

**Universidad de Talca**

**Instituto de Investigación y Desarrollo Educativo**

# Formatos de ítemes para el dominio cognitivo del aprendizaje

Prof. Gustavo Hawes B.

2005

## INDICE

Presentación

### **ITEMES DE RESPUESTA ABIERTA.**

1. Descripción general.	5
2. Ítemes de ensayo con respuesta extensa.	5
3. Ítemes de ensayo de respuesta restringida.	6
4. Ventajas relativas del ítem de ensayo.	7
5. Inconvenientes relativos de la prueba de ensayo en cuanto a los ítemes.	8
6. Inconvenientes relativos de la prueba de ensayo en cuanto al instrumento.	8
7. Sugerencias para la construcción de ítemes de ensayo: en cuanto a la construcción.	9
8. Sugerencias para la construcción de ítemes de ensayo en cuanto a la redacción y presentación.	9
9. Reglas generales para la calificación y puntuación: procesos previos	10
10. Reglas generales para la calificación y puntuación: proceso de puntuación.	10
11. Casos especiales.	10

### **II. ITEMES DEL TIPO VERDADERO-FALSO.**

1. Descripción.	11
2. Ventajas de los ítemes del tipo V-F.	12
3. Inconvenientes.	13
4. Aspectos relacionados con la construcción.	14

### **III. ÍTEMES DE SELECCION MUL TIPLE.**

1. Descripción general.	15
2. Principales ventajas relativas del ISM.	17
3. Principales desventajas relativas del ISM.	18
4. Construcción y diseño de los I.S.M.	19

### **IV. ÍTEMES DE APAREAMIENTO.**

1. Descripción.	20
2. Ventajas relativas de los ítemes de apareamiento.	23
3. Desventajas relativas de los ítemes de apareamiento.	23
4. Aspectos relacionados con la construcción.	24

## **PRESENTACIÓN.**

Este documento ha sido elaborado para apoyar el trabajo de docentes que no poseen formación pedagógica profesional. Ello explica el carácter eminentemente directo y aplicado de su contenido.

Como puede advertirse, este texto se caracteriza porque está referido de manera exclusiva a ítemes del tipo “papel y lápiz”, no considerando ítemes o reactivos de ejecución, tales como exámenes orales o bien ejecuciones propiamente tales.

A la vez, su referencia principal –y prácticamente exclusiva- es al ámbito de lo cognitivo, enfatizando tanto los saberes declarativos (componentes de conceptos, principios, hechos) como de procesos intelectuales (comprensión, análisis, síntesis, evaluación), al tenor de la taxonomía de los objetivos educacionales desarrollada por Bloom y Krathwohl (1956). Esto implica que hay una multitud de otros comportamientos a nivel individual (afectivos, valóricos, inter e intrapersonales, etc) que no pueden ser observados a través de dispositivos como los que se presentan aquí.

## **I. ÍTEMES DE RESPUESTA ABIERTA.**

### **1. Descripción general.**

Los ítemes de respuesta abierta, también denominados "de ensayo", se caracterizan por medir el aprendizaje del alumno según éste produzca o suministre una respuesta, en lugar de identificarla y marcarla, como es el caso de los ítemes de respuesta cerrada.

Se distinguen en esta familia tres tipos de ítemes: de COMPLETACION, de ensayo con RESPUESTA RESTRINGIDA, y de ensayo con RESPUESTA EXTENSA.

### **2. Ítemes de ensayo con respuesta extensa.**

La pregunta de ensayo de respuesta extensa (a diferencia de la de respuesta restringida que se verá más adelante) tiene la propiedad de no limitar explícitamente el campo de acción dentro del cual el alumno puede actuar para proporcionar la respuesta. Este amplio espacio de libertad permite al alumno desarrollar y mostrar su capacidad para seleccionar información o hechos que él considere pertinentes a lo requerido, a fin de organizar la respuesta de acuerdo a su propio criterio (en el caso más extremo), e integrar y evaluar las ideas según él cree apropiado, dentro del contexto en que se plantea la respuesta.

Esta misma dependencia del alumno en relación a su criterio hace a este tipo de ítem relativamente ineficaz para la medición de productos más específicos de aprendizaje, introduciendo obviamente dificultades para la calificación. Esto restringe fuertemente ciertos usos que se pueden dar a los resultados de la evaluación.

