

UNIVERSIDAD CATOLICA DE CHILE

MANUAL DE EVALUACION

FACULTAD DE  
INGENIERIA

PROGRAMA  
INTERDISCIPLINARIO  
DE INVESTIGACIONES  
EN EDUCACION

1974

## I N D I C E

	<u>Pág.</u>
I INTRODUCCION	1
II LA EVALUACION SUMATIVA Y FORMATIVA	2
III ASPECTOS GENERALES	5
i) El proceso	5
ii) La supervisión	7
iii) La evaluación	8
iv) La medición	10
v) La calificación	11
IV QUE EVALUAR	13
i) Efectividad	13
ii) Eficiencia	14
iii) Evaluación de los esfuerzos y de los recursos disponibles	18
iv) Criterios para evaluar el material didáctico	19
v) Causas del fracaso escolar	20
V COMO MEDIR	21
i) Diferentes tipos de instrumentos	
ii) Características de algunos tipos de ítem	23
iii) Etapas en la construcción de instrumentos de evaluación	33
VI COMO CALIFICAR	39
i) La nota relativa	39
ii) Ejemplo para pasar de puntajes a notas	40

	<u>Pág.</u>
iii) Otros aspectos de la nota relativa	41
iv) Las notas, sus limitaciones y su significación	42
VII COMO ANALIZAR LOS INSTRUMENTOS	45
i) Confiabilidad	45
ii) Validez	46
iii) El análisis de ítem	48
iv) Corrección de los puntajes	50
VIII CONSTRUCCION DE ITEMS PARA LOS DIFERENTES NIVELES DEL CONOCIMIENTO	55

## LA SUPERVISION EN EL PROCESO DE ENSEÑANZA APRENDIZAJE

### I INTRODUCCION.

Cuando se construye un nuevo tipo de submarino, se emplean técnicas bastante sofisticadas para controlar que los materiales cumplan las especificaciones, para que la precisión de sus mecanismos sea la indicada y para que los plazos sean los estipulados. Si se detecta un error, inmediatamente se corrige y si el proyecto no tiene éxito, o sólo se lo-gran algunas de las metas, se cuestionará todo; los materiales, los operarios y muy especialmente al autor del proyecto y al constructor.

En la enseñanza cuando el proyecto es la formación de un grupo de personas, lo que parece ser más complejo y más trascendente, los errores se consideran poco y de los fracasos normalmente se responsabiliza a los alumnos.

Nuestra intención es motivar al profesor para aplicar en el proceso de enseñanza-aprendizaje un "sistema de control del proceso", entregándole ejemplos o sugerencias de instrumentos de medida y modelos de flujo de información que faciliten la labor del profesor.

## II LA EVALUACION SUMATIVA Y LA EVALUACION FORMATIVA.

La experiencia realizada por B. Bloom, ha comprobado que el 90% de los estudiantes puede alcanzar las metas prefijadas en tiempos promedios similares a los actuales, siempre que se respeten las diferencias individuales. Para el profesor la labor importante no es "pasar la materia" sino determinar cómo la mayor proporción de estudiantes puede aprender eficientemente los conocimientos y la materia de estudio, su éxito como profesor se medirá, justamente, por su capacidad, su creatividad para idear los métodos y los medios para obtener el aprendizaje de todos los alumnos.

Aparece pues como un defecto importante la evaluación selectiva, es decir, fijar niveles de exigencia que sólo pueden alcanzar unos pocos, sin importar el resto de los estudiantes que podría aún alcanzar niveles altos de aprendizaje usando mejores métodos y estableciendo una elasticidad en los tiempos de aprendizaje. No debe olvidarse que los alumnos universitarios ya han sido cuidadosamente seleccionados y que se han realizado cuantiosas inversiones en ellos. Tampoco hay que olvidarse que las frustraciones crean frustraciones e impiden el desarrollo de una educación permanente. El sistema de enseñanza personalizada da a cada alumno la posibilidad de seguir su propio ritmo de aprendizaje, utilizando todas las ayudas que estime conveniente; pero basándose principalmente en la auto-instrucción y evaluándose finalmente por la cantidad de materia que domina, sin dejar la posibilidad de pasar a una ma-

teria nueva sin dominar la anterior.

Un sistema similar a la enseñanza personalizada permite cumplir los objetivos del curso a una abrumadora mayoría de los estudiantes.

Según los datos disponibles, dependiente de su aptitud el alumno se puede demorar hasta seis veces más en una determinada materia, pero con un buen aprendizaje se puede reducir esta relación a 3:1.

Normalmente se ha planteado el proceso de enseñanza aprendi-zaje como algo separado de la evaluación, y de hecho es así para el caso de la evaluación acumulativa, en que se "mide", mediante algunos instrumentos, lo que ha aprendido el estudiante, y se obtiene un promedio acumulativo final; de hecho este procedimiento está levemente ligado al proceso de enseñanza ya que el aprendizaje se verá reflejado en la evaluación. La evaluación formativa, en cambio, forma parte del proceso de aprendizaje mismo. Los tests formativos deben detectar si el estudiante ha logrado su aprendizaje, indicar cuál es el método recomendable para cumplir con el objetivo, si aún no lo ha logrado, reforzar el aprendizaje mediante estímulos que se dan prácticamente junto con el aprendizaje, preparar al estudiente para obtener un éxito en la asignatura, indicar las fallas puntuales a los alumnos que no cumplen con el objetivo, y finalmente, indicar-le al profesor cuáles son las fallas y los aspectos positivos de cada uno de los alumnos que está tomando el curso. El diagnóstico debería acompañarse de una prescripción específica si es que se desea que los alumnos puedan mejorar su comportamiento.

Se ha demostrado empíricamente que los alumnos se esfuerzan en realidad mucho más por superar sus dificultades cuando se les hacen sugerencias específicas sobre lo que ellos deben hacer.

Si se va a hacer una evaluación acumulativa paralela a la evaluación formativa es importante que el nivel de exigencia para el test formativo, sea similar al de la prueba acumulativa. Además en ambos casos es recomendable disminuir la subjetividad del profesor para calificar usando pruebas de tipo objetivo.

El empleo de la evaluación formativa permite que tanto el profesor como el alumno, tengan alguna idea de los criterios de rendimiento establecidos y ambos deben ser capaces de dar evidencias del progreso hecho en su consecución. Es conveniente que el alumno conozca también el orden de rango que ocupa en el grupo; ya que se sabe que la competencia puede ser un buen aguijón para el estudio, siempre que no se llegue a excesos.

Finalmente, debe dejarse en claro que no pueden darse estrategias únicas sino más bien incentivar a los profesores para que busquen cuáles son los métodos más adecuados para lograr un buen aprendizaje, ésto es que un 90% de los alumnos cumplan con todos los objetivos (según los datos disponibles) al fin y al cabo, es, por excelencia su labor como profesor.

Podríamos resumir diciendo que el objeto de la evaluación

formativa es determinar el grado de dominio de una conducta específica en el aprendizaje y detectar los aspectos puntuales no dominados, es decir, ayudar tanto al profesor como al alumno a centrarse en un aprendizaje determinado hasta lograr su dominio. En este sistema la nota no interesa sólo interesa lograr la conducta en un tiempo que no sea demasiado largo. Se usa permanentemente durante el proceso de enseñanza aprendizaje y forma parte de él.

La evaluación acumulativa se refiere en forma más amplia al grado en que se ha logrado el aprendizaje por un curso o por cada alumno sobre una parte importante de la materia. La evaluación acumulativa está destinada generalmente sólo para calificar y se realiza dos o tres veces por semestre.

