentes. Cuando éste es el caso, conviene estipular el marco teórico, para ajustar el análisis. Siempre que haya desacuerdo entre los expertos que leen la prueba es conveniente revisarla, porque con seguridad hay alguna deficiencia.

**Dificultad:** la dificultad de una prueba se establece en función de la cantidad de sujetos que la respondieron correctamente. Tradicionalmente los tratados sobre evaluación dedicaban mucha atención a este punto y, generalmente, aparecían planteamientos estadísticos para determinar cuál era la dificultad esperable en una prueba y cómo programarlo. Con el énfasis puesto en la evaluación del rendimiento en función de los objetivos y como retroalimentación del proceso de enseñanza, el tema del nivel de dificultad de una prueba tiene menos peso y características distintas.

La calidad de una prueba no está determinada por su dificultad, salvo en casos extremos. Si nadie la puede contestar correctamente, la prueba no es buena en el sentido de que no sirve para proveer información acerca del aprendizaje. Si todos lo pueden resolver, es muy poca la información que da y, probablemente, se hubiera podido evitar, porque el docente ya debería tener indicios de que ese aprendizaje estaba superado. En algunos casos, muy pocos, se justifican pruebas así.

Fuera de estas situaciones límites, el problema de la dificultad no tiene relevancia para la construcción de la prueba. En tanto debe ajustarse al objetivo, hacerla más fácil o más difícil, implicaría variar el objetivo. La prueba, en cambio, debe reflejarlo lo más exactamente posible.

Una vez administrada se puede averiguar el nivel de dificultad de cada ítem, como una señal de rendimiento. Este dato sí resulta significativo, junto con todas las otras características que estamos analizando. Si un ítem resultó más difícil que otros, y su calidad se comprueba de acuerdo con los otros criterios, debemos averiguar qué problemas ha habido en ese aprendizaje. Es un llamado de atención para revisar enseñanza y aprendizaje, o para un diagnóstico más fino.

**Discriminación:** la discriminación es el poder que tienen los ítem para diferenciar entre los alumnos "buenos" (o que han llegado a dominar el objetivo) y los alumnos "malos" (que no han llegado).

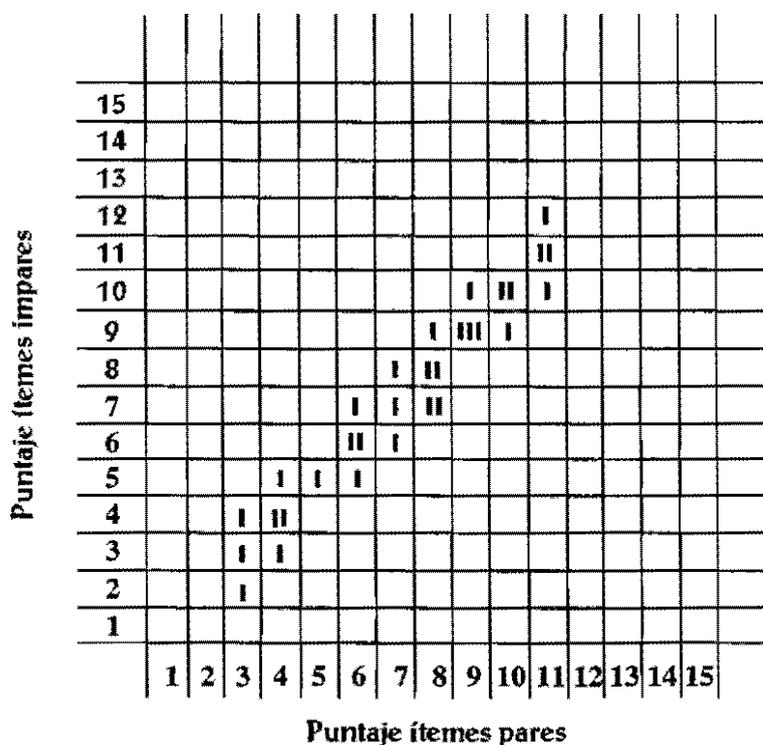
Idealmente si todos los ítem discriminaron deberían ser contestados correctamente por los alumnos que han alcanzado el objetivo, y no serían resueltos por los que no lo lograron. Esta situación extrema es difícil de encontrar, pero a menudo se hallan situaciones semejantes. El poder de discriminación sólo se puede averiguar una vez que la prueba ya ha sido tomada, y se tienen los resultados.

El análisis se realiza a partir del supuesto de que el rendimiento en la prueba como totalidad, refleja exactamente el grado de dominio del aprendizaje en aquello que la prueba pretende medir. Se trata entonces de comparar los resultados de cada ítem en cada uno de los dos grupos, los de alto y bajo rendimiento en la prueba. Se pueden ordenar las pruebas, de acuerdo con los resultados, de mayor a menor; separar en dos mitades y comparar, para cada ítem, la cantidad de sujetos del grupo alto que lo contestó bien, con la cantidad del grupo bajo que lo contestó bien. Si esta diferencia es significativa, en favor del grupo alto, el ítem discrimina. Esto significa que el ítem es homogéneo en relación con el resto de la prueba, y no ha habido ningún problema ni en la elaboración ni en la aplicación que lo haya deteriorado. Vamos a ejemplificar este proceso cuando veamos el análisis de los resultados.

**Confiabilidad:** la confiabilidad es la precisión con que un instrumento realiza una medición.

Se refiere a la calidad de la medición. Poco valor pueden tener los resultados si se han utilizado instrumentos poco confiables. Nadie usaría a sabiendas un metro elástico para medir una longitud. En el tipo de mediciones físicas, es más o menos simple averiguar si el instrumento es confiable; si, repitiendo las mediciones, no obtenemos diferencias significativas podemos pensar que el instrumento es confiable. Este procedimiento no puede proponerse para las pruebas educacionales. No podemos someter a los alumnos a ese tratamiento.

Los tratados de medición proponen varios métodos para comprobar el grado de confiabilidad del instrumento usado: aplicar dos veces la misma prueba y comparar los resultados, aplicar dos formas equivalentes de pruebas, o dividir la prueba en dos partes de la misma extensión. En todos estos casos, la confiabilidad va a estar dada por la correlación entre ambos grupos de puntajes. Estas formas son un poco complicadas y no es posible pedir a un docente que aplique estos procedimientos cada vez que toma una prueba. Es más, podrían llegar a ser perniciosas dentro del proceso de enseñanza las dos primeras formas. La última, compara las dos mitades de una prueba, se puede realizar sin grandes dificultades. (Las fórmulas para trabajarlas se pueden encontrar en cualquier tratado). Una manera práctica de realizarlo, que parece suficiente para las situaciones comunes de clase, es llevar los datos a un diagrama de dispersión que permita visualizar el valor de correlación entre las dos mitades de la prueba. Para obtener el diagrama, se divide la prueba en partes iguales tomando, por ejemplo, los ítem pares y los impares para lograr dos formas lo más homogéneas posibles. Se halla el puntaje de cada alumno para cada una de las partes de la prueba y se traslada al gráfico. Cada marca representa el puntaje del mismo alumno para cada una de las partes, leídos en los ejes horizontal y vertical.\*



\* Cada una de las marcas identifica el puntaje de un alumno representado por la cantidad de ítemes respondidos en forma correcta para cada una de las mitades de la prueba.

Pero estos procedimientos técnicos no están justificados para la mayoría de las situaciones de pruebas escolares; toman valor para situaciones límites, por ejemplo, para una prueba integradora que pueda tener consecuencias para la promoción. En ese caso, conviene asegurarnos de que el instrumento usado tiene una confiabilidad adecuada. Usar los puntajes de las pruebas como medidas exactas para tomar decisiones, sin averiguar la confiabilidad de los resultados, representa un acto de fe, sin bases racionales.

Esto no significa que en todas las otras situaciones trabajemos con instrumentos no confiables, sino que no es necesario averiguar exactamente cuál es la confiabilidad. Debemos tomar ciertos recaudos para hacer que nuestras pruebas sean lo más exactas posibles.

- 1) **Objetividad.** Cuanto más objetiva sea una prueba, en el sentido que lo explicamos antes, más confiable será.
- 2) **Extensión.** Una prueba más extensa es más confiable que una más corta; hay mayor probabilidad de que las respuestas no sean por azar.
- 3) **Corrección rigurosa.** Siempre se debe tratar de reducir al mínimo los errores del lector o corrector; las pruebas que no tienen respuesta fija, es conveniente hacerlas controlar por más de una persona.