En cuanto a la calificación, ésta normalmente se hace sobre una base dicotómica de correcto / incorrecto. No obstante, pueden presionar sobre el evaluador aspectos a veces no considerados entre los elementos de criterio como, por ejemplo, grado de lucidez en la exposición, buen manejo del recurso lenguaje en la redacción de la respuesta. Al respecto, indica Gronlund (1965) que "las calificaciones de las respuestas extensivas es tan poco confiable que tales preguntas no deben usarse para finalidades de medición sino como dispositivos para la enseñanza". (pp. 250).

Siendo razonable la posición de Gronlund citada, es discutible, toda vez que se piensa que su enfoque básico es el del psicometrista, cuyo interés primordial está puesto sobre aspectos tales como confiabilidad y validez. Sin embargo, desde el punto de vista edumétrico, el problema que representa la confiabilidad de los resultados de las pruebas deja de tener el peso que se le atribuye en el contexto psicométrico, adquiriendo la máxima relevancia el aspecto relacionado con la validez de contenido de las preguntas. Si a lo anterior se agrega la concepción de la evaluación como instancia instruccional u oportunidad de aprendizaje, más que como instancia verificadora y sumativa, entonces la prueba de ensayo de respuesta extensa tiene un potencial mucho mayor que el asignado por el autor .

Algunos aspectos de los tipos de productos de aprendizaje que son susceptibles de medirse mediante las pruebas de ensayo de respuesta extensa son:

- capacidad para analizar una comunicación, distinguiendo sus partes componentes, las interrelaciones entre las partes, su principio organizativo
- capacidad para producir, organizar, y expresar ideas;
- capacidad para integrar aprendizajes de áreas diferentes; capacidad para crear formas originales;
- capacidad para evaluar críticamente las ideas.

En consideración al alto nivel de desarrollo cognitivo que implican las conductas más susceptibles de medir con este tipo de ítem, parece razonable pensar que su uso debería plantearse más bien con alumnos mayores.

### **3. Ítemes de ensayo de respuesta restringida.**

El ítem de ensayo que implica una respuesta restringida o corta contribuye a minimizar algunas de las debilidades de la pregunta de respuesta extensa.

#### Ítem de completación

Una restricción máxima la hace eficiente para la medición de conocimientos de hechos y reduce fuertemente la dificultad para calificar. Es el caso del ítem de COMPLETACION.

En este tipo de ítem, la tarea estructurada más finamente que se presenta al estudiante, la hace menos efectiva en cuanto instancia para la medición de la capacidad para seleccionar, organizar e integrar ideas, la que es precisamente uno de los propósitos únicos y fundamentales para los que se usa la prueba de ensayo.

La pregunta de respuesta restringida establece límites explícitos en cuanto a la extensión y género de las respuestas que son requeridas al alumno. Además, es más corta y posee una direccionalidad definida por las mismas especificaciones de la pregunta. Al definir con mayor precisión el problema, el estudiante tiene la ventaja de poder demostrar ciertos tipos definidos e identificables de aprendizaje explícitos.

El ítem de respuesta restringida tiende a limitar tanto el contenido como la forma de la respuesta del estudiante. El CONTENIDO se limita por la especificación y restricción del alcance del tópico de la materia que se evalúa. La FORMA de las respuestas del estudiante está limitada por la formulación misma de la pregunta, la que trae explicitadas las especificaciones de requerimiento.

La pregunta de respuesta restringida es más útil para la medición de productos de aprendizaje que requieren interpretación y aplicación de los datos en un área específica de la materia. Sin embargo, no se la puede identificar con el "ejercicio interpretativo", por cuanto éste tiende a medir el nivel de destrezas ejecutadas sobre una base de información dada explícitamente, preferentemente a través de ítemes de respuesta cerrada.

Algunos ejemplos de los tipos de productos de aprendizaje que son susceptibles de medirse a través del ítem de respuesta restringida (Gronlund, 1965), son:

- explicar relaciones de causa-efecto;
- describir aplicaciones de los principios;
- presentar argumentos pertinentes;
- formular hipótesis sostenibles;
- formular conclusiones válidas;
- expresar los supuestos necesarios;
- describir las limitaciones de los datos;
- explicitar métodos y procedimientos; y
- en general, productos similares que estén basados en la capacidad del alumno para suministrar o producir la respuesta.