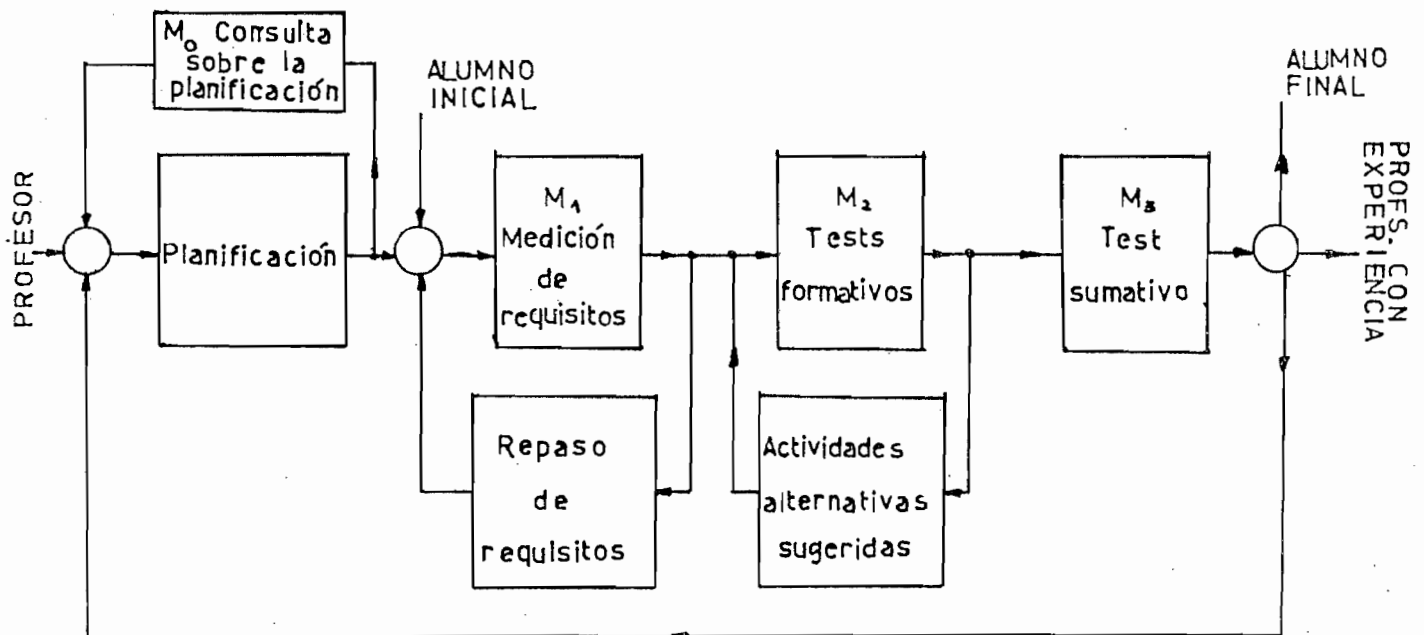


### III Aspectos Generales

#### 1) El Proceso

Para que un proceso funcione bien deben estar muy claros los objetivos o la calidad del producto final. Mientras más precisas sean las especificaciones más fácil es discriminar si ellos se cumplen, y se pueden identificar mejor los errores puntuales, los que podrán corregirse oportunamente si la frecuencia de las mediciones es adecuada.

Al Supervisar un proceso se debe disponer de los instrumentos de medición adecuados, de los canales de información y de los mecanismos para enmendar errores y corregir las deficiencias. En el proceso de enseñanza-aprendizaje un modelo útil podría ser el siguiente.



CGR.

En el modelo se puede observar que a cada etapa y a cada actividad le corresponde una medición, indicándose además el sentido de la información para perfeccionar el modelo. El producto final no es tan sólo el alumno si no también el profesor con una nueva experiencia.

ii) La Supervisión.

Podemos resumir lo dicho anteriormente en 5 presupuestos básicos de la Supervisión.

- 1 Claridad máxima en los objetivos (especificación y jerarquización).
- 2 Conocimiento de los objetivos por parte del alumno (compromiso con ellos).
- 3 Evaluación previa para discriminar sobre el cumplimiento de los objetivos. y sobre los prerequisites.
- 4 Contar con los instrumentos adecuados para lo que se quiere medir.
- 5 Tener criterios sobre la interpretación y utilización de los resultados.

En la Supervisión, que es la forma de controlar un proceso pueden distinguirse algunas actividades fundamentales.

- En primer lugar está la medición, que es la actividad destinada a proporcionar la información para evaluar y dar las pautas para clasificar a los alumnos y los resultados del profesor.

La medición consiste en detectar la información asignándole símbolos a las observaciones de una manera significativa y coherente, comparando los resultados con patrones absolutos o relativos. La medición se hace mediante una serie de "instrumentos" que se verán más adelante.

- En segundo término está la evaluación, que consiste en emitir juicios sobre la información recibida a través de una o más mediciones.
- En tercer lugar está la calificación que consiste en comunicar, usualmente a través de códigos, la evaluación.

### iii) La Evaluación.

#### Algunos elementos Básicos

- 1) La evaluación es un método para adquirir y procesar la información sobre las necesidades evidentes del aprendizaje y de la enseñanza.
- 2) La Evaluación es un proceso permanente que permite juzgar el valor o cantidad de cambio obtenido según los objetivos propuestos.
- 3) La Evaluación permite valorar los trabajos del alumno, conocer sus aspectos fuertes y sus aspectos débiles.

- 4) La evaluación permite valorar el actuar del profesor y le muestra hasta donde ha logrado sus propósitos.
- 5) La evaluación permite valorar el ambiente, las dificultades sociales, económicas y su influencia en el aprendizaje.
- 6) La evaluación permite valorar el material didáctico que se está usando.
- 7) La evaluación permite descubrir cuán bien se están cumpliendo los objetivos.
- 8) La evaluación permite descubrir las causas de éxitos y fallas específicas.
- 9) La evaluación permite descubrir los principios subyacentes a los programas que tienen éxito.
- 10) La evaluación permite sentar las bases para una investigación sobre las causas de los éxitos relativos de las diversas técnicas empleadas en la enseñanza.
- 11) La evaluación permite orientar las experiencias con técnicas que aumentan la efectividad.
- 12) La evaluación permite redefinir los medios usados para alcanzar los objetivos y aún redefinirlos a la luz de los datos característicos de la medición.

iv) Características de la medición.

- 1.- La medición es generalmente indirecta, por ejemplo, se mide la aplicación de un principio a través de un sólo ejemplo.
- 2.- Normalmente se trata de cuantificar la información recopilada en la medición.
- 3.- Al medir se producen errores debido a:
  - a: la persona que hace la medición.
  - b: los instrumentos inadecuados.
  - c: a la falta de consistencia, esto es, a la variación de criterios de comparación.
- 4.- Las mediciones son relativas, debido a problemas de los alumnos, de los profesores y al período en que se hace.

iv La calificación.

1. Es muy común entre los profesores confundir todo el proceso de supervisión con la calificación. La calificación es la forma de informar al alumno sobre su aprendizaje y usualmente se asocia a una nota.
2. "Existe concenso general en que las notas deben colocarse de acuerdo con una base de comparación". Esto significa que los mejores alumnos deberían obtener la nota 7, los siguientes mejores la nota 6, y así sucesivamente. Esto parece más justo que tomar un puntaje arbitrario y reprobar a los que no lo alcancen.
3. "Las notas deben estar basadas, en cuanto sea posible, en mediciones objetivas". No siempre es posible reducir la evaluación a pruebas objetivas; pero en la medida en que deban usarse pruebas de ensayo (preguntas de respuestas abiertas), conviene definir previamente los criterios de evaluación de cada pregunta.
4. "En lo posible, las notas deben expresar el logro obtenido en objetivos específicos, más bien que dar resultados sobre apreciaciones subjetivas". No deberíamos tratar de informar mediante una sola nota un conocimiento en una determinada materia, y a la vez, otros aspectos de la disciplina. Puede servir más como información al alumno distinguido los logros alcanzados en diversos aspectos de la disciplina.
5. "Se obtendrán un mejor sistema de colocación de notas empleando va-

rias formas de medición". No sólo convendrá utilizar pruebas objetivas, sino también interrogaciones, trabajos individuales, etc.

6. "No utilizar una escala de notas demasiado amplia". Es más fácil colocar notas con una escala de 7 puntos que con una de cien puntos. Se ha demostrado que los profesores pueden discriminar entre seis o siete niveles de conocimientos, pero no entre cien. Por ello deberíamos tener mucho cuidado al colocar una nota con uno o dos decimales.

Por lo demás la precisión no la dan los decimales sino los instrumentos de medida.

## IV QUE EVALUAR.

Una sugerencia para determinar que se puede evaluar en un proceso docente se muestra a continuación.