**Validez:** por validez se entiende la exactitud con que una prueba mide lo que pretende medir. Es una condición de la prueba que sólo se puede cumplir con relación a un criterio externo.

En la enseñanza, se trata de analizar hasta qué punto el instrumento representa al objetivo que se desea medir. Es, sin lugar a dudas, la condición fundamental cuando se trata de elegir, o elaborar una prueba. De nada serviría poseer un instrumento objetivo, discriminante, confiable, si no mide lo que intentamos medir. Sería como poseer un excelente termómetro cuando deseamos medir la presión atmosférica. Dicho en estos términos parece un despropósito. Pero lo que ocurre es que en ese campo sabemos que existen barómetros y termómetros y qué podemos obtener utilizándolos en las condiciones adecuadas. Si estuviéramos en un lugar donde no existen y tuviéramos que construirlos, tendríamos que asegurarnos de que las variaciones que el instrumento acusa se deben realmente a cambios en la presión o en la temperatura, y no a otros factores. Este es el problema de la validez.

Se trata de lograr certeza de que los datos que ofrece una prueba corresponden a la variable que se desea estudiar. Esta preocupación podría parecer infundada, pero, muchas veces, los rasgos que se desean evaluar no están definidos claramente, y no es fácil discernir si las conductas observables corresponden o no a esos rasgos.

Se han ideado diversas formas de determinar la validez a partir de dos posiciones diferentes: 1) buscar que la prueba represente fielmente las conductas propuestas -validez directa- y; 2) relacionar los puntajes con otros que ya tienen validez probada -validez derivada-.

La validez derivada es muy difícil de buscar en la mayoría de las pruebas escolares. Las dos formas más comunes son: la concurrente (correlación entre los puntajes de la prueba y otros ya conocidos en ese momento) y la predictiva (correlación entre los puntajes de la prueba y posteriores medidas de rendimiento).

La validez concurrente es muy difícil de plantear con pruebas de rendimiento, porque habitualmente no se dispone de otras medidas ya validadas para las mismas conductas. La predictiva podría llegar a

trabajarse, especialmente para pruebas de selección, pero no sería tarea de los docentes, sino de equipos especializados.

Las pruebas de rendimiento exigen una validez directa. Tradicionalmente se dice que estas pruebas deben poseer validez curricular. Se juzga hasta qué punto es representativa de los objetivos y materiales del curso. La única manera posible de buscar esta validez es lograr una definición operacional precisa de los objetivos que se desean evaluar. Dice Ebel: "Muchos de los problemas de validación que causan perplejidad a los autores de pruebas son en realidad problemas de definición operacional del rendimiento o rasgos que se quiere evaluar".

Si se acepta que la prueba mide un conjunto de tareas y hay acuerdo sobre las tareas que se quieren medir, la validez deja de ser problema. Probablemente el problema se traslada de la dificultad para la validación, a la dificultad para enunciar los objetivos. Un objetivo enunciado operacionalmente permite determinar las tareas que deben formar parte de la prueba. A veces ésta es un muestreo de las tareas posibles, cuando los objetivos a evaluar son muchos, por ejemplo cuando se evalúa con fines integradores. En ese caso la prueba debe estar formada por una muestra representativa del universo posible a evaluar.

**La tabla de especificaciones es un recurso apto para proyectar esa muestra. Ajustar una prueba a la tabla de especificaciones es la forma de asegurarle validez curricular.**

Aunque hemos analizado por separado las condiciones que deben reunir los instrumentos, debemos destacar las relaciones entre ellas. En nuestra definición de validez, hasta ahora, sólo hemos puesto énfasis en la relación con los objetivos, pero dijimos que una prueba válida mide con exactitud. Es decir que una prueba para ser válida, debe ser confiable. Inútil sería que pretendiera medir el objetivo, si en realidad no midiera nada. La confiabilidad es una condición para la validez. La inversa no es cierta. La validez no es condición para la confiabilidad. Un instrumento puede ser preciso para medir una cosa distinta de aquella que necesitamos.

#### ◆ Pruebas de rendimiento

Vamos a analizar las pruebas. La prueba en el sentido más amplio es el instrumento más utilizado en la situación escolar. Permite un mayor control de la situación para que no aparezcan variables difíciles de integrar o que no formen parte de la situación pero que puedan alterarla. Por ejemplo, si se deseara que el alumno fuera capaz de leer un gráfico y se esperara que apareciera en un periódico para ver si puede hacerlo, no se podría controlar la mayor o menor dificultad del gráfico (por ejemplo el número de variables en juego), la información que pudiera agregar el artículo periodístico y que facilitara o dificultara la lectura, etc. Aun cuando nuestro objetivo pidiera expresamente que el alumno explicara verbalmente las ideas fundamentales del gráfico, con la situación apuntada no sabríamos exactamente cómo puede hacerlo en una situación típica. Análisis semejante se pueden hacer para la mayoría de las situaciones de observación directa.

La prueba, en cambio, si está bien planeada permite la organización de la situación controlando todos los elementos que resultan importantes para nuestro objetivo: situación de tiempo y espacio, contenido, actuación propuesta, materiales.

Tradicionalmente se conocen dos variables fundamentales de prueba: verbales, tanto orales como escritas, y de ejecución o de desempeño, donde el sujeto responde mediante una acción manifiesta, básicamente motora.

Pruebas verbales son aquéllas donde el lenguaje tiene un papel protagónico, juega como mediatizador de otras conductas. De alguna manera, el uso de las pruebas verbales para objetivos distintos al uso del lenguaje parece apoyarse en el supuesto de que todo aprendizaje es verbalizable y verbalizar es demostrar que se ha aprendido.

Los conocimientos o información de cualquier tipo se muestran a través de la posibilidad de expresarlos. Poseer conocimientos significa poder comunicarlos. El hecho de tenerlos almacenados pero no poder usarlos es a los fines del aprendizaje semejante a no poseerlos.

Las habilidades intelectuales implican la posibilidad de manejarse con símbolos. Las habilidades intelectuales se pueden manifestar también a través del lenguaje. Con ciertos recaudos, las pruebas verbales pueden ser aptas para evaluar las habilidades intelectuales.

Pero para otros tipos de productos de aprendizaje, tales como actitudes y destrezas, las pruebas verbales sólo pueden dar aproximaciones lingüísticas que no agotan ni reemplazan a los resultados buscados. Supongamos que nos propusiésemos como objetivo que nuestros alumnos gustaran de la música sinfónica y al cabo de cierto tiempo los enfrentáramos a una prueba donde le preguntáramos por ejemplo, ¿qué momentos le gustaron más de la Sinfonía N° 5 de Beethoven? ¿Por qué? Frente a la mejor de las respuestas podemos dudar de si el alumno escuchó la sinfonía o si solamente oyó o leyó un comentario sobre la obra. Su actuación nos muestra información verbal pero no apreciación de la música. No es difícil encontrar situaciones como ésta en la escuela. El peligro se hubiera evitado si antes de armar la prueba se hubiera definido operacionalmente el objetivo. Inmediatamente habría surgido la discrepancia entre el objetivo y la conducta propuesta. Se habría visto la necesidad de una prueba de desempeño: por ejemplo, marcar, mientras se escucha la sinfonía, los momentos sobresalientes o regrabar sólo esos momentos y pasarlos para que los escuchen los compañeros.

Las pruebas deben tratar de ser una muestra de la tarea pedida en el objetivo. Cuando esto no puede lograrse en la escuela, entonces se recurre a situaciones simuladas. La conducta obtenida en esa situación se supone semejante a la que se daría en la situación real.

Toda prueba implica, evidentemente, pedir una ejecución. Redactar es una ejecución también lo es embocar en el cesto, pero, al contestar preguntas o resolver un problema matemático, la ejecución explícita de escribir no es la más importante sino que, en realidad, buscamos los procesos de pensamiento implícito en esa actuación. Esta diferencia es la que se quiere marcar cuando se habla de pruebas verbales y de ejecución.

Las pruebas deben reflejar de la manera más directa posible las ejecuciones propuestas en la definición del objetivo. Si tuviéramos clara esta norma no sería necesario hacer ninguna clasificación de pruebas, se desprenderían de los tipos de objetivos. Pero, como no siempre se conocen esas maneras directas y como el uso ha establecido ciertos tipos de pruebas, vamos a revisarlos para discutir sus valores con respecto a los objetivos y los criterios para su utilización.