#### **4. Ventajas relativas del ítem de ensayo.**

Constituye el método más apropiado y directo para examinar cierto tipo de aprendizajes. En algunos objetivos, los requerimientos del instrumento de ensayo pueden resultar casi idénticos al comportamiento crítico (máximo) supuesto por el objetivo. Al proporcionar una medida directa de los productos complejos del aprendizaje, que no pueden medirse por otros medios, la prueba de ensayo se convierte en un recurso insustituible en la evaluación del aprendizaje. No obstante, se debe tener sumo cuidado en su planeación y redacción .

La prueba de ensayo constituye una base adecuada para lograr inferencias en cuanto al nivel de competencia del alumno respecto a capacidad para definir términos, para explicar claramente los procedimientos para llevar a cabo una demostración, etc.

Puede estimular al alumno a estudiar más significativamente, en el uso de métodos de estudio superiores para la preparación a la prueba. Parece haber una relación positiva entre el uso de pruebas destinadas a medir los logros más importantes en la educación y la tendencia en los estudiantes a estudiar más eficaz y significativamente (Goring, 1971 ).

Da al alumno la oportunidad de poner en práctica destrezas gramaticales y de composición. En este sentido, la prueba de ensayo es la única forma de medir el aprendizaje en gramática, ortografía y composición. No obstante, no es tan evidente que sea una buena forma de medir la originalidad, en particular considerando las condiciones de tensión a que está sometido el alumno por el hecho mismo de tratarse de una prueba.

Da al alumno la posibilidad de crear y ser original, afirmación que posee las limitaciones indicadas en el punto anterior .

La prueba de ensayo ofrece mayor economía de tiempo en su preparación en lo referente a tiempo que insume. Sin embargo, la economía de tiempo no es tan evidente cuando se pretende una buena redacción (cuestión que, por lo demás, está relacionada con el tiempo

insumido en la corrección y calificación de las respuestas). De acuerdo a Goring (1971) una prueba de ensayo de respuesta restringida con 10 preguntas insume al alumno tanto tiempo como para contestar 50 ítems de una prueba de respuesta cerrada. Además, la puntuación de las pruebas de ensayo exige muchísimo más tiempo que las pruebas de respuesta cerrada .

## **5. Inconvenientes relativos de la prueba de ensayo en cuanto a los ítems.**

Las preguntas de ensayo, particularmente las de respuesta extensa, pueden ser mal interpretadas por el estudiante. Aun cuando ello puede suceder también en un instrumento de respuesta cerrada, el riesgo es menor por el cuidado en la preparación. Además, la mala interpretación de un ítem en una prueba de ensayo trae mayores consecuencias por cuanto, debido a su escaso número, cada una de ellas constituye una proporción mayor sobre el total del puntaje.

La corrección, evaluación, asignación de puntajes a los resultados de los ítems en una prueba de ensayo, consumen un tiempo excesivo al evaluador, lo cual puede incidir en el proceso mismo, especialmente por el efecto de la fatiga o del olvido de los criterios establecidos (por lo general implícitamente) para la asignación de puntaje a las respuestas.

El ítem de ensayo tiende a ser calificado sin ayuda de una clave de puntuación por respuesta previamente establecida y adecuada. En consecuencia, las puntuaciones que reciben los alumnos se ven afectadas particularmente por:

- el efecto de halo (positivo o negativo) de los resultados alcanzados por la pregunta anterior de la prueba, sobre la evaluación de la(s) respuesta(s) siguiente(s);
- características de la expresión escrita del alumno: caligrafía, ortografía, sintaxis, efectividad de la redacción, etc., todo lo cual influye en cierto modo en el evaluador

## **6. Inconvenientes relativos de la prueba de ensayo en cuanto al instrumento.**

La prueba de ensayo puede ser poco confiable en sus resultados, especialmente por:

- lo limitado de la muestra de conocimientos o destrezas intelectuales que puede abarcarse con unas pocas preguntas. Al reducirse el número de ítems, se reduce su cobertura y, por lo tanto, la probabilidad de que la prueba sea una muestra representativa de las conductas del estudiante;
- la subjetividad de la puntuación, debido fundamentalmente a la subjetividad contenida en el valor que el docente asigna a cada respuesta, como asimismo por el efecto de halo que puede producirse.

En consecuencia con lo anterior, los resultados también pueden ser poco válidos, debido a:

- lo limitado de la muestra de comportamientos;
- la baja confiabilidad, según se anotó en el punto anterior .

La prueba de ensayo exige a los alumnos dedicar demasiado tiempo al examen. En el caso de que el tiempo esté fijado de antemano, ello incide sobre los alumnos de ritmo más bajo, aumenta la tensión y, naturalmente, incide sobre la calidad de las respuestas.