Se sugieren tres aspectos en torno a los cuales se puede realizar una evaluación completa:

- i) Efectividad o Evaluación de los efectos: que implica medir las acti-  
tudes logradas en torno a los objetivos previamente fijados.

Este aspecto está íntimamente relacionado con la praxis (inte-  
gración de teoría y práctica), si los objetivos se formulan en térmi-  
nos de conductas observables.

Algunas preguntas sugeridas para evaluar efectos son las si-  
guientes:

- 1) ¿Cuáles de las conductas importantes que se querían lograr se han  
logrado?
- 2) ¿En qué proporción se ha logrado cada uno de los efectos importantes?
- 3) ¿Cuáles de las conductas importantes han fallado abiertamente?
- 4) ¿Cuáles son las posibles causas de las fallas de los efectos impor-  
tantes?
- 5) ¿Cuáles de las conductas secundarias se han logrado?



- 6) ¿Cuáles de las conductas secundarias han fallado?
- 7) ¿Cuáles pueden ser las causas de las fallas?
- 8) ¿Es conveniente modificar los objetivos iniciales?
- 9) ¿Qué efectos no previstos han aparecido? (se podrían desglosar en cognitivos, de actitudes y de acciones)
- 10) ¿Cuáles pueden ser las causas de estos efectos?
- 11) Estos efectos, ¿son positivos?
- 12) ¿Qué modificaciones deben hacerse al programa, para evitar estos efectos?
- 13) ¿Se han producido efectos en personas aisladas?
- 14) Los alumnos que pretendían rendimientos más allá de los patrones exigidos ¿han podido realizarlos con éxito?

\* Ver diagrama de control de efectividad, pág. N° 15.

ii) Eficiencia o Evaluación de la ejecución de los procesos y de las actividades: que implica medir como se han logrado los objetivos previamente fijados. Algunas preguntas sugeridas al respecto son las siguientes:

- 1) ¿Cuáles actividades han sido más efectivas en torno a los objetivos?
- 2) ¿Cuáles de las actividades planificadas se han realizado?
- 3) Los caminos (procesos o métodos) ¿son mejores que otros experimentados anteriormente?
- 4) En base a la experiencia lograda, ¿qué modificaciones se sugieren?
- 5) ¿Se cumplieron los plazos programados? indique las posibles causas de falla.

DIAGRAMA DE CONTROL DE EFECTIVIDAD  
(USANDO EVALUACION INFORMATIVA)

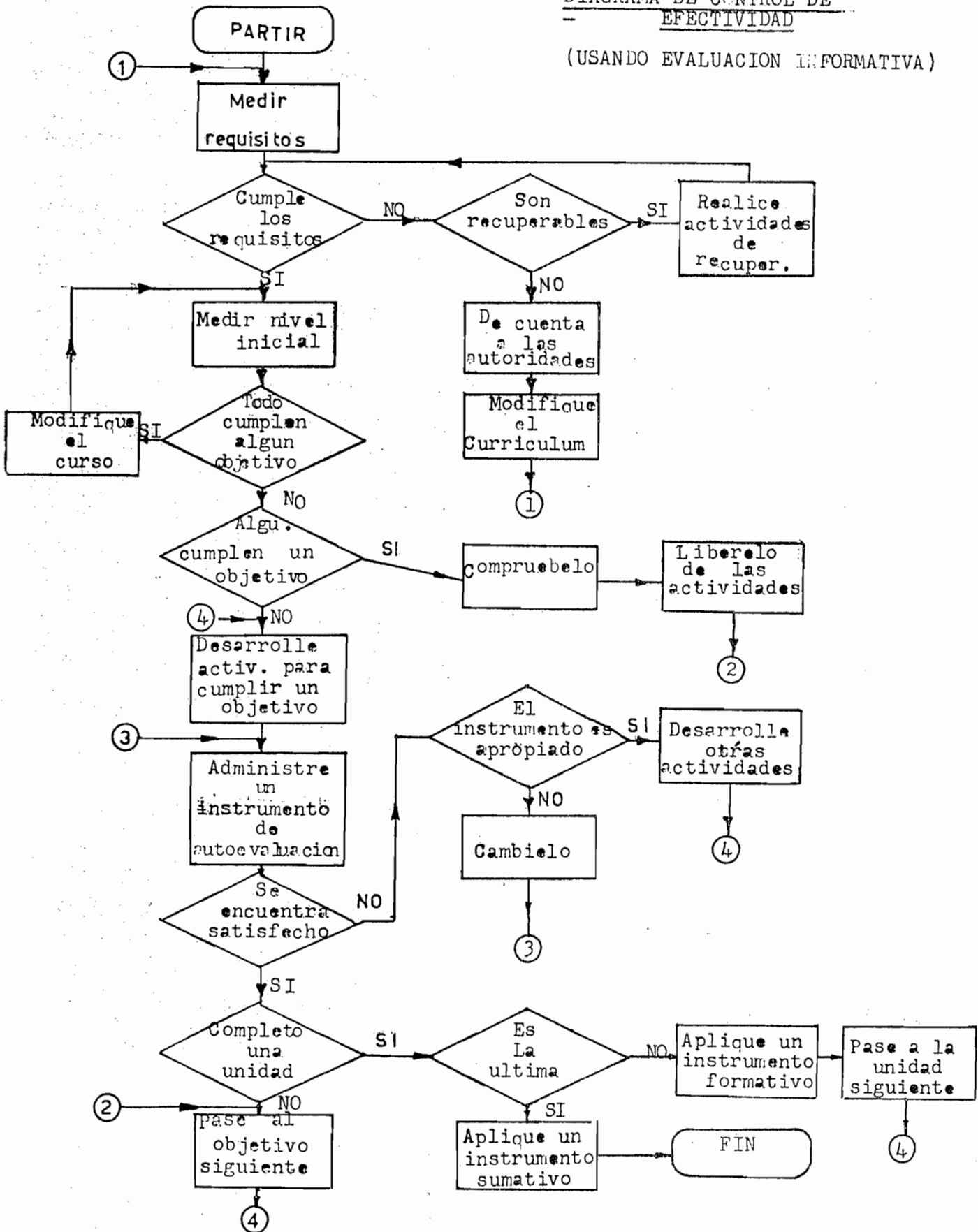
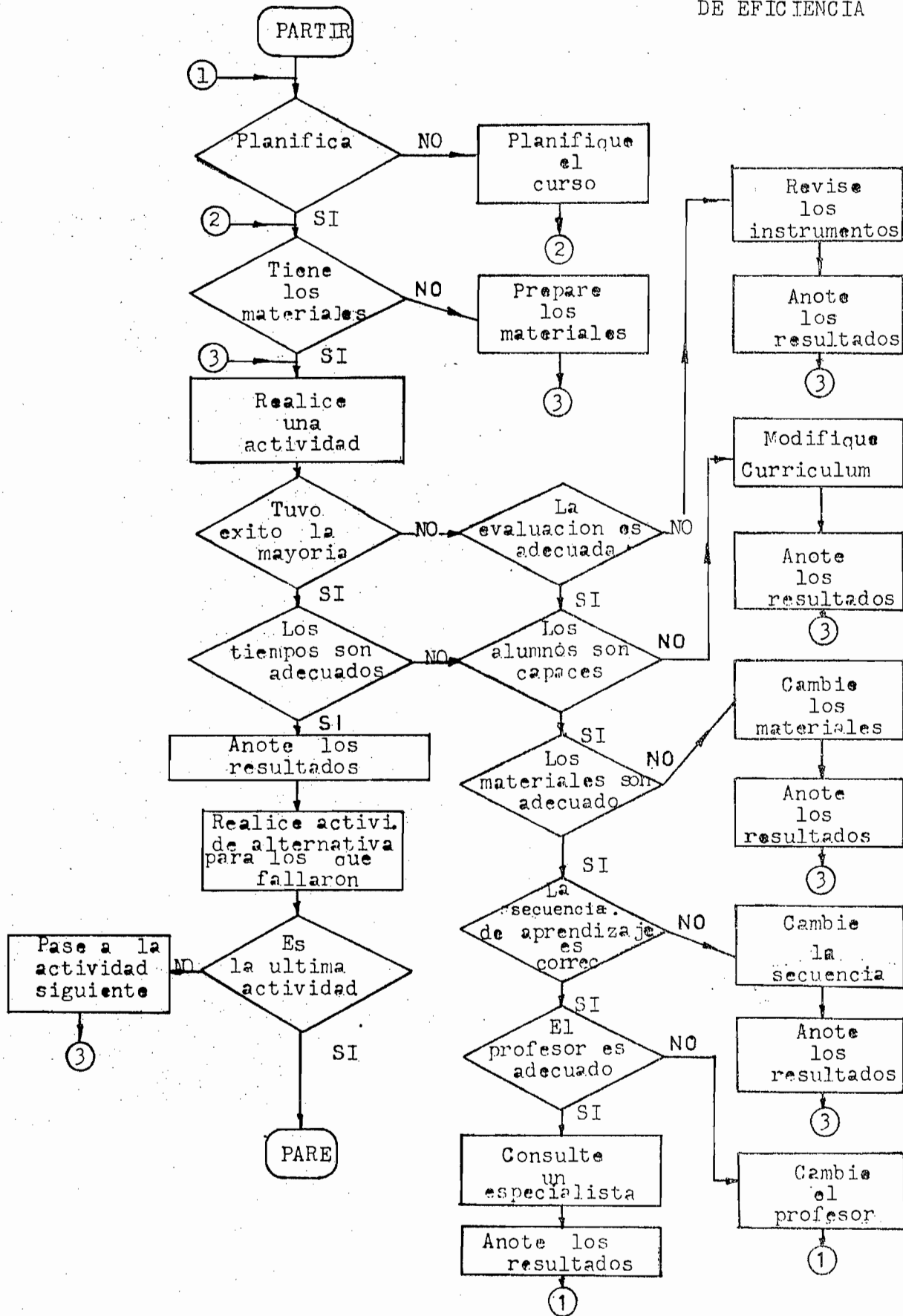