Las pruebas verbales pueden ser orales y escritas.

La prueba oral ha sido tradicionalmente el método de examen más usado y todavía cuenta con gran

prestigio, especialmente en los niveles de enseñanza superior. Como cualquiera de los instrumentos, la prueba oral puede estar justificada cuando resulta una muestra de la tarea señalada en el objetivo. Frente al objetivo de expresarse oralmente en una lengua extranjera o el de manejar recursos de comunicación oral, por ejemplo para un docente, las pruebas orales no pueden ser sustituidas por otro instrumento. Tratar de medir esos objetivos con pruebas escritas sería distorsionar la actuación. Pero no está justificada su uso cuando se busca apreciar otros objetivos, como el manejo de conocimientos o las habilidades intelectuales.

La prueba oral es, en sí misma, poco confiable. Es difícil controlar la situación y poder hacer un análisis y una interpretación adecuada de sus resultados, puesto que es difícil el registro en el momento en el que se efectúa la prueba, salvo que se utilicen recursos especiales tales como grabación en audio o videocassette (tal como se hace en la microenseñanza).

Un problema especial que surge con las pruebas orales es que generalmente se tratan de usar para más de un objetivo simultáneamente; por ejemplo, ver como el alumno "expresa" "los conocimientos". La falta de confiabilidad se agrava frente a esta situación. El alumno no sabe bien qué se espera de él y el docente no sabe dónde focalizar su observación. Llaman su atención tanto los errores de expresión como los problemas de conocimiento y no puede atender bien a los logros en ninguno de los dos campos.

En los casos en que sea necesario recurrir a las pruebas orales, debe elaborarse algún patrón de análisis que permita la observación focalizada y, en lo posible, tratar de lograr un registro de la situación para un posterior análisis más fino.

Las pruebas escritas son sin duda el instrumento más usado en la escuela. Una razón para esto es que uno de los fines primordiales de la escuela es formar un hombre pensante. Pensamiento y lenguaje pueden ser evaluados a través de pruebas escritas. Pero a veces se hace una extensión indebida de este uso y se utilizan pruebas escritas en reemplazo de otro tipo de ejecuciones. Pedir a un futuro docente que describa como enseñaría a sumar dígitos no nos dice nada acerca de como actuará realmente cuando deba enseñar. Los informes de experimentos científicos que los alumnos han visto no nos informan acerca de su capacidad para moverse en un laboratorio.

Hay dos tipos básicos de pruebas escritas: las que piden elaboración de las respuestas y la que proveen la respuesta que es necesario elegir. Estas últimas son las que habitualmente se conocen con el nombre de **pruebas objetivas**. Este nombre requeriría el opuesto de pruebas subjetivas. Pero en realidad, la objetividad es una característica que debería tener todas las pruebas. Las llamadas pruebas objetivas resultan así porque, dada la forma de construcción, deben tener una sola posibilidad de respuesta correcta. Esto hace que su corrección sea más objetiva que la de aquellas que piden que el examinado elabore su respuesta. En estas últimas aparece una variedad de respuestas susceptibles de ser juzgadas de manera diversa. La diferencia básica está entonces en la forma de la respuesta que se pide a los alumnos. En un caso sólo debe reconocerlo o seleccionarla; las llamaremos **pruebas de selección** o de respuesta fija. En el otro, debe buscar, entre todo el bagaje de información almacenada, aquella que sirve para la pregunta o el problema y debe darle forma adecuada para que los demás comprendan su expresión; las llamaremos **pruebas de elaboración**. Es evidente que se piden capacidades distintas en cada uno de estos tipos de prueba. Antes de elegir uno u otro es necesario ponderar sus ventajas en función del objetivo a evaluar.

## ◆ Pruebas de selección

Las pruebas de selección son de uso más reciente que las de elaboración. Surgieron como una solución para resolver los graves problemas de confiabilidad que éstas presentaban. Es uno de los medios de la enseñanza que mayor cantidad de reacciones afectivas ha despertado, tanto entre especialistas como en la sociedad en general. Tienen sus grandes amantes y sus grandes detractores. Se han convertido en artículos de fe. Trataremos de encontrar las pautas racionales para poder actuar entre los que "creen" o "no creen" en las pruebas objetivas.

Vamos a analizar cada uno de los tipos de ítems (\*) que pueden formar parte de una prueba de selección. La más conocida y utilizada es la llamada de **selección múltiple**. Consta de un cuerpo que puede ser una pregunta o un enunciado incompleto, que se debe responder o completar con alguno de los cuatro o cinco enunciados que siguen. Por ejemplo:

*La mayoría de las pruebas elaboradas por los maestros ponen énfasis en la información pura y dejan de lado los otros objetivos del curso para los cuales están destinadas. En términos técnicos les falta.*

- a) validez
- b) discriminación
- c) objetividad
- d) confiabilidad

Analicemos este ítem. El enunciado plantea una situación en la cual se pone en juego la relación entre objetivos y pruebas, estableciendo una falta de acuerdo. Lleva a pensar cuál o cuáles de las características técnicas apunta esa relación. Si el examinado sabe que la confiabilidad es una característica del instrumento en sí mismo, la objetividad tiene que ver con el acuerdo en la corrección o puntuación, la discriminación es la posibilidad de diferenciación entre los sujetos que responden, sólo la "validez" puede ser la respuesta correcta.

Un proceso de este tipo es el que se pone en juego frente o a un ítem de selección múltiple. Se parte de un análisis del significado del enunciado general, que debe guiar la búsqueda de la respuesta correcta. Ese enunciado debe ser lo más completo posible para que de su análisis se obtengan los elementos para una anticipación de la respuesta. En este caso, la respuesta correcta debería referirse a la relación entre pruebas y objetivos. Si están claros los cuatro conceptos propuestos, la respuesta se diferencia claramente.

Hay varias acciones implicadas en esa respuesta: la comprensión del enunciado, la información sobre las características y la posibilidad de discriminar entre ellas. En todo ítem de selección múltiple se piden procesos semejantes.

Siempre que nuestros objetivos nos lleven a buscar estos tipos de procesos se pueden usar ítems de selección múltiple. Este ítem se podría haber elaborado como de completamiento, por ejemplo, sin proveer las opciones. En ese caso el sujeto debería buscar en su memoria cuál es la característica técnica

---

(\*) Ítem es la unidad mínima de una prueba que requiere una respuesta. Cada blanco que se debe llenar es un ítem, cada propuesta que requiere una elección es un ítem.

que se refiere a la relación prueba-objetivo. Sería posible que recordara perfectamente la diferencia entre validez y confiabilidad y escribiera inmediatamente la respuesta correcta. Podría ocurrir que la conociera pero no pudiera recordar el nombre, por una falta de fijación de las denominaciones; en ese caso, la presentación de los nombres sería una ayuda. También sería posible que la presentación de los cuatro nombres le hiciera notar que no puede discriminar exactamente entre los cuatro conceptos y surgieran dudas, que no hubieran surgido ante la pregunta directa.

La intención de este análisis es que seamos conscientes de los procesos que podemos desencadenar ante la presentación de ítemes que estimulan memoria, reconocimiento, búsqueda, ajuste de conceptualización, etc.

Para que un ítem de selección pueda cumplir estas funciones es necesario que esté bien construido. La calidad de un ítem de selección múltiple depende de dos características fundamentales: 1) la claridad y complejidad del cuerpo del ítem; 2) la homogeneidad de las opciones. Estas dos características se relacionan directamente con los procesos que se desean medir; de ahí su importancia. Las otras indicaciones que se suelen dar para mejorar la calidad de los ítemes son válidas para todos los tipos: claridad en la redacción, adecuación del vocabulario, pertinencia del tema, etc.

1. El cuerpo del ítem debe llevar a pensar en una respuesta posible a todo aquel que tenga la información necesaria. Los enunciados amplios o abiertos tales como "La educación propone..." "La docencia implica..." no sirven para focalizar un campo de información ni para orientar una búsqueda en la estructura cognoscitiva del sujeto. La calidad del enunciado no depende de la extensión, sino del uso de palabras específicas que delimiten un campo. Generalmente esta precisión hace que las opciones sean breves, puesto que la fundamental del contenido ya está en el enunciado.
2. La homogeneidad de las opciones pide que todas las respuestas propuestas pertenezcan a la misma "clase" de cosas. En el ítem propuesto como ejemplo, las cuatro opciones son características técnicas de las pruebas en relación con algún otro elemento del proceso de evaluación. Si entre las opciones hubieran aparecido términos como calificación o promoción, el ítem hubiera exigido menor poder de discriminación.