La prueba de ensayo tiende a ser construida de prisa. Por lo general, se trata de un conjunto de preguntas que el profesor improvisa pocos minutos antes de tomar el examen. Como consecuencia de ello:

- las preguntas son ambiguas;
- pueden ser tan generales que el alumno puede salirse del tema o bien dedicarse a improvisar;
- el nivel de dificultad para cada pregunta es irregular porque no ha sido ponderado previamente;
- la selección de las preguntas no es representativa de los conocimientos o conductas esenciales.

### **7. Sugerencias para la construcción de ítemes de ensayo: en cuanto a la construcción.**

La primera regla es restringir el uso de las preguntas de este tipo a aquellos comportamientos producto del aprendizaje que no pueden ser medidos satisfactoriamente mediante instrumentos más estructurados. Esto es, al menos las preguntas de respuesta extensa deben reservarse para medir la producción de respuestas originales y complejas.

Los tipos de pregunta (de respuesta restringida o extensa) deben ser determinados de acuerdo a los productos específicos de aprendizaje que se quiere medir (isomorfismo con los objetivos instruccionales).

Definir con precisión para cada pregunta la dirección y finalidad de las respuestas que se requieren. Esto es, la pregunta debe especificar lo más exactamente posible la naturaleza de la tarea requerida al alumno.

Indicar para cada pregunta el límite o cantidad aproximada de tiempo que deberá dedicar a la respuesta. La suma no debe exceder el tiempo total destinado a la prueba.

### **8. Sugerencias para la construcción de ítemes de ensayo en cuanto a la redacción y presentación.**

Es preferible la formulación de varias preguntas de tipo restringido a unas pocas de tipo respuesta extensa, siempre que las conductas a medir se presten a ello. Siempre que sea posible, por tanto, hay que especificar al máximo la tarea requerida al alumno.

Evitar el uso de preguntas opcionales, por ejemplo, de cinco contestar tres a elección. Con ello, la baja capacidad de muestreo termina por anularse, porque se tendrían tantas pruebas como combinaciones pudieran formularse.

Indicar el valor en puntos de cada ítem o pregunta.

Advertir a los alumnos en el caso de que aspectos tales como ortografía, redacción, caligrafía, sintaxis, etc., tengan alguna influencia en la puntuación de las respuestas.

En lo posible, someter la redacción previa de las preguntas a la opinión de otras personas entendidas en la materia, como evaluación previa del instrumento.

### **9. Reglas generales para la calificación y puntuación: procedimientos previos.**

Preparar de antemano un esbozo de la respuesta que se espera del alumno.

Preparar una hoja de cotejo, o con escalas de calificación, para hacer posteriormente la puntuación, siempre que el método de calificación por puntos sea el más adecuado.

Decidir el método de calificación más adecuado:

(a) por puntos: asignar puntaje a las respuestas según el patrón establecido previamente;

(b) por tasación: ordenar las hojas de las pruebas según el grado de calidad estimada de la respuesta y luego asignar puntajes según la distribución ordinal.

Optar por una decisión en cuanto a factores irrelevantes para el objetivo de la prueba. Pueden ser, por ejemplo, caligrafía, sintaxis, etc., siempre que ello sea pertinente.

### **10. Reglas generales para la calificación y puntuación: proceso de puntuación.**

Evaluar todas las respuestas a una pregunta determinada en todas las hojas de prueba antes de pasar a la siguiente.

Evitar conocer el nombre del alumno al cual pertenece la prueba que se está calificando.

### **11. Casos especiales.**

En el caso de que los resultados de las pruebas tengan proyecciones excepcionalmente importantes para los alumnos (por ejemplo, ser aceptado o no a una beca), es recomendable obtener dos o más calificaciones independientes de parte de expertos en la materia.

## II. ÍTEMES DEL TIPO VERDADERO-FALSO.

### 1. Descripción.

El ítem llamado de "verdadero-falso" (VF) es un tipo del ítem de respuesta estructurada, escrito, que presenta un enunciado que el sujeto debe calificar sobre una escala dicotómica como, por ejemplo, verdadero-falso, correcto-incorrecto, sí-no, acuerdo-desacuerdo. En consecuencia, debe notarse que aunque se llame VF, su ámbito es más amplio.

Formalmente, se trata de un enunciado gramaticalmente correcto y semántica mente significativo (no vacío), para el cual se ofrece una respuesta dicotómica en la que las clases de la dicotomía son exhaustivas y complementarias.