DIAGRAMA DE FLUJO PARA CONTROL DE EFICIENCIA



6) ¿Cuáles son las etapas siguientes?

7) ¿Hubo dificultades con algún alumno en especial? ¿Cuáles pueden ser las causas de su falla?

(Ver diagrama control de eficiencia Pág. 17)

iii) Evaluación de los esfuerzos y de los recursos disponibles:

Implica medir la eficiencia en la utilización de recursos en cuanto ellos sirven para cumplir los objetivos fijados.

Algunas preguntas sugeridas al respecto son las siguientes:

- 1.- ¿Cuáles son los éxitos y las fallas del sistema?
- 2.- ¿Cuáles son las causas específicas de éxitos y fracasos (Esto implica desglosar cuidadosamente el programa).
- 3.- ¿Cuán afectado ha sido cada persona por el sistema?
- 4.- ¿Bajo qué condiciones el sistema ha logrado mayor éxito?
- 5.- ¿Cuán efectivo ha sido el sistema en torno a las necesidades reales, se mide por el coeficiente.

$$C_1 = N^\circ \text{ personas servidas con éxito} / N^\circ \text{ personas disponibles.}$$

- 6.- El sistema que se ha aplicado a un grupo ¿da resultados al aplicarlo masivamente?
- 7.- El costo total (Profesores ayudantes, materiales, etc. es menor que otras alternativas con rendimiento similar. Se puede medir por el coeficiente:

$$C_2 = \text{Costo del sistema B} / \text{Costo del sistema A}$$

se usará el sistema A si  $C_2$  es mayor que 1

- 8.- ¿Cuál es el rendimiento en torno al esfuerzo desplegado? Se puede medir por algunos coeficientes tales como:

$C_3 = \text{N}^\circ \text{ alumnos aprobados} / \text{Costo sistema}$

$C_4 = \text{N}^\circ \text{ alumnos aprobados} / \text{Horas profesor empleadas}$

$C_5 = \text{N}^\circ \text{ alumnos aprobados} / \text{Hora alumnos dedicadas al curso.}$

En general estos y otros coeficientes servirían para establecer patrones de rendimientos comparativo, para lo cual se podrían standarizar, modificando en los coeficientes la cantidad de alumnos aprobados por % de alumnos aprobados.

iv) Criterios para evaluar el material didáctico.

- i) ¿Se está usando I.P.? Introducción Programada (de que tipo).
- ii) ¿Están divididas las materias en pequeñas unidades con objetivos explícitamente formulados?
- iii) ¿Existen los ejemplos y contra ejemplos suficientes? (indicar criterios para encontrarlos suficientes).
- iv) ¿Está usando el profesor algún tipo de guías en sus escritos? (resumen al final)
- v) ¿Usa el profesor algún sistema de auto control en sus apuntes?
- vi) Los ejemplos o contra ejemplos del profesor, ¿están asociados a cambios conductuales?
- vii) ¿Existe ayuda audiovisual?
- viii) ¿Se justifica su costo en relación al rendimiento y al N° de alumnos?

v) Causas del fracaso Escolar.

Después de ver ¿qué evaluar? se puede concluir que las causas de los fracasos en la enseñanza, del punto de vista del profesor, son muy variados y se pueden agrupar en tres aspectos a saber.

- a) fallas en la motivación
- b) fallas en las actividades de los alumnos para lograr los objetivos.
- c) fallas de los materiales empleados.

## V COMO MEDIR.

### i) Diferentes tipos de instrumentos.

Existen diferentes tipos de instrumentos atendiendo a su forma de administrar, a su construcción y corrección, o a las condiciones en que aplica.

#### a) Según las condiciones en que se aplica existen dos tipos fundamentales.

- Aplicado en condiciones naturales, se observa la forma de actuar del sujeto con ciertas pautas y sin que éste se de cuenta que está siendo observado, un ejemplo típico es la "apreciación" que el profesor tiene del alumno de acuerdo a su actitud durante el desarrollo del curso.
- Aplicado en condiciones manipuladas bajo esta clasificación caben las pruebas objetivas y las no objetivas, es decir todos los tipos de pruebas usados comunmente en la enseñanza Universitaria, en que se fijan condiciones.

#### b) Según su forma de administración.

- Test de potencia: En este tipo de test interesa que el alumno conteste el máximo de preguntas, se usan para medir el dominio sobre una materia sin que interese mayormente la velocidad de respuesta. Normalmente se usan ordenados de menor a mayor grado de dificultad.



- Test de velocidad. Interesa fundamentalmente la rapidez del alumno y son útiles para medir destreza.
  - Test mixto es una mezcla de un test de potencia y un test de velocidad.
- c) Según el sistema de construcción y corrección.
- Test informal es aquel test sin mayor análisis construido usualmente por el profesor.
  - Test Standarizado: es un test que ha sido analizado por especialistas, se ha validado, se ha estudiado su confiabilidad, y el análisis de cada ítem.
- d) Por su forma de corregir.
- Test objetivo: No influye la opinión del corrector.
  - Test Sujetivo: influye en forma importante el juicio del corrector, pues la respuesta viene incluida en el texto, o es sumamente específica.

Algunos ejemplos del tipo de ítem para pruebas objetivas son:

- el de verdadero - Falso
- el de selección múltiple
- el de términos pareados
- el de respuesta breve
- el de completación
- el problema (en algunos casos).

Para pruebas no objetivas se tiene.

- El ensayo
- el test oral
- el problema (en algunos casos).

ii) Características de algunos tipos de items.

Se verá a continuación la forma en que se pueden construir algunos tipos de item para los instrumentos de medición.

a) El item de dos alternativas.

Consiste en una proposición, frente a la cual, el alumno debe discriminar si ella es correcta (verdadera) o errónea (falsa).

Ejemplo:	Verdadero	Falso
Un campo conservativo		
disipa energía	-----	-----

Ventajas del item de dos alternativas.

- 1) Permite evaluar una cantidad importante de conocimientos en un período muy breve, ya que se puede construir una gran cantidad de items y se puede exigir rapidez en las respuestas.
- 2) Los items son fáciles de construir.
- 3) Da rápidamente una "impresión general del nivel de aprendizaje del curso.
- 4) Son muy objetivos.