La homogeneidad de las opciones se logra tratando de buscar una base estricta de clasificación como serían, en nuestro ejemplo, las características técnicas de la evaluación. Si hubiéramos querido hacer un ítem para verificar si los alumnos poseen el concepto de calificación, las opciones podrían haber sido calificación, promoción, evaluación, medición. Cualquier criterio de clasificación de los conceptos que aparecen en las opciones puede ayudar a lograr la homogeneidad.

Cuando no es posible lograr en la construcción esa homogeneidad, conviene repensar si este tipo de ítem es el más adecuado para nuestros objetivos. Probablemente se esté tratando de forzar un proceso que no corresponde a lo que queremos evaluar.

No siempre es fácil encontrar cuatro o cinco opciones con sentido. Esa cantidad se aconseja para poder contrarrestar la incidencia del azar, pero cuando no se pueden hallar buenos distractores (distractores son las opciones no correctas) es preferible poner menos, si estamos convencidos de que la pregunta es conveniente. Es más honesto frente al alumno poner menos opciones que ponerlas sin valor. Cuando todos los enunciados tienen sentido, el examinado tiende a tratarlos con más seriedad que cuando puede eliminar alguno a priori, porque obviamente no es la respuesta correcta.

Todas estas consideraciones deben llevarnos a pensar que la construcción de ítems de selección múltiples es difícil, que requiere tiempo y habilidad. Una prueba de este tipo mal elaborada deteriora el proceso de evaluación, porque se convierte en un burdo juego de adivinanzas. Una buena prueba, en cambio, interesa a los alumnos, los desafía, no los impulsa a la adivinación ni a la copia. Cualquier docente puede llegar a construir pruebas si está dispuesto a dedicar tiempo a esa tarea, o revisarlas y hacerlas revisar por expertos antes de administrarlas.

A veces, cuando no podemos encontrar suficiente cantidad de opciones o cuando las opciones son muchas, se puede pensar en elaborar ítems de **correspondencia** que constituyen un tipo especial de selección múltiple. La selección múltiple plantea un problema seguido de varias soluciones posibles. La correspondencia plantea varios problemas y un conjunto de opciones que pueden ser válidas para todos los problemas. Es decir, son varias preguntas de selección múltiple, pero con el mismo conjunto de opciones. Suelen ser más fáciles de construir que las anteriores. Probablemente resultan más fáciles de contestar, porque proveen más indicios para la elección. Veamos un ejemplo.

*Cuando se evalúa para controlar los logros inmediatos del aprendizaje antes de continuar con la enseñanza, se está usando la evaluación como*

- a) diagnóstico*
- b) formativo*
- c) integradora*

Sólo pueden aparecer tres opciones válidas, porque no hemos hablado más que de tres funciones de la evaluación en relación con la enseñanza. La posibilidad de acertarla por azar es grande. El examinado puede dudar sobre la respuesta correcta y tener dificultad en identificar la situación como formativa.

Si pensamos en otras situaciones de evaluación y las hacemos corresponder con las funciones es probable que, al ver todo el panorama junto, le resulte más fácil diferenciar las funciones. Por ejemplo:

- |   |                            |
|---|----------------------------|
| 1) Controlar los logros inmediatos del aprendizaje.     |                            |
| 2) Ubicar a los alumnos de acuerdo con las dificultades | a) Evaluación diagnóstica. |
|   | b) Evaluación formativa.   |
| 3) Determinar el nivel de los alumnos.                  | c) Evaluación integradora. |
| 4) Seleccionar alumnos para una determinada tarea.      |                            |
| 5) Promover a los alumnos.                              |                            |
| 6) Realimentar el proceso de aprendizaje.               |                            |

En otro marco, es más fácil identificar las funciones; al mismo tiempo, hay menor probabilidad de que se resuelva por azar. Se trata, como en la selección múltiple, de lograr homogeneidad; en este caso, dentro de cada una de las listas. Es necesario cuidar de que todas las situaciones tengan opciones. En este ejemplo, cada una de las funciones puede usarse como respuesta todas las veces que sea necesario; por lo tanto, ninguna respuesta está determinada. Esta forma puede llamarse de **clasificación**. Podríamos ampliar la lista de la izquierda tanto como quisiéramos, para asegurarnos la comprensión.

Otras formas de correspondencia pueden pedir formas pares, donde cada uno de los elementos de la lista de opciones se use una sola vez. En ese caso, conviene que ambas listas no tengan el mismo número de elementos para evitar la respuesta forzada, la que resulta por eliminación.

Las formas de clasificación resultan aptas para una gran variedad de objetivos, para todos aquellos que representan el manejo de conceptos. Cada ítem implica pedir al sujeto que señale los casos que pertenecen a un concepto diferenciándolo de otros semejantes.

Otro tipo de ítem de reconocimiento es el de **doble alternativa**, conocido también como de verdadero (V) o falso (F). La tarea que se pide es juzgar la verdad o falsedad de un enunciado.

Consideramos que un objetivo importante de la escuela es que los alumnos aprendan información verbal. La información verbal se expresa en forma de proposiciones. Poder distinguir entre las proposiciones verdaderas o falsas indica dominio de la información verbal. Para este tipo de objetivos está plenamente justificado el ítem de doble alternativa.

Muchos docentes rechazan este tipo de ítem con el argumento de que presentar un error a los alumnos puede guiarlos a fijar ese error. Este argumento es insostenible por que en esta clase de ítemes no se presentan errores como si fueran verdades, sino que se presentan dentro de un contexto de duda, con una postura crítica y desafiante. El inducir a error podría llegar a darse cuando se usara la autoridad del docente para postular algunas posiciones incorrectas como si fueran ciertas. Por el contrario, tratar de que el alumno dude, frente a afirmaciones que pueden tener el aspecto de verdaderas, puede ser un objetivo valioso. Claro que no cualquier enunciado puede ser sometido a este juicio. Es preciso seleccionar muy bien cuáles pueden ser aptos para esta forma. La proposición debe ser indudablemente verdadera o falsa para cualquier persona experta en el tema.

Otra acusación habitual, que reciben los ítemes de doble alternativa, es que apuntan a hechos aislados y muy específicos, que resultan poco importantes. Esta acusación puede ser válida también para las otras formas de ítemes que hemos visto, si no se cuida de atender a las especificaciones de la tabla.

Pero en el caso que estamos analizando es más probable que se caiga en ese error porque, por su forma, se recomienda elaborar una gran cantidad para cada prueba y es entonces más fácil caer en detalles sin importancia y muy aislados.

Todos estos argumentos, más la mayor incidencia del azar que puede darse en cada ítem de doble alternativa, hace que los docentes no utilicen suficientemente este tipo de ítem, y muchas veces se esfuerzan por elaborar ítemes de selección múltiple cuando sería más adecuado uno de otro tipo. Por ejemplo:

¿Para medir cuál de los siguientes resultados es menos adecuada la prueba de respuesta fija?

- a) aplicación del conocimiento
- b) organización de pensamiento
- c) reconocimiento de información
- d) comprensión de la información

La respuesta correcta sería la "b", pero es evidente que las opciones no son homogéneas y, que, además, es difícil responder a preguntas planteadas por la negativa. Con ese mismo material, se podrían elaborar ítems de alternativa de mayor calidad. Por ejemplo:

- |    |   |                         |                         |
|----|---|-------------------------|-------------------------|
| 1. | La prueba de respuesta fija, por lo común, es adecuada para medir la organización del pensamiento ..... | <input type="radio"/> V | <input type="radio"/> F |
| 2. | La prueba de respuesta fija no es adecuada para medir comprensión de la información .....               | <input type="radio"/> V | <input type="radio"/> F |

Otro ejemplo: si deseamos elaborar un ítem de selección múltiple para evaluar si los alumnos manejan claramente el concepto de enseñanza y pueden establecer relaciones con el de educación, nos resulta muy difícil encontrar opciones homogéneas, porque se caracterizan por atributos heterogéneos. Algunas proposiciones que definan a ambos procesos y cuya verdad o falsedad se juzguen, pueden darnos una evaluación más fiel de nuestro objetivo. Se podrían seleccionar algunos de estos enunciados:

- |    |   |                         |                         |
|----|---|-------------------------|-------------------------|
| 1. | Un proceso puede ser enseñanza y educación a la vez .....         | <input type="radio"/> V | <input type="radio"/> F |
| 2. | Enseñar y educar son partes del mismo proceso .....               | <input type="radio"/> V | <input type="radio"/> F |
| 3. | Cuando alguien educa, enseña, el que enseña, educa .....          | <input type="radio"/> V | <input type="radio"/> F |
| 4. | La enseñanza requiere métodos eficaces, la educación, valiosos .. | <input type="radio"/> V | <input type="radio"/> F |

Estos y otros enunciados posibles a partir de los conceptos de enseñanza y educación pueden permitir un buen análisis del ajuste de estos conceptos con poco esfuerzo y tiempo. No son difíciles de elaborar y exigen muy poco tiempo para responder. Cuando se hacen varios ítems para el mismo objetivo, es muy difícil que el resultado se vea afectado notablemente por el azar. Analizando las respuestas se puede llegar a ver si el alumno domina el objetivo o si sus respuestas fueron aleatorias.