Que el enunciado sea gramaticalmente correcto significa que el ítem, en cuanto enunciado, debe poseer algunas características básicas como, por ejemplo, constar de sujeto, verbo y predicado; consistente en cuanto a número, género y persona; etc. A la vez se exige que posea significado no sólo en sí mismo o para el redactor, sino que para quienes se espera que lo lean.

En cuanto a las opciones de respuesta, éstas pueden ser básicamente dos: en el sentido más amplio, una de las opciones a lo menos debe llevar asociado un juicio como, por ejemplo, "verdadero", "correcto", "humano", etc. En esta perspectiva, la alternativa es complementaria en el mismo sentido de lo anterior: "no verdadero", "no correcto", "no humano", etc. Asimismo, no debe existir una tercera respuesta posible, a menos que la ausencia de respuesta sea considerada como una forma de respuesta. Esta acepción dicotómica básica es posible ampliarla hacia esquemas clasificatorios más amplios, como se ilustra más adelante.

El ámbito de aplicación y uso es amplio. Por ejemplo, se los puede aplicar en campos de respuesta tales como:

(A) VERDADERO - FALSO.

*"Indique si el enunciado a continuación es verdadero o falso:*

*( ) La acción de la tragedia Hamlet, de Shakespeare, se desarrolla en el castillo de Elsinor, en Dinamarca "*

(B) CORRECTO - INCORRECTO .

*"Indique si el enunciado a continuación es correcto o incorrecto desde el punto de vista gramatical:*

*( ) Durante el terremoto pasado hubieron muchas personas que se desmayaron. "*

(C) OPINION - HECHO.

*"Indique si los enunciados a continuación corresponden a opiniones o si representan hechos:*

*( ) Las hormigas son insectos himenópteros.*

*( ) Las mujeres no deberían estudiar carreras de ingeniería. "*

**(D) ACUERDO - DESACUERDO.**

*"Indique si está de acuerdo o en desacuerdo con los enunciados siguientes:*

*( ) Algunas carreras, como Obstetricia o Educación Parvularia, deberían ser exclusivamente femeninas.*

*( ) El derecho a voto solamente debería corresponder a los mayores de 21 años que sepan leer y escribir.*

**(E) Clasificaciones.**

*"Indique para cada uno de los objetos a continuación si corresponden al reino Animal, Vegetal, Mineral, o Moneras:*

*Agua*

*Tiburón*

*Ameba*

*Alga."*

## **2. Ventajas de los ítemes del tipo v-f.**

Este tipo de ítem posee algunas ventajas importantes que hacen adecuado su uso para medir cierto tipo de conductas. Entre éstas se encuentran las siguientes:

Pueden ser contruidos rápidamente. Como se trata de un único enunciado por lo general, presenta la ventaja relativa de la rapidez con que pueden construirse. En muchas ocasiones, dichos enunciados pueden extraerse directamente de textos o bien mediante variaciones sistemáticas de uno o varios de los elementos constitutivos del enunciado.

Son fáciles de calificar: la existencia de una única respuesta correcta proporciona no solamente velocidad a quien califica, sino que además puede ser realizado por cualquier persona, aunque no domine la materia de la prueba, siempre que disponga de una clave de las respuestas. Con todo, esta facilidad de la calificación puede decirse solamente de aquellos ítemes que apuntan directamente a aspectos cognoscitivos (como verdadero-falso, o correcto-incorreto). Tratándose de ítemes que abordan dimensiones afectivas, creencias,

actitudes, valores, etc., la situación cambia, ya que en este ámbito no es posible hablar de respuestas correctas o incorrectas.

Por lo general, son comprensibles para los alumnos que tienen destrezas lectoras limitadas. Esta ventaja es relativa, ya que se presenta solamente en referencia a ítemes que implican una mayor cantidad de lectura como, por ejemplo, los ejercicios interpretativos. Por otra parte, está condicionada por la redacción del ítem: un enunciado puede estar redactado en forma tan compleja o utilizando un vocabulario tan sofisticado que, incluso para sujetos de buena capacidad lectora, puede representar una dificultad adicional.

Son el mejor tipo de ítem cuando se trata de enunciados o preguntas para las que solamente hay dos respuestas alternativas. En efecto, se hace difícil y artificioso intentar generar más de dos opciones de respuesta cuando realmente el enunciado no da para más. Además de exigir una cantidad de trabajo mucho mayor en la construcción, crear más respuestas artificialmente no proporcionará al evaluador un mejor conocimiento de las destrezas o conductas del alumno.