Desventajas.

- 1) El azar puede influir notoriamente (existe 50% de posibilidades de acertar en cada ítem).
- 2) La medición es muy poco matizada, sabe ó no sabe, aunque en algunos casos esto constituye una desventaja.

La forma de corregir la influencia del azar es aumentando el número de ítems y aplicando descuento.

#### Características de construcción.

- 1) Las respuestas deben ser claramente verdaderas ó falsas, por ningún motivo deben dar lugar a confusiones, de ahí que se recomienda preparar las pruebas en equipo.
- 2) Lo falso no debe estar en los detalles, sino en lo substancial.
- 3) El juicio debe ser conciso, usando palabras precisas.
- 4) Usar preferentemente términos cuantitativos (tres de) y no cualitativos (varios de)
- 5) Evitar los juicios negativos (El calcio no es un no metal)
- 6) Existen algunos términos tales como "todo", "siempre", "nunca", "ninguno", "nadie", que tienden a usarse en juicios falsos, ya que generalmente hay excepciones (Ej. todos los hombres son fieles a sus esposas) por eso conviene evitar estos términos.
- 7) Otros términos tales como "algunos", "rara vez", "a menudo", se usan con mayor frecuencia en aseveraciones verdaderas, de ahí

**Ventajas:** 1) permiten medir todos los niveles cognitivos  
 2) a través de un sólo problema se puede abarcar una parte importante de la materia.

**Desventajas:** 1) Los criterios de evaluación permiten mucha subjetividad.  
 2) Son piramidales, de tal forma que pequeños errores al iniciar el problema pueden conducir a resultados falaces.

**Algunas sugerencias de construcción.**

- 1) El profesor siempre debe resolver el problema para estimar los tiempos y comprobar que no falten datos y que en general sean factibles de resolver.
- 2) El problema debe poder resolverse con los elementos que el alumno dispone por su nivel de aprendizaje.
- 3) El planteamiento del problema debe ser claro y conciso no dando posibilidades a interpretaciones erróneas.
- 4) Debe construirse una pauta de evaluación, que es preferible que el alumno conozca, indicando el puntaje que corresponde a cada etapa.
- 5) Cuando la respuesta contempla muchos aspectos es preferible desglosar la pregunta final en preguntas parciales.

iii) Etapas en la construcción de instrumentos de evaluación

Algunas sugerencias para construir items son las siguientes:

- 1) Consulte los objetivos. Como se ha visto anteriormente la medición de la efectividad consiste en discriminar sobre el cumplimiento de los objetivos luego no se puede construir un instrumento de medición sin que este ligado directamente a los objetivos del curso.
- 2) Reúna los materiales que puedan servir como base para los ítems (gráficos, modelos, bancos con los ítem anteriores).
- 3) Escriba los ítem en fichas indicando en la parte superior del anverso el objetivo y en la inferior el ítem y al reverso la fecha, el curso y deje un espacio para el análisis del ítem.
- 4) Analice cada ítem viendo si efectivamente mide lo que se quiere medir, revise si la redacción es correcta y si esta claro para el alumno.
- 5) Consulte con sus colegas cada ítem y sólo use los ítems que todos consideren aceptables.
- 6) Prepare las instrucciones que va a dar a los alumnos.
- 7) Prepare una hoja de respuestas estas hojas pueden standarizarse para toda la facultad colocando por ejemplo 50 números y 5 alternativas. Se adjunta una hoja de respuestas tipo.
- 8) Prepare el folleto completo del test.
- 9) Administre el test.
- 10) Confeccione una plantilla perforando una hoja de respuestas en el lugar de las claves.
- 11) Raye con una línea horizontal de color (puede ser rojo) que atraviese todas las alternativas de los ítem no contestados (omitidos)

- 12) Corrija el test colocando la plantilla sobre la hoja de respuestas de cada alumno, cuente las respuestas acertadas y las omitidas que aparecerán con un color distinto en los huecos de la plantilla, y anótelos en los casilleros correspondientes.
- 13) Coloque el puntaje final (haciendo las correcciones si es necesario) y coloque la nota correspondiente.
- 14) Analice cada ítem midiendo su dificultad y anote sus resultados en la ficha de cada ítem.
- 15) Analice el rendimiento completo del curso, trate de descubrir las fallas fundamentales.
- 16) Aliente a los alumnos que han tenido éxito y ofrezca ayuda a quienes hayan fracasado.
- 17) Haga las anotaciones correspondientes en su sistema de control si está usando evaluación formativa.

Ejemplo de Ficha para la escritura de items

(ANVERSO)

<p>OBJETIVO:</p> <p>Enunciado .....</p> <p>.....</p> <p>A)</p> <p>B)</p> <p>C)</p> <p>D)</p> <p>E)</p>
--

(REVERSO)

<p>Clave:</p> <p>Curso:</p> <p>Fecha:</p> <p>Dificultad:</p>
--

HOJA DE RESPUESTAS

Nombre Completo: ----- CURSO: -----

Fecha:-----

## RESPUESTAS

Correctas: ----- Puntaje sin corregir: -----

Omitidas : ----- corregido: -----

Malas : ----- Porcentaje : -----

Total : ----- NOTA FINAL : -----

Marque la respuesta que Ud. estima que es la correcta en esta forma:

EJEMPLO: A B C ~~D~~ E

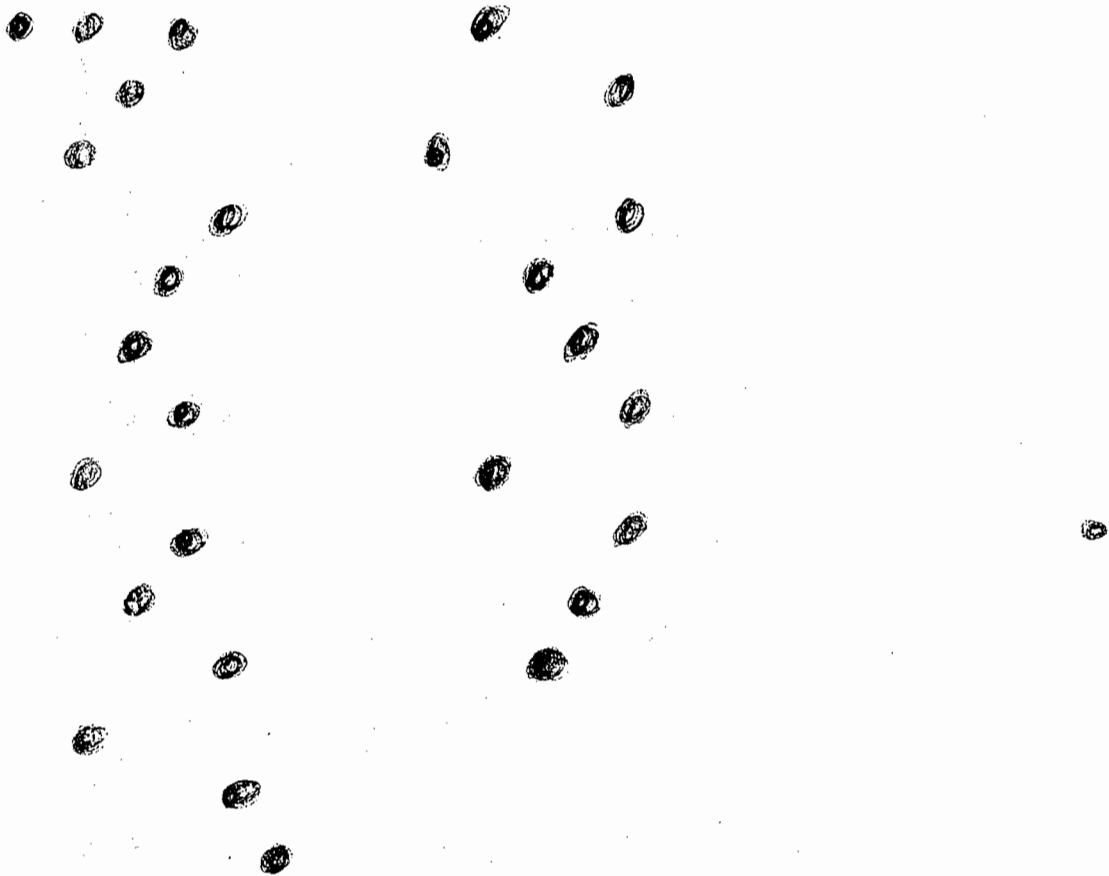
Marque sólo una respuesta para cada pregunta