Para esta forma, más todavía que para las otras, hay que cuidar los enunciados para evitar que repitan una sucesión de palabras conocidas, que el alumno puede identificar como tal, sin conocer completamente el significado. También hay que cuidar que los enunciados no sean evidentes por sí mismos, que alguien que no maneje los conceptos subyacentes pueda contestarlos por sentido común.

Llamamos prueba de elaboración a aquéllas en que se pide al estudiante que organice por sí mismo la respuesta y la exprese con sus propias palabras. Da al alumno más libertad en sus respuestas que las formas que analizamos anteriormente.

La forma de la pregunta puede poner más o menos límites a esa libertad, pero siempre es mucho mayor que marcar la respuesta. Es necesario elegirla, buscarla en la memoria y luego redactarla. El valor de

este proceso dependerá de la calidad de la pregunta.

La pregunta puede pedir desde el simple recuerdo de hecho o nombres, hasta la solución de un problema con la elaboración de un producto original. Al contrario de las otras formas, donde la pregunta exigía un determinado proceso, el único proceso necesariamente presente acá es la redacción de la respuesta. Obviamente, si el objetivo es evaluar la elaboración de respuestas, es ésta la única forma válida. Se han propuesto formas más objetivas para evaluar esas acciones. Pero el producto han sido pruebas sofisticadas, cuyo uso es difícil de extender a todos los docentes y con dificultad para demostrar su validez. Es más fácil y más claro para los alumnos, medir la posibilidad de organizar las ideas por escrito, pidiéndolo directamente, que inferirlo de una cantidad de ítems de selección múltiple a través de los cuales se trata de ver cómo lo harían.

La búsqueda de otras formas para reemplazar a las pruebas de elaboración surge de la indiscutible dificultad de hacer, de estas pruebas, un instrumento confiable. La falta de confiabilidad reside fundamentalmente en 1) la poca representatividad de las preguntas; y, 2) la corrección.

1. A igual tiempo destinado a la evaluación, es posible, obtener más respuesta de los alumnos cuando se utilizan pruebas de respuesta fija, que cuando se piden respuestas abiertas. El tiempo para la redacción es mayor que para la lectura. Hay menores posibilidades entonces de cubrir el campo de los objetivos que se desea evaluar. Cuando el campo no se cubre, se infiere que los resultados obtenidos son semejantes a los que se obtendrían en el resto de las ejecuciones posibles; las conductas actuales representan a las posibles. Esta inferencia exige un análisis exhaustivo de las conductas elegidas que dé garantía de que cada tarea esencial ha sido evaluada. Cuando las respuestas requieren poco tiempo, podemos pedir más de una pregunta para cada ejecución importante, de manera de obtener certeza de que la respuesta no fue por azar o por alguna variable que no controlamos. Con las preguntas abiertas, esto resulta bastante difícil.
2. La causa fundamental por la cual este tipo de prueba recibe las mayores críticas es su dificultad de corrección. Mientras que en las pruebas de selección, una vez fijada la respuesta correcta, no existen problemas de corrección, en estas pruebas hay que discriminar permanentemente qué es lo correcto; cuándo está correcto; las ideas pueden parecer adecuadas como respuesta, pero la forma puede no serlo; o a la inversa, la forma dificulta el análisis del contenido. ¿Qué es lo que se debe tomar en cuenta?

Estas dudas surgen frente a cada una de las respuestas y el examinador debe tener gran cuidado para poder lograr una corrección confiable y defendible. Esta tarea se verá facilitada con dos precauciones que puede adoptar el docente; la elaboración cuidadosa de la pregunta y la preparación previa de una clave de análisis de las respuestas.

La pregunta bien elaborada debe dar un marco para que el examinado fije los límites de su respuesta, tanto en relación con la capacidad que se busque, como en cuanto al contenido.

El alumno debe saber si se espera de él una elaboración personal, o la interpretación o reproducción de lo aprendido, si se quiere que lo escriba en el lenguaje técnico o cotidiano, hasta dónde debe profundizar en el contenido, etc.

Estas delimitaciones permite que el examinado sepa qué se espera de él y pueda mostrar su rendimiento.

La elaboración previa de una clave de corrección permite ajustar la pregunta. Después de armar la clave, es conveniente cotejar la respuesta esperada con la pregunta, para ver si realmente ésta conduce a esa respuesta y hacer las modificaciones necesarias.

Veamos un ejemplo: en un curso para docentes, se han discutido los conceptos de enseñanza y educación. Deseamos averiguar si los alumnos pueden manejarlos. Ítems como los que presentamos para los ejemplos de doble alternativa pueden servir para ver si diferencian ambos conceptos. Pero no garantiza que puedan entender situaciones específicas a partir de ellos. Por eso, parece más válido plantear una pregunta de elaboración para comprobar esa actuación. Podría plantearse una tarea como ésta:

**"Compare los procesos de educación y enseñanza"**

Este pedido puede llevar al alumno a analizar ambos procesos desde todos los puntos de vista posibles. Comparar, por ejemplo, la educación, que puede ser sistemática o asistemática, con la enseñanza que es sistemática. Al elaborar la clave de respuesta, aparece que el interés del evaluador es que se los compare viendo en qué condiciones un mismo proceso puede ser enseñanza y educación. En la formulación propuesta, no estaban claras las bases de análisis para la comparación. Podría mejorarse planteando la pregunta así,

**"¿Cuándo la enseñanza es educativa?"**

Acá está más focalizada la pregunta y por lo tanto la respuesta puede ser más breve y más estricta. Una respuesta podría ser: "la enseñanza puede ser educativa cuando guía el aprendizaje hacia objetivos valiosos y de una manera moralmente inobjetable". La respuesta tipo tiene que incluir los conceptos de: "guía del aprendizaje", "objetivos valiosos" y "forma moralmente inobjetable".

Usada de esta manera la pregunta abierta no aporta mucho más que lo que podría aportar una de selección con el agravante de que esos enunciados podría haberlos aprendido en forma estereotipada. La pregunta abierta facilita el uso del estereotipo. La pregunta de selección se podría elaborar con palabras distintas para apuntar al mismo concepto y evitar la repetición.

Pero habría una forma mejor en la cual se podrían aprovechar las ventajas de la prueba de elaboración: presentar la descripción de una clase, donde se vean objetivos, contenidos y actividades de un grupo dirigido por un docente. A continuación aparecen las preguntas:

1. ¿Considera Ud. que en esta situación descrita se ha dado un proceso de enseñanza?, ¿por qué?

2. ¿Ha tenido lugar allí un proceso educativo?, ¿por qué?

Establezca las relaciones necesarias con el ejemplo para fundamentar su respuesta. Defina los términos técnicos que utiliza.

¿Qué interesaría evaluar?: si el alumno es capaz de identificar las características de esos dos procesos y puede mostrar que los comprende relacionándolos con el ejemplo.

La clave de análisis de la pregunta 2., podría ser ésta:

Alumnos	Proceso educativo	Contenido valioso	Relaciones c/ejemplo	Forma valiosa	Relaciones c/ejemplo

Probablemente esta sea la forma más adecuada para medir este objetivo. Las relaciones que pueda establecer entre el ejemplo y la caracterización de los procesos pueden demostrar claramente la comprensión de los conceptos.