Su campo de aplicación es amplio en cuanto a conductas. Pueden usarse para medir, entre otras, conductas tales como

- actitudes, creencias, opiniones, apreciaciones valóricas; \* habilidad para identificar la corrección de enunciados;
- habilidad para reconocer relaciones de causa y efecto;
- habilidad para distinguir entre hechos y opiniones.

Permite una amplia capacidad de muestreo de contenidos de conocimiento .

### **3. Inconvenientes.**

Las mayores desventajas que presenta la utilización de este tipo de ítem son:

Son difícilmente utilizables para medir aprendizajes complejos. En el caso del dominio cognoscitivo, su utilización se encuentra restringida fundamentalmente a la categoría de Conocimiento Simple (B.S.Bloom). Intentar construirlos para medir conductas más complejas como Comprensión, Análisis, etc., presenta tan serias dificultades que, por lo general, llevan a optar por otro tipo de ítem.

Es difícil construir ítemes de este tipo que no sean triviales, ambiguos o susceptibles de interpretación. Evitar la trivialidad solamente puede lograrse incorporando nuevos elementos al enunciado. No obstante, ello lleva al otro extremo de la ambigüedad, o bien abre la posibilidad para más de una interpretación, con lo cual el ítem deja de ser el mismo para todos quienes lo lean: podrá haber más de una lectura del mismo enunciado.

En el caso de la medición de conductas cognoscitivas, estos ítemes no son adecuados para contenidos que no tengan respuestas verdaderas y falsas, correctas o incorrectas, claramente distintas.

Su valor diagnóstico es limitado. El hecho de que haya sólo dos respuestas alternativas

permite detectar si el sujeto se equivoca o no, pero, en caso de error, no es posible dónde reside la equivocación. Solamente ítemes que comprendan esquemas de respuesta más complejos o más amplios permiten diagnosticar la fuente y naturaleza de los errores de los sujetos.

Su uso exclusivo o muy reiterado estimula en los alumnos la memorización mecánica simple, sin incentivar el desarrollo de destrezas intelectuales superiores.

Estimulan la copia y la adivinanza. En cuanto a la copia, la situación es similar a la de los otros ítemes de respuesta cerrada: el hecho de estar impresos en un mismo formato facilita la ubicación de las respuestas de un alumno a otro durante el desarrollo de la prueba. También posibilita el desarrollo de códigos no lingüísticos a fin de comunicarse las respuestas. Por otra parte, el hecho de que haya solamente dos respuestas posibles, estimula la adivinanza, ya que la probabilidad de acertar adivinando de 0.5, o sea, de un 50%.

Estimulan la generación del "patrón de respuestas" o forma sistemática de responder cuando no se conoce la respuesta. De esta forma, un sujeto puede responder sistemáticamente a todos los ítemes como "V" . Si existe igual número de ítemes verdaderos que de falsos, entonces por este simple recurso ya se asegura la mitad del puntaje. Otro patrón de respuesta es contestar alternativamente: V - F - V - F - . . . .

#### **4. Aspectos relacionados con la construcción.**

Las principales recomendaciones que deben observarse para el diseño, construcción y presentación de estos ítemes son las siguientes :

Asegurarse de que el enunciado sea completamente verdadero o completamente falso. En una situación de ambigüedad, modificar el enunciado hasta que sea posible, o descartarlo si no admite modificación sin caer en la trivialidad .

Evitar el uso de enunciados compuestos. Cuando en un ítem existen dos enunciados conectados por conjunciones como "y" y "o", pueden producirse algunas situaciones como las siguientes, que afectan la calidad del ítem:

- que uno de los enunciados sea verdadero y el otro falso;
- que siendo ambos enunciados verdaderos o correctos, la unión de ambos tenga un valor contrario.

Hay que evitar el uso de palabras inductoras de respuesta del tipo SOLAMENTE, TODOS, NINGUNO, SIEMPRE, NUNCA, términos calificadores que por lo general denotan respuestas falsas. Por otra parte, hay términos como PODRIA, QUIZAS, GENERALMENTE, ALGUNOS, etc., que por lo general denotan respuestas verdaderas.