- |               |               |               |               |
|---------------|---------------|---------------|---------------|
| 1. A B C D E  | 15. A B C D E | 28. A B C D E | 42. A B C D E |
| 2. A B C D E  | 16. A B C D E | 29. A B C D E | 43. A B C D E |
| 3. A B C D E  | 17. A B C D E | 30. A B C D E | 44. A B C D E |
| 4. A B C D E  | 18. A B C D E | 31. A B C D E | 45. A B C D E |
| 5. A B C D E  | 19. A B C D E | 32. A B C D E | 46. A B C D E |
| 6. A B C D E  | 20. A B C D E | 33. A B C D E | 47. A B C D E |
| 7. A B C D E  | 21. A B C D E | 34. A B C D E | 48. A B C D E |
| 8. A B C D E  | 22. A B C D E | 35. A B C D E | 49. A B C D E |
| 9. A B C D E  | 23. A B C D E | 36. A B C D E | 50. A B C D E |
| 10. A B C D E | 24. A B C D E | 37. A B C D E |               |
| 11. A B C D E | 25. A B C D E | 38. A B C D E |               |
| 12. A B C D E | 26. A B C D E | 39. A B C D E |               |
| 13. A B C D E | 25. A B C D E | 40. A B C D E |               |
| 14. A B C D E | 27. A B C D E | 41. A B C D E |               |



PLANTILLA PARA CORREGIR

Nota: -----



## VI.- COMO CALIFICAR

## i) La nota relativa.

Si se supone en un curso que los alumnos se distribuyen normalmente en torno a un promedio se podría pensar en colocar las notas siguiendo la distribución normal de tal modo que haya unos pocos mejores con un 7 la gran mayoría en el 4 y unos pocos con muy bajo rendimiento con un 1.

Con 6 desviaciones standard se barre el 100% del área bajo la curva normal (3 positivos y 3 negativos). Si la curva se divide en tantas partes como notas haya es decir 7 se tiene que a cada parte corresponde  $\frac{6}{7} = 0,857 \approx 0,86$  con lo cual se puede construir la siguiente tabla

Tabla que muestra la relación entre dos límites de la curva normal, los porcentajes y las notas correspondientes:

Límites		% alumnos	Nota
Inferior	Superior		
De - 3,01 a -	2,15	2%	1
- 2,15 a -	1,29	1%	2
- 1,29 a -	0,43	23%	3
- 0,43 a	0,43	34%	4
0,43 a	1,29	23%	5
1,29 a	2,15	8%	6
2,15 a	3,01	<u>2%</u>	
		100%	

ii) Un Ejemplo Práctico para pasar de puntajes a notas.

Examinemos ahora la forma en que podría aplicarse este procedimiento a un ejemplo concreto. Supongamos que se ha aplicado un test de matemáticas a un Primer Año que tiene 40 alumnos. Los puntajes, ordenados en forma decreciente, obtenidos por los alumnos fueron:

73 72 69 67 66 63 59 56 54 54 53 52 52 51 49 48 48  
47 47 46 46 46 45 45 44 42 42 41 40 38 37 35 34 34  
32 32 28 25 23 18.

Debemos ahora calcular la media aritmética ( $\bar{X}$ ) y la desviación standard ( $\sigma$ ) de los 40 puntajes anteriores.

El medio aritmético es

$$M = 46,1 \approx 46$$

y la desviación standard es 13 multiplicando  $\sqrt{\quad}$  por 0,86 (para dividirla en 7 parte y reducirla a puntos) se tiene

$$13 \cdot 0,86 = 11,18$$

aplicando la tabla anterior se tiene:

Límites superiores	Notas
$46 - 2,15 \times 11,18 = 21,70$	1
$46 - 1,29 \times 11,18 = 31,58$	2
$46 - 0,43 \times 11,18 = 41,20$	3
$46 + 0,43 \times 11,18 = 50,80$	4
$46 + 1,29 \times 11,18 = 60,42$	5
$46 + 2,15 \times 11,18 = 70,03$	6

Es decir

entre 18 y 22 puntos:	Nota	1
" 23 y 32 "	"	2
" 33 y 41 "	"	3
" 42 y 51 "	"	4
" 52 y 60 "	"	5
" 61 y 70 "	"	6
" 70 y 73 "	"	7

iii) Otros aspectos de la nota relativa.

Ventajas de la nota relativa.

- Estimula la competencia y puede disminuir la "copia" ya que al ayudar a otro alumno suben los promedios y se autoperjudica.
- Ayuda a discriminar entre alumnos de diferentes rendimiento.
- Evita la influencia de la dificultad de la prueba.

Desventajas:

- Las distribuciones no son usualmente normales, sobre todo en cursos pequeños debe recordarse que la distribución normal es el límite de una distribución binomial con infinitos casos. Por lo demás si hay aprendizaje perfecto todos debieran tener 100%.
- Pueden ocurrir casos extremos tales como un curso excelente o pésimo que a pesar de no obtener un nivel aceptable pueden te-

ner buenas notas.

Un sistema que ha dado mejores resultados es establecer instrumentos standarizados, patrones, usándolos reiterativamente. Así un alumno que obtenga por ejemplo sobre un 80% tendrá nota siete independiente del rendimiento del curso y como se sabe de algunas experiencias usando evaluación formativa se pretende que el 90% del curso llega a cumplir con 100% del aprendizaje lo que daría a la distribución un sesgo altamente negativo.

iv) Las notas, sus limitaciones y su significación.

Hay que tener en cuenta que la asignación de puntajes es generalmente algo muy subjetivo; además los profesores se ven a menudo obligados a juzgar del rendimiento mediante la comparación. Un capítulo aparte significaría analizar los problemas planteados por los efectos de las notas en la personalidad y vida del alumno.

Por todo esto, es importante que los profesores reconozcan la arbitrariedad casi inevitable de los sistemas de calificación: las diferencias entre puntos y décimas de puntos no mejoran las cosas. Muchos se inclinan con razón a usar el criterio de aprobado-reprobado para calificar rendimiento. Como decisión final parece adecuada para expresar el resultado, pero supone de todos modos la formulación y aplicación a ciertos criterios que deben quedar establecidos en alguna parte.

Uno de los inconvenientes de las notas expresadas en números a letras, es que no están definidas significativamente. La significación que se les suele atribuir comúnmente (muy bueno, bueno, regular, malo, etc.), como se ha comprobado, permite una variedad o interpretación aún entre profesores de una misma institución.

Como una línea de solución, se hace necesario encontrar significados más precisos en función de los objetivos que se planean a un proceso docente determinado, donde se establezcan niveles mínimos y grados de resultados posibles en relación con esos objetivos. Se podría proponer una significación de las notas con número que expresen los grados de realización o no realización de los objetivos. Por ejemplo: aprovechando la escala de 2 a 7, las nota podrían significar: (7).

- 7 - Dominio del mínimo de elementos esenciales y el grado más alto de resultados en el desarrollo de los objetivos.
- 6 - Dominio del mínimo de elementos esenciales y alto grado de resultados en el desarrollo de los objetivos.
- 5 - Dominio del mínimo de elementos esenciales y el grado más bajo de resultados en el desarrollo de los objetivos.

- 4 - Dominio del mínimo de elementos esenciales solamente.
- 3 - Casi dominio de elementos esenciales mínimos.
- 2 - Necesita un trabajo considerable para lograr dominio del mínimo de elementos esenciales.

Nótese que las calificaciones bajo la aprobación tienen un sentido indicador para que el alumno mejore su rendimiento.

En este código de significados, es más fácil ponerse de acuerdo sobre el dominio del mínimo de elementos esenciales que sobre el grado (más alto, alto o más bajo) de resultados en el desarrollo de los objetivos. Sin embargo, en la medida en que los objetivos han sido operacionalizados, es posible aplicar estos criterios en forma más adecuada.