Otras veces, la pregunta abierta no presupone una respuesta que puede ser correcta, frente a otra que no lo es, sino que busca una respuesta original, cuyo valor puede estar en la coherencia interna de la resolución (por supuesto que el resultado también tiene que ser coherente; medios adecuados no pueden guiar a soluciones descabelladas).

Por ejemplo, si planteamos una lista de materiales de enseñanza y una situación de enseñanza con sus objetivos, podemos pedir:

1. ¿En qué orden utilizaría estos materiales para tratar de lograr su objetivo? Fundamente cada paso en la teoría del aprendizaje y de la comunicación que maneja. Especifique las condiciones del curso para el cual lo planifico.

Cada alumno puede producir una secuencia distinta; no es tan importante la secuencia lograda como la trama de relaciones que vaya presentando. Una pregunta de este tipo, en un curso de formación docente, propone una tarea muy semejante a lo que deberán hacer los alumnos cuando planifiquen sus propios cursos. Resultaría una prueba válida para evaluar las habilidades para planificar. Por supuesto, no nos dice nada de como utilizaría realmente los materiales, para eso deberíamos pensar alguna prueba de actuación.

En síntesis, las pruebas abiertas permiten medir no sólo los contenidos "cognitivos" de los sujetos, sino además el tipo de relaciones que puede establecer entre ellos, a través de la expresión verbal.

Dentro de las pruebas de rendimiento, la forma menos utilizada es la **prueba de ejecución**, aunque quizás se la utilice en algunos casos sin demasiada conciencia de su calidad de "prueba". Por ejemplo, las "prácticas docentes", los trabajos en el laboratorio, la elaboración de gráficos o cuadros pueden ser situaciones de prueba. Los profesores de educación física o de música saben que deben pedir una "muestra de tareas" para evaluar el rendimiento pero no siempre pondrían el nombre de prueba a lo que hacen. Entenderlo como situación de prueba puede ser útil porque puede orientar para buscar mayor validez, confiabilidad y objetividad.

Cuando los objetivos apuntan a destrezas motoras o habilidades específicas requieren el uso de pruebas de ejecución; esto significa plantear una situación, real o simulada, en condiciones típicas, donde el alumno deba mostrar su eficiencia en la habilidad.

Es importante la elección de la situación en tanto va a determinar directamente la validez de la prueba. Generalmente la habilidad o destreza que se desea evaluar está enunciada en términos tan generales que hacen difícil planificar bien la situación. Tomemos como ejemplo las prácticas docentes. Al mismo

tiempo que se califica al alumno-docente, se trata de que se ejercite en dirigir a un grupo de alumnos para el aprendizaje de un tema; "dirigir un aprendizaje" es un enunciado tan amplio que difícilmente el alumno o el supervisor conocen claramente qué deben tener en cuenta para manejar o para evaluar. Esa habilidad puede subdividirse en varias habilidades más específicas, como "organizar la situación de aprendizaje", "favorecer la comunicación", "guiar el proceso de aprendizaje", "organizar el trabajo grupal e individual", "utilizar medios de enseñanza". Cada una de estas requiere a su vez, una definición operacional para permitir su evaluación. La situación de desempeño prevista, la clase en este caso, tiene que estar organizada de manera que permita mostrar el dominio en las tareas requeridas: el tema, los objetivos, las condiciones, el tiempo, tienen que adecuarse a esa finalidad.

Las pruebas de ejecución exigen mucho tiempo para ser administradas porque son individuales (a veces se pueden aplicar a pequeños grupos). Si la situación no está bien regulada, además de perder validez, consumen más tiempo para permitir observar lo que se busca.

Otro problema de estas pruebas es la confiabilidad. Puede haber distorsiones, tanto en la falta de explicitación de las condiciones que hagan poco objetiva la evaluación, como en la observación de los resultados. Definiendo claramente las características que debe presentar una buena ejecución se aumenta la confiabilidad.

Podemos ayudarnos para esto con el uso de instrumentos que guíen la observación y faciliten el registro de lo observado. Las listas de control y las escalas cumplen esta función. Las **listas de control** representan un esquema para registrar la secuencia de actos o el empleo de técnicas aceptadas como valiosas para esa situación. Las **escalas** exigen apreciar analíticamente las etapas o aspectos de la tarea. Pueden tener diversas formas pero siempre se pide que el evaluador ubique su juicio en un continuo de calidades fijadas, con números o adjetivos.

Por ejemplo, en una escala para evaluar el desempeño de un docente en la habilidad de guiar el aprendizaje de conceptos puede aparecer:

	Imposible de consignar	Nulo	Pobre	Medio	Bueno	Muy Bueno
1. Establece discriminaciones para ajustar significados	0	1	2	3	4	5
2. Utiliza ejemplos para precisar el significado	0	1	2	3	4	5
3. Ayuda a formular definiciones	0	1	2	3	4	5
4. Cuida la correspondencia entre lenguaje y conceptos	0	1	2	3	4	5

Lo que sigue es una parte de una escala para evaluar los elementos de la comunicación en una clase. Se presentan dos formas extremas de un tipo de conducta. La conducta adecuada sería el punto medio entre las dos explicitadas.

1. La cantidad de ejemplos e ilustraciones es excesiva	1	2	3	4	5	La cantidad de ejemplos e ilustraciones es escasa
2. Utiliza demasiados materiales no verbales	1	2	3	4	5	Usa sólo la expresión verbal
3. Modifica exageradamente los tonos de voz	1	2	3	4	5	Es monótono

Todo el análisis que hemos realizado de los distintos tipos de pruebas puede dar los elementos para el segundo paso del proceso de evaluación: decidir cuál es la situación de evaluación y los instrumentos que podemos utilizar. Creemos conveniente sintetizarlos en un cuadro, agregando algunas consideraciones sobre la construcción y corrección de los distintos tipos de pruebas.

## ANÁLISIS DE RESULTADOS

La etapa del análisis de los resultados tiene una importancia fundamental en el proceso de evaluación cuando los consideramos en relación con la enseñanza. Es el momento de preguntarnos como docentes: ¿qué significa esto que obtuve?, ¿qué han aprendido realmente los alumnos? ¿cómo debe seguir? Es la reflexión sobre el aprendizaje, la enseñanza y la evaluación misma. Un análisis cuidadoso nos muestra las fallas de nuestra evaluación y nos ayuda a mejorarla. Habitualmente no se dedica a esta etapa ni el tiempo ni el cuidado suficiente y no se obtiene toda la información que ésta puede proveer.

Vamos a ver qué conclusiones podemos sacar a partir de la tabulación de los datos. Del análisis podemos obtener información sobre: 1) la calidad del instrumento usado; 2) el rendimiento individual y grupal; y, 3) las necesidades de la enseñanza.

	Pruebas verbales de elaboración	Pruebas verbales de selección	Pruebas de ejecución
Capacidades que miden	Recordación de información verbal. Manejo de habilidades intelectuales a través de expresión propio.	Reconocimiento de información verbal. Habilidades intelectuales en situaciones estructuradas.	Destrezas motoras. Habilidades complejas, tales como las de comunicación o estrategias docentes.
Calidad de la medición	Poco confiable, a menos que se fengan cuidados especiales en la elaboración y análisis del registro.	Confiable, si está bien elaborada.	Poco confiable, mejorable, con cuidados para registrar y analizar los datos (registros grabados o filmados, listas de control).
Relación con el aprendizaje	Guía hacia una necesidad de expresión correcta y segura.	Guía hacia un análisis discriminativo de los contenidos "cognitivos".	Provee retroalimentación para apreciar actuaciones poco controlables.
Elaboración	Requiere elaborar pocas preguntas. Deben estar claramente definidas, ser específicas para poner límites pero sin coartar la expresión.	Requiere elaboración muy cuidadosa de muchos ítems. Exige ingenio y entrenamiento.	Requiere claridad y especificación en el planteo de la situación, que debe estar muy bien delimitada.
Análisis de los resultados	Lento y difícil. Requiere cuidados especiales para lograr objetividad (Clave de corrección, relectura). Énfasis en la calidad de enseñanza y aprendizaje.	Rápido y sencillo. Permite análisis técnicos muy elaborados para juzgar la calidad de la prueba, además de la calidad de enseñanza y aprendizaje.	Lento. Exige un análisis muy cuidadoso de la tarea. Informa sobre calidad de enseñanza y aprendizaje.
Registro	Sencillo. No necesita cuidados especiales.	Conviene utilizar material impreso para facilitarlo.	(Aparatos especiales para grabación o filmación en video-cassette). Listas de control u hojas de registro impresos.