El uso de enunciados negativos o dobles negativos es poco recomendable, ya que dificultan la lectura y la comprensión. Con ello se estaría midiendo la capacidad lectora del sujeto en lugar de su dominio del material evaluado. Se justifica su uso solamente cuando se pretende medir si el alumno sabe lo que NO es algo. En todo caso, cuando se usan ítemes negativos, es conveniente destacar la palabra "no", subrayándola o escribiéndola con mayúsculas.

Aproximadamente la mitad o menos de los ítemes deben ser falsos o incorrectos.

La presentación de los ítemes debe evitar dar claves a los sujetos. Esto se produce cuando, por ejemplo,

Se ubican alternadamente los ítemes de respuesta V y F; cuando se ubican todos los ítemes de respuesta V al comienzo y luego los F (o a la inversa).

### **III. ÍTEMES DE SELECCION MULTIPLE.**

#### **1. Descripción general.**

El ítem de selección múltiple (ISM) pertenece a la categoría de los ítemes objetivos, esto es, de respuesta cerrada. Para muchos autores este es el tipo de ítem más versátil de su género, por cuanto permite abarcar una gran cantidad de categorías de conducta, no sólo en el dominio cognitivo sino también en el afectivo.

En su formato estándar, se compone de dos elementos :

- un CUERPO, enunciado que presenta una cantidad de información y un problema que el sujeto debe resolver, y
- una lista de OPCIONES o posibles respuestas para el problema planteado.

El sujeto debe elegir entre las Opciones aquella que represente la respuesta correcta o la más apropiada, o la que mejor refleja su propia opinión. En el plano cognitivo, cuando es posible hablar de respuestas correctas o incorrectas, las opciones que corresponden a las respuestas incorrectas se denominan DISTRACTORES. Su función es precisamente distraer a aquellos sujetos que no están plenamente seguros de la respuesta correcta. Tienen una importante función en el enfoque diagnóstico de la medición, como se verá más adelante.

Existen muchas variantes del ISM. Entre las más corrientes se tiene:

- SELECCION MULTIPLE ESTÁNDAR: todas las opciones son incorrectas menos una (Ver Ejemplo 1)
- SELECCION MULTIPLE REVERSA: todas las opciones son correctas menos una (Ver Ejemplo 2)
- RESPUESTA COMBINADA: junto con el cuerpo se presenta una lista de opciones. Las respuestas sugeridas son combinaciones de éstas (Ver Ejemplo 3).
- ANALOGÍAS: Se dan tres partes de una analogía y el estudiante elige la cuarta, a

partir de las soluciones sugeridas en el listado de opciones (Ver Ejemplo 4) .

**EJEMPLO 1:**  
**SELECCION**  
**MULTIPLE ES-**  
**TANDAR.**

La media aritmética se caracteriza porque:

- (a) divide a la muestra en dos grupos de igual tamaño o mitades;
- (b) representa el valor que más se repite en una distribución;
- (c) hace nula la suma de las desviaciones respecto a ella;
- (d) es el valor más representativo de un conjunto de datos.

**EJEMPLO 2:**  
**SELECCIÓN**  
**MÚLTIPLE RE-**  
**VERSA**

Una de las diferencias más importantes de la Evaluación referida a Criterio (ERC) con respecto a la Evaluación referida a Norma (ERN) es que:

- (a) la ERC apunta a tomar decisiones respecto a los sujetos;
- (b) la ERC tiene un referente externo al grupo;
- (c) la ERC posee un carácter sumativo y calificador;
- (d) la ERC pone énfasis en la validez de los instrumentos.

**EJEMPLO 3:**  
**SELECCION**  
**MULTIPLE**  
**COMBINADA.**

"Son características generales de la Evaluación referida a Criterio las siguientes:

1. La fuente de significación es el dominio
2. Su representatividad depende del grado de congruencia entre ítems y objetivos
3. Sus resultados se distribuyen en un amplio rango de ejecución.

- (a) solamente 1
- (b) 1 y 2
- (c) 1, 2 y 3
- (d) 2 y 3

- EJEMPLO 4. "El viento es al barco como el combustible es a:  
SELECCIÓN  
MÚLTIPLE DE  
ANALOGÍA
- (a) la estufa
  - (b) el automóvil
  - (c) el motor
  - (d) las ruedas"

Estructuralmente, entonces, el ISM se define como un estímulo que se proporciona al sujeto, consistente en dos conjuntos de elementos: (a) un conjunto de información y problemas, y (b) un conjunto cerrado de respuestas posibles y plausibles. El sujeto debe elegir entre las respuestas posibles y marcar una de ellas, no producirla. La elección que haga el sujeto podrá después ser evaluada como "correcta" o "incorrecta" en el caso del dominio cognoscitivo, o bien como "acuerdo-desacuerdo", "aceptación-rechazo" en el dominio afectivo.