## VII.- COMO ANALIZAR LOS INSTRUMENTOS.

Se darán en este folleto algunos de los aspectos más importantes para analizar un instrumento entre ellas se destacan las siguientes.

### i) Confiabilidad:

La confiabilidad mide el rol del factor suerte, es decir si un instrumento está libre de errores por azar. Por ejemplo una regla hecha con un elástico no es confiable ya que el azar influye mucho en la forma que el usuario la aplique.

Se puede también decir que la confiabilidad es el grado en que un instrumento es consistente consigo mismo al medir lo que se quiere medir. Existen tres criterios aparte del "sentido común" para fijar la consistencia interna de un instrumento a saber.

- a) Test-Retest: consiste en aplicar una prueba dos veces a los mismos alumnos, el tiempo transcurrido entre las dos pruebas debe ser adecuado para que no varíen los criterios por nuevo conocimiento de los alumnos lo que ocurre si el tiempo es demasiado largo y no tan corto para que los alumnos recuerden sus respuestas anteriores, Si el coeficiente de correlación de puntajes de una y otra medición es alta el instrumento es consistente es decir confiable.
- b) Formas equivalentes: Se construyen dos instrumentos que son equivalentes en contenido, dificultad por ítem, extensión y forma.



Se aplican simultáneamente o sucesivamente ambos instrumentos.

El coeficiente de correlación alto nos dará la confiabilidad.

c) Mitad-mitad: Se establecen la correlación entre los puntajes de una mitad de los alumnos del curso con la otra mitad (normalmente se usa pares e impares) y a mayor coeficiente de correlación mayor confiabilidad.

d) La fórmula de Richardson.

Una relación bastante simple para calcular la confiabilidad está dado por:

$$r = 1 - \frac{M(n-M)}{n^2}$$

en donde

r = coef. de correlación (ptaje. items-ptaje pba.)

M = promedio notas (o puntajes) total

n = número de items

= desviación standard de las notas.

También puede medirse la confiabilidad de una prueba a través del análisis de item como se verá un poco más adelante.

ii) Validez.

Validez de un instrumento es la forma en que el instrumento mide realmente los objetivos. Por ejemplo una regla perfectamente construida es válida para medir longitudes pero no es válida para

medir el alza del costo de la vida.

Existen varios tipos de validez.

- a) Validez de contenido: los ítem versan sobre los tópicos que se quieren medir.
- b) Validez de construcción: El instrumento debe ser consecuente con la concepción de lo que se quiere medir. Por ejemplo: Debe definirse lo que se entiende por inteligencia antes de construir un test de C.I. que debe ser consecuente con la definición.
- c) Validez de concurrencia: Cuando la información que proporciona el instrumento es confirmado por otro instrumento u otra información que han sido probados. Este tipo de instrumento sirve para tomar decisiones presentes.
- d) Validez predictiva: Si las predicciones realizadas a partir del instrumento son corroboradas posteriormente, quiere decir que el instrumento es confiable y puede continuar usándose. Por ejemplo si existe una alta correlación entre la prueba de aptitud académica y el rendimiento en la Universidad, quiere decir que la P.A.A. es válida.

Los dos primeros tipos corresponden a una Validez de pertenencia lógica y los dos últimos de pertenencia empírica.

Algunas preguntas útiles sobre la validez de los instrumentos son:

- ¿Qué medir por medio de los instrumentos?
- ¿Con qué grado de eficiencia mide una prueba lo que debería medir?
- ¿Constituye la información proporcionada por la prueba el tipo de datos que necesitamos para nuestros fines?

La confiabilidad es un requisito necesario pero no suficiente para la validez de un instrumento. Por ejemplo un profesor se pesa todos los días en una romana del Laboratorio en cuyo manual dice que es solo apto para pesos superiores a 200 Kg. Los pesos del profesor oscilan entre los 300 Kg. y 13 grs. y 300 Kg. y 15 grs., la romana aparentemente es confiable (sólo si se aplica test-retest) pero no es válida, puesto que en las instrucciones dice que no es apto para medir peso inferiores a 200 Kgs.

### iii) El análisis de ítem.

Existen algunas formas para determinar los resultados de la eficiencia de cada ítem, algunos de ellos son:

a) Dificultad de un ítem: Una forma muy sencilla para determinar la dificultad de un ítem es la siguiente.

- 1) Ordene en orden decreciente los puntajes del curso en el test.
- 2) Separe el 25% superior (AS) y el 25% inferior (Ai)

- 3) Sea PS el % de alumnos del cuarto superior (AS) que contestó correctamente el item.
- 4) Sea PI el % de alumnos del cuarto inferior (AI) que contestó correctamente el item.
- 5) La dificultad de un item se calcula como

$$D = \frac{PS + PI}{2}$$

- 6) a menor D mayor dificultad. Ejemplo el item 4 lo contestaron el 35% de los alumnos de puntaje alto y un 5% de los alumnos de puntajes bajos la dificultad del item es

$$D = \frac{0,35 + 0,05}{2} = 0,20$$

Se puede decir que el item es más bien difícil.

Usualmente se excluyen de pruebas futuras los items inferiores a 0,20 y superiores a 0,70 por ser muy difíciles y muy fáciles respectivamente.

#### b) Discriminación de un item.

Es la forma como la pregunta ayuda a discriminar entre alumnos exitosos o que hayan tenido un alto puntaje en el test total y los sin éxito o con un bajo puntaje en el test total para lo cual basta con hacer una correlación biserial entre los que acertaron o no en el item (puntaje dicotómico) y el puntaje total del test.

La discriminación también podría hacerse con respecto al promedio anterior del alumno u otros indicadores.

Otra forma de calcular confiabilidad a través de los items es estableciendo una correlación entre los puntajes de los items pares y de los items impares la fórmula de la correlación está dada por

$$r_{pi} = \frac{M_{pi} - M_p M_i}{\sqrt{p} \sqrt{i}}$$

$r_{pi}$  = correlación par impar

$M_{pi}$  = promedio total del curso

$M_p$  = promedio en los items pares

$M_i$  = promedio en los items impares

$p$  = desviación standard par

$i$  = desviación standard impar

Esta relación se corrige por

$$R \text{ corregido} = \frac{2 r_{pi}}{1 + r_{pi}}$$

(fórmula de Sperman Brown)

iv) Corrección de los puntajes de un instrumento.

Para disminuir el "carrileo" o la adivinación en un instrumento se acostumbra a "castigar" al alumno que lo hace disminuyen

do su puntaje final. Del punto de vista del aprendizaje se podría decir que "sabe más" un alumno que conoce su ignorancia que otro que cree saber y está errado, de ahí que sea lógico puntuarlo en forma diferente.

Por otra parte como información también es importante que el alumno discrimine lo que no sabe, esta información le ayudará en su aprendizaje.

El descuento está dado por

$$\frac{E}{K-1}$$

en donde

E = Suma de las respuestas erróneas

K = N° alternativas de cada ítem.

De modo que el puntaje corregido es

$$P \text{ corregido} = \text{Buenas} - \frac{E}{K-1}$$

Ejemplo en un test con ítem de 4 alternativas un alumno obtiene 21 ítem correctos y 9 erróneos y 10 no contesta su puntaje corregido será

$$P \text{ con} = 21 - \frac{9}{4-1} = 24 \text{ puntos}$$

v) El puntaje mínimo de aprendizaje.

Una vez que se ha construido un test se requiere saber cuantos items debe contestar un alumno para que haya aprendizaje, ya que un alumno podría contestar al "azar" cada item y obtener una buena nota.

Si tenemos  $n$  item y siendo  $p$  la probabilidad de acertar en cada item ( $p=1/k$ ) y  $q$  la probabilidad de no acertar ( $q = 1-p$ ) se tiene que con un 5% de confiabilidad, usando la curva del azar (curva normal) que

$$P.M.A. (\text{azar}) = M(\text{azar}) + 1,64 \text{ azar}$$

en donde

P.M.A. = puntaje mínimo de aprendizaje.

$M$  = medio aritmético (o promedio)

$\sqrt{\quad}$  = desviación standard.