Tomaremos como ejemplo una parte de una tabla de doble entrada que se elaboró con las respuestas de 44 alumnos a una prueba de selección múltiple. En el eje vertical, pueden leerse las respuestas a cada ítem; en el eje horizontal, las respuestas de cada alumno a todos los ítems. Los sujetos aparecen numerados en orden correlativo, de mayor a menor, de acuerdo con el puntaje total obtenido en la prueba. Cada ítem correcto se valoró con un punto. Se reproduce la tabla sólo para 12 ítems. Ver a continuación la Tabla I.

TABLA I

Ítem	1	2	3	4	5	6	7	8	9	10	11	12
Alumno												
1	d	b	c	b	d	a	a	c	c	a	a	b
2	d	b	c	b	d	a	a	c	c	d	a	b
3	d	b	c	b	d	a	a	c	c	a	d	b
4	d	b	c	b	d	b	a	c	b	a	a	b
5	d	b	c	b	d	b	d	c	a	a	a	c
6	d	b	c	b	d	a	c	c	c	d	d	b
7	d	b	c	b	d	a	a	a	c	a	c	c
8	d	b	c	b	d	a	a	c	a	a	d	c
9	d	b	c	a	d	b	d	b	c	d	c	b
10	d	d	c	b	d	b	a	c	a	d	a	a
11	d	b	c	b	d	a	d	c	a	d	d	c
12	d	b	c	b	a	b	d	c	c	d	a	c
33	d	b	c	b	d	b	b	a	d	c	c	d
34	d	a	a	b	c	b	c	c	a	d	c	c
35	d	a	c	a	d	a	c	c	d	b	c	c
36	d	b	a	b	a	a	a	a	d	b	b	a
37	a	b	a	a	c	a	a	a	d	a	d	d
38	a	b	c	b	b	a	d	a	b	d	d	c
39	d	d	a	b	a	b	a	c	b	c	d	c
40	d	a	a	b	b	c	b	a	d	b	d	d
41	d	a	a	b	a	c	b	c	a	b	c	a
42	d	d	b	a	c	b	b	a	d	c	d	c
43	c	d	b	b	a	a	b	c	a	a	c	a
44	b	d	a	b	b	a	d	a	d	c	c	a

## ◆ Calidad del instrumento

**Discriminación:** habíamos visto que para hallar la discriminación debíamos comparar los sujetos de alto rendimiento con los de bajo rendimiento. Necesitamos formar dos grupos extremos significativos. Si tomamos el 50% más alto y el 50% más bajo estamos haciendo un corte muy arbitrario porque, evidentemente, diferencias de un punto en el medio de la tabla no son significativas. El sujeto N° 22 podría estar tanto en el grupo alto como en el grupo bajo. Hay que armar los grupos tratando de tener la mayor cantidad posible de sujetos, bien diferenciadas. Estadísticamente se aconseja tomar el 27% de cada extremo. En nuestro caso, está representado por 12 sujetos. Si bien podemos hacer la tabla registrando a todo el grupo (cuando no es muy grande), tomamos para analizar sólo los 12 primeros y los 12 últimos.

Para cada ítem consignamos la cantidad de sujetos del grupo alto -A- que lo contestó correctamente y la cantidad de sujetos del grupo bajo -B- que lo contestó correctamente. Si marcamos con color las respuestas erróneas, a simple vista podremos tener una apreciación de la discriminación de los ítemes. La mayoría de los ítemes, especialmente los primeros, discriminan.

La discriminación se puede calcular fácilmente restando del total de respuestas correctas para el grupo alto - $C_A$ - el total de respuestas correctas del grupo bajo - $C_B$ - y dividiendo por la cantidad de casos de uno de los grupos  $N_A$  -cualquiera de los dos puestos que son iguales- o sea

$$\frac{C_A - C_B}{N_A}$$

**Dificultad:** ya dijimos que la "dificultad" no era un dato importante en sí mismo, pero sí en relación con el aprendizaje. La "dificultad" se informa habitualmente en términos del porcentaje de sujetos que contestaron bien cada ítem; en realidad es un índice de facilidad, como algunos autores prefieren llamarlo. Lo obtenemos dividiendo el número de respuestas correctas por el total de alumnos. Como en este caso tenemos consignadas sólo las respuestas de los grupos extremos, la "dificultad" podemos hallarla sumando la cantidad de respuestas correctas de cada grupo y dividiendo por el total de casos analizados, en este caso 54% (27% + 27%) o sea

$$\frac{C_A + C_B}{N_{(A+B)}} \times 100$$

Este cálculo no proporciona una medida exacta sino aproximada de la dificultad, porque no está considerado todo el grupo. Pero la distorsión que produce no es significativa para los fines del diagnóstico de aprendizaje que nos interesa.

Con este cómputo obtenemos el siguiente cuadro para los 12 ítemes analizados.

Ítem	Dificultad	Discriminación
1	83	0.33
2	58	0.66
3	62	0.75
4	83	0.16
5	54	0.75
6	54	0.08
7	41	0.33
8	54	0.58
9	29	0.58
10	37	0.41
11	25	0.50
12	25	0.50

Analizando estos resultados surge que ninguno de los ítemes tiene una dificultad extrema. Los primeros son más fáciles y los últimos considerablemente más difíciles. Veremos después la relación de estas dificultades con la ubicación de esos ítemes en la tabla de especificaciones, para ver qué conclusiones podemos sacar en relación con el aprendizaje.

En cuanto a la discriminación, se consideran índices aceptables los que están por encima de 0,40. Este límite no es estricto porque la discriminación depende de varios factores, especialmente la homogeneidad de la prueba total, que es lo que se usa como criterio para diferenciar los grupos altos y bajos. La mayoría de nuestros ítemes muestran una buena discriminación.

Analicemos algunos casos especiales. Comparemos el ítem 1 con el 4. Ambos son fáciles. En un ítem fácil no es posible obtener una discriminación alta porque, en realidad, cuando la mayoría del grupo lo responde correctamente no pueden darse diferencias entre los altos y los bajos. Por lo tanto las diferencias de discriminación no son significativas, de hecho, revisando la tabla de resultados, la diferencia que aparece es que un alumno del grupo alto contestó mal el ítem 4. En los ítemes con dificultades cercanas a los extremos, pequeñas diferencias aparecen como muy significativas, esos datos deben ser tomados con cuidado. Si en nuestra evaluación puede ser útil usar preguntas para constatar que todos los pueden responder bien, o sea preguntas muy fáciles, debemos saber que los otros valores de esa pregunta van a resultar poco significativos. Sin embargo en este caso, ítemes 1 y 4, hay otro dato significativo. Veamos las respuestas:

Ítem	Respuestas			
	a	b	c	d
1	2	1	1	20*
4	4	20*	0	0

\* Respuesta correcta

En el ítem 1 las respuestas incorrectas se distribuyen entre todas las opciones; en cambio en el ítem 4 la única opción que funcionó fue "a". Esto debe llevarnos a analizar el ítem: las opciones "c" y "d" pueden ser obviamente incorrectas, heterogéneas y eso facilitó la respuesta, o "a" es un error común que todavía subsiste en algunos alumnos. En cualquier caso, el ítem 4 merece ser revisado si se desea volver a usarlo, o se debe revisar el aprendizaje del concepto expresado en "a", por lo menos con esos cuatro alumnos.

Tomemos otra situación: los ítems 5 y 6. Ambos tienen una dificultad intermedia. Si todo el grupo A los hubiera contestado bien y mal todo el grupo B, tendríamos el máximo de discriminación. En el ítem 5, la discriminación es muy alta. No se ven problemas con este ítem. En el 6, en cambio, la discriminación es muy baja. Revisemos.

Ítem	Grupo	Respuestas			
		a	b	c	d
6	A	7*	5	0	0
	B	6*	4	2	0

\* Respuesta correcta

Casi tantos alumnos del grupo A como del B lo contestaron correctamente. El análisis de las opciones nuevamente es útil. La opción "d" no fue elegida por nadie en ninguno de los dos grupos. Evidentemente no era buena. El ítem resulta con tres opciones. La opción "b" fue elegida por muchos alumnos de A, es probable que sea demasiado buena. Puede haber detalles en la redacción que haya llevado a los alumnos que más dominan el contenido a pensar que era correcta. Conviene revisar el ítem: preguntar a los alumnos cómo la entendieron, rehacer por lo menos la opción "d". Si no se encuentra nada en la construcción que explique estos resultados, la revisión debe ir a los contenidos de aprendizaje implícitos en cada uno.