1,64 = valor de tabla para el 5%

o bien

$$P.M.A. = np + 1,64 \sqrt{npq}$$

Ejemplo para un test de 2 alternativas y 30 items al 5% de confiabilidad se tiene

$$P.M.A. = 30 \cdot \frac{1}{2} + 1,64 \sqrt{30 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 19,49$$

Esta es: se puede decir con un 95% de certeza que los alumnos que obtengan 20 items correcto o más han aprendido algo.

Esta relación es sólo útil en pruebas de ítem con igual N° de alternativas, en pruebas con ítem heterogéneos 2 y 4 alternativas por ejemplo o se aplica corrección por azar la relación se complica un poco más.

vi) Instrumentos standarizados.

Cuando un instrumento ha sido aplicado a un curso piloto, cuando se ha realizado un análisis exhaustivo de él, se convierte en un instrumento patrón o standarizado, lo que le da mucho mayor jerarquía. Un caso típico es la prueba de aptitud académica.

Cuando un instrumento se ha standarizado en otro país o con una muestra muy diferente se debe validar con la nueva muestra.

Algunas Características son:

- tienden a dar más importancia a las aptitudes y habilidades intelectuales que a la memorización de hechos (pruebas de rendimiento)
- abarcan los principales conceptos comunes a los planes generales de un país o de una parte considerable de un país, y no se limitan al contenido específico de un currículum o programa local (de rendimiento)
- están redactados por personas muy versadas en los procedimientos necesarios para construir pruebas y además muy idóneas en los temas correspondientes.
- se suelen preparar en dos o más formas semejantes.



- las normas que acompañan a las pruebas standarizadas dan una base para comparar el rendimiento de grupos locales con el de los grupos que representan la población nacional
- como todas las secciones que componen una batería de pruebas o tests de rendimientos o ejecución han sido standarizados sobre la misma población, se puede trazar un gráfico para cada subprueba con las normas de la población. Esto servirá para dar una indicación sobre sus puntos fuertes y débiles con respecto al promedio de la población que sirvió para fijar las normas.

nota: una batería consiste en un grupo de pruebas de rendimiento, relativamente cortas, todas las cuales se han standarizado con la misma población. (referida a la categoría de rendimiento).

VIII.- CONSTRUCCION DE ITEMS PARA LOS DIFERENTES NIVELES DEL  
CONOCIMIENTO

Aunque no existen reglas definitivas al respecto, algunos estudios realizados permiten recomendar distintos tipos de items para las distintas conductas del dominio cognitivo categorizadas en la Taxonomía de Bloom.

Se entrega a continuación una serie de ejemplos que pueden ser de utilidad para el profesor que debe construir instrumentos de evaluación.

1) Para el nivel conocimiento se sugiere.

a) items de completación.

Ejemplo:

Objetivo: El alumno enumerará, sin errores, las tres características del vector fuerza.

Item : Las características del vector fuerza son:

\_\_\_\_\_ , \_\_\_\_\_ , \_\_\_\_\_

nótese que se da una ayuda al indicar con las rayas que son tres.

b) Los items de respuesta breve, también son útiles para el nivel conocer.

Ejemplo:

Objetivo: El alumno será capaz de indicar el rango de valores del coeficiente de roce con error máximo del 5%.

Item: indique que entre que valores varía el coeficiente de roce.  
condición para aceptar la respuesta correcta: 5% sobre los valores máximos o mínimos dados durante la enseñanza

Ejemplo: El alumno describirá las condiciones límites de roce de un cuerpo en el plano horizontal al que se le aplica una fuerza variable. Sólo si aceptaran errores de redacción.

Item: describa en no más de tres líneas las condiciones límites de roce para un cuerpo en el plano horizontal sometido a diversas fuerzas.

condiciones para aceptar la respuesta, debe expresar a lo menos las 3 ideas principales contenidas en el texto programado.

2) Para el nivel comprensión se sugieren diferentes tipos de items según el objetivo, pero en general se usa el item de selección múltiple.

a) item de selección múltiple

Objetivo: El alumno será capaz de interpretar en casos reales los efectos de la corriente inducida. No se aceptarán errores en 4 ejemplos diferentes.

Item : Se induce una corriente "i" moviendo un conductor de longitud  $l$  en una zona donde existe un campo  $B_0$  el campo magnético inducido  $B_i$  tiende a oponerse al movimiento.

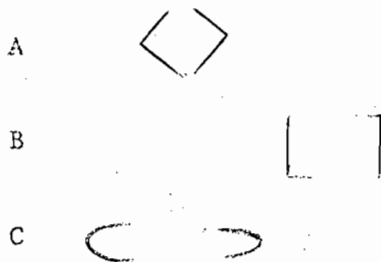
- cual de las siguientes afirmaciones interpreta mejor esta situación.

- A Una bobina que atrae a un clavo
- B Un generador
- C El movimiento de la aguja de un compás
- D Un timbre eléctrico

Nota: este ítem corresponde a uno de los cuatro ejemplos indicados en el objetivo. Se supone además que el ejemplo no ha sido dado durante la enseñanza.

Objetivo: El alumno será capaz de traducir una pregunta lógica de 2 alternativas a un símbolo en el diagrama de flujo.

Item : cual de los siguientes símbolos indican mejor la pregunta.  
¿Es la última tarjeta?



b) El ítem de respuesta breve con gráfico.

Objetivo: El alumno será capaz de traducir en un gráfico cualquier ejemplo de fuerzas (estática) dadas. Se aceptará un error en seis ejemplos diferentes.

Item : dibuje el esquema de fuerzas para un cuerpo que se empuja sobre una mesa de superficie rugosa, considere el peso, el roce, la reacción de la mesa y la fuerza aplicada en cualquier dirección.

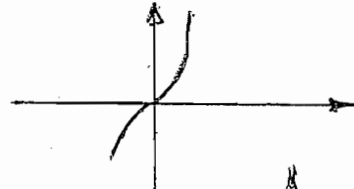
c) El pareo con fórmula o con gráficos también es útil para medir traducción o interpretación.

Objetivo: El alumno será capaz de discriminar la curva correspondiente a cada uno de las 15 funciones elementales dadas:

Item: indique en su hoja de respuestas la alternativa que corresponda a cada uno de los números indicados, indicando la curva que mejor represente la ecuación.

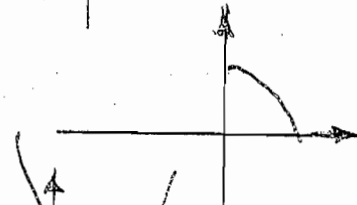
A  $3x + 25y = 0$

51)



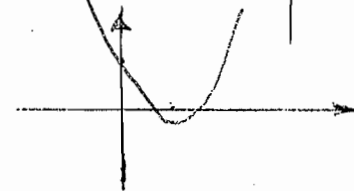
B  $4\text{sen}x = 0,7$

52)



C  $2x^2 + 6x + 10 = 2y$

53)



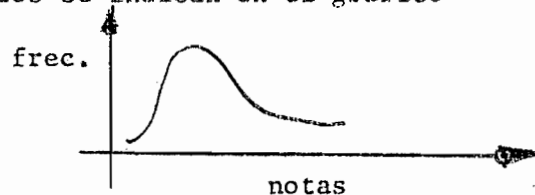
D  $\text{tg}x = 43$

E Ninguna

d) También puede usarse, el ítem V - F usando un gráfico o una tabla y estableciendo una o más conclusiones.

Objetivo: Los alumnos serán capaces de interpretar correctamente un gráfico con los resultados de un test no se aceptaran inferencias erradas

Item : Se aplica un test objetivo en un curso los resultados obtenidos se indican en el gráfico



indique si las conclusiones basadas en el gráfico son verdaderas o son falsas

	Verdadero	falso
- Los alumnos estaban muy mal preparados	- - - -	- -
- En opinión de los alumnos la prueba fue fácil	- - - -	- -
- El profesor está desconforme por el bajo rendimiento del curso	- - - -	- -

### 3) Para el nivel de Aplicación.

Usualmente en la categoría aplicación se distingue la aplicación de principios y la aplicación de generalizaciones a casos concretos.

Para que la conducta expresada por el alumno sea realmente aplicación la situación de evaluación debe ser nueva, pero debe estar al alcance de los alumnos.