¿Qué pasa con los ítem 11 y 12?

Ítem	Grupo	Respuestas			
		a	b	c	d
11	A	6*	0	2	4
	B	0*	1	6	5
12	A	1	6*	5	0
	B	4	0*	5	3

\* Respuesta correcta

Ambos son ítemes fáciles, con buena discriminación; en realidad muy buena, dado el nivel de dificultad. Las opciones funcionaron todas. Con esos datos no se puede pensar en fallas en la construcción. Aquí solamente se justifica el análisis en función de aprendizaje.

El ítem 7 no alcanza el límite de discriminación que hemos fijado; es medianamente difícil.

Ítem	Grupo	Respuestas			
		a	b	c	d
7	A	7*	0	1	4
	B	3*	5	2	2

\* Respuesta correcta

Aparentemente no se ven problemas; las opciones han sido elegidas todas. Quizás convendría revisar la "d"; una cantidad de alumnos de A la eligieron y pocos de B; tendría que ser a la inversa. Pero si no se encuentran problemas, puede pensarse que la baja discriminación se debe a que este ítem es heterogéneo dentro de la prueba como totalidad. Si pensamos que aquello que mide es relevante para nuestros propósitos, no es conveniente pensar en descartarlo ni modificarlo.

Estos análisis ejemplificados para una prueba de selección múltiple pueden hacerse también para otras formas de prueba, por ejemplo, de doble alternativa o de elaboración. La clave de corrección de la página 40 puede servir de base para una tabla de resultados.

#### ◆ Diagnóstico del aprendizaje

Para estudiar los resultados en función del aprendizaje es necesario primero revisar los ítemes en relación con la tabla de especificaciones, o, lo que es lo mismo, en relación con los objetivos.

Varios ítemes de una prueba pueden ser muestra del mismo objetivo, pueden ser representantes de un mismo casillero de la tabla, o pueden apuntar a un mismo contenido en distintos niveles de conducta.

El análisis en función de estas relaciones nos indica la estabilidad o dominio de lo aprendido, la coherencia de los logros o de las fallas, tanto a nivel individual como de grupo.

La prueba de la cual formaban parte de los ítemes que estuvimos analizando era tipo formativo, buscaba diagnosticar el progreso de los alumnos y sus posibles fallas mientras trabajaban en una secuencia de aprendizaje. Era parte de un curso de formación docente. El tema era evaluación. Deseábamos investigar qué nivel de dominio tenían los alumnos de los conceptos fundamentales antes de entrar a

los aspectos técnicos. La tabla de especificaciones tenía los encabezamientos que aparecen en nuestro ejemplo y los ítemes estaban ubicados de esta forma:

	Conocimiento de términos	Interpretación
Evaluación y enseñanza	1	7
	2	8
	3	
	4	
Evaluación y objetivos	5	9
	6	12
	10	
	11	

Tenemos acá la estructura conceptual y conductual de la prueba, y deseamos encontrar la configuración de errores y logros grupales e individuales. La lectura de la Tabla I nos permite interpretar la forma en que el grupo se comportó frente a cada ítem. Los ítemes 1, 2, 3 y 4, que aparecen en el mismo casillero, son los que resultaron más fáciles y con buena discriminación. El conocimiento de esos términos básicos está logrado. En cambio, los ítemes 7 y 8, que evaluaban la posibilidad de relacionar esas ideas y ubicarlas en otro contexto, resultaron más difíciles. Ese nivel de conducta no está todavía logrado para la mayoría del grupo. En cambio la terminología del tema objetivos todavía tiene dificultades, y más aún la interpretación de esos temas. El aprendizaje se está desarrollando normalmente en el grupo. Cada línea horizontal señala el rendimiento de un alumno, en la tabla de análisis de resultados. Podemos, a través de esa lectura, determinar las fallas personales y elaborar las formas de acción. Resulta interesante notar que el ítem 6 pedía el conocimiento del mismo término que después se usa en el ítem 12. Si bien el 6 es más fácil que el 12, en las líneas horizontales no hay concordancia. Hay alumnos que contestaron mal el 6 y bien el 12; esto no resulta coherente. Otra vez, esto hace pensar que la falla está en el ítem.

### **Diagnóstico de necesidades de la enseñanza**

Del diagnóstico anterior podemos determinar que hay aspectos, los ítemes 1, 2, 3 y 4, que grupalmente ya están dominados; no es necesario volver sobre ellos. Pero sí hace falta volver a trabajar, con metodología nueva, las relaciones entre ellos. Es indudable que todavía no dominan este aprendizaje. Para algunos alumnos, los que aparecen últimos en la tabla, es necesario plantear formas individuales de rever los errores. No podrán seguir adelante si no logran superar esos errores.

En cambio en el grupo de ítemes que sigue -5, 6, 10 y 11- la revisión de estos conceptos debe ser completa; el grupo entero muestra rendimiento bajo. La estrategia debe variar; las formas usadas anteriormente no han dado el resultado esperado y debemos volver atrás.

Planteamientos como éstos son los que se pueden obtener del análisis de resultados. La retroalimentación para el docente surge del análisis de los éxitos y los fracasos, tanto en cantidad como en calidad.

Esta tarea de análisis se enriquece cuando se realiza en colaboración con los alumnos. Ellos pueden intervenir en todos los pasos, y al mismo tiempo que facilitan la labor del docente, obtienen una retroalimentación mayor. Cobra mayor significado cada uno de los datos del análisis cuando se sabe cómo se obtuvo y de dónde surgió.

Es especialmente interesante reelaborar con los alumnos las claves de corrección para las pruebas abiertas y pedir a ellos que revisen sus respuestas en función de la clave. Las evaluaciones y las notas pierden carácter punitivo cuando forman parte de un proceso de autoevaluación integrado en la enseñanza. La autoevaluación guiada por un esquema de análisis, ayuda a tomar distancia de los propios productos, a verlos con mayor objetividad. Es indudable que las funciones de motivación, fijación y orientación del aprendizaje se cumplen mucho más eficazmente cuando todos se sienten responsables de la evaluación. Cuantos más momentos del proceso compartan docentes y alumnos, mejor se van a cumplir los objetivos de la enseñanza y de la educación. Si comparten todo el proceso, la evaluación va a ser considerada como una tarea más enriquecedora y no persecutoria.

R.S. Peters señala que para que un proceso pueda llamarse educativo debe cumplir con dos requisitos básicos: su contenido debe ser valioso y sus procedimientos moralmente inobjetables, en tanto apelen a la intencionalidad y a la conciencia del que se educa. El educando debe saber lo que está haciendo, debe tener conciencia de lo que está tratando de aprender, de dominar. La evaluación es el mejor camino para que el proceso de enseñanza sea más educativo, sea más un enseñar a aprender que un condicionamiento.

## BIBLIOGRAFÍA

1. Adams, Georgia Sachs. **Medición y Evaluación.** Barcelona, Herder, 1970
2. Block, James M. **Cómo aprender para lograr el dominio de lo aprendido.** Buenos Aires, El Ateneo, 1975.
3. Bloom, B., Hastings, J. y Madaus, G. **Evaluación del aprendizaje.** Buenos Aires, Troquel, 1975.
4. Ebel, Robert L. **Fundamentos de la medición educacional.** Buenos Aires, Guadalupe, 1977.
5. Gagné, Robert M. **Las condiciones del aprendizaje.** 3a. edición. México, Interamericana, 1979.
6. Peters, Richard S. comp. **El concepto de educación.** Buenos Aires, Paidós, 1969.
7. Sawin, Enoch I. **Técnicas básicas de evaluación.** Madrid, Magisterio Español, 1970.
8. Smith, Othanel B. y Ennis, Robert M. **Lenguaje y conceptos en la educación.** Buenos Aires, El Ateneo, 1971.



Subsecretaría de Evaluación de la Calidad Educativa  
Dirección Nacional de Evaluación  
Paraguay 1657 3°/4° (1062) Buenos Aires - Argentina  
Tel. / Fax: (541) 811-3297 / 812-6520 / 812-4684 / 813-5638