## **2. Principales ventajas relativas del ISM**

A pesar de la gran discusión existente en torno a estos ítemes, los ISM poseen algunas indiscutibles ventajas que los hacen muy atractivos para su uso. Entre éstas, se pueden nombrar:

Es posible construir ítemes de alta calidad que midan incluso aprendizajes muy completos. En referencia a las categorías que indica Bloom para el dominio cognoscitivo, los ISM han sido aplicados mayoritariamente a la medición del Conocimiento Simple. No obstante, éste es un defecto que debe atribuirse a los constructores y usuarios de los ISM y no a la naturaleza del mismo. Una construcción cuidadosa permite medir aprendizajes de destrezas intelectuales como Comprensión, Análisis, Evaluación, etc. En el dominio cognoscitivo, el ISM tiene un campo de aplicación máximo. Sucede lo mismo con el dominio afectivo, si bien el potencial del uso del ISM no ha sido explorado en profundidad.

Pueden usarse para evaluar aprendizaje de material no demasiado homogéneo o claramente distinto. A diferencia de los ítemes de apareamiento, no es requisito básico del ISM la homogeneidad del material. Pueden medirse distintas conductas sobre distintos contenidos en una misma prueba.

Esto representa un alto nivel de flexibilidad. El problema de la capacidad de los sujetos parece resistir variaciones en el material que comprende cada uno de los ISM que configuran una prueba, es algo que debe ser tenido en cuenta por el constructor de los ítemes. Muchas veces se decide por la homogeneidad en algún respecto, por ejemplo, en cuanto al tipo de conductas o bien respecto a los contenidos.

Por lo general, se trata de un ítem menos ambiguo que el de VF o de apareamiento. La ambigüedad dice relación principalmente con el conocimiento que obtendrá el evaluador a partir de las respuestas del sujeto. En el ítem VF existe una gran pérdida de información en cuanto a que, si bien el ítem VF dice si el alumno sabe o no sabe, en el caso de no saber carece de poder para identificar dónde radica la falla del estudiante. Por el lado del ítem de

apareamiento, la ambigüedad se presenta en relación a la probabilidad creciente de adivinanza que se produce cuando el sujeto va respondiendo. De esta manera, el evaluador no está en condiciones de saber el peso relativo que posee cada una de las respuestas.

Puede proporcionar valiosa información diagnóstica. En el caso del dominio cognoscitivo como del afectivo, la posibilidad de estructurar situaciones de respuesta que vayan más allá del correcto-incorreto, o del acuerdo-desacuerdo, permite al evaluador conocer las desviaciones de los sujetos respecto a las respuestas correctas, o respecto a los estándares en el caso del dominio afectivo. Ello implica, necesariamente, que los distractores deben estar contruidos de manera que permitan la detección de los comportamientos anómalos, lo que se discute más adelante, en el punto referido a Diseño. En todo caso, el examen de las respuestas incorrectas permite hallar claves acerca de la naturaleza de los errores o desviaciones de los sujetos y, consecuentemente, permite orientar la instrucción ulterior en función de la modificación de dichos errores o desviaciones.

Disminuyen los problemas de "patrones de respuesta". La abundancia relativa de opciones de respuesta (4 o 5 por ítem) disminuye la potencial tendencia de los sujetos a preferir sistemáticamente una respuesta particular cuando no saben la correcta. Esto, que es un defecto serio en los ítemes VF, no se produce en los ISM.

Los efectos de la adivinación son bajos comparados con los ítemes VF. La existencia de mayor cantidad de respuestas posibles produce dos efectos muy importantes:

- en cuanto al alumno, dificulta a éste la adivinación, por cuanto presenta más de dos respuestas posibles;
- en cuanto a los resultados, el peso de una respuesta correcta obtenida por adivinación es mucho menor, porque es menor también la probabilidad de haberla obtenido. En un ítem VF la probabilidad de acertar por adivinación es de  $p = 0.5$ . En un ISM de cuatro opciones, la probabilidad de acertar adivinando es de  $p = 0.25$ .

### **3. Principales desventajas relativas del ISM**

Son los más difíciles de construir. Esta dificultad se presenta en dos niveles:

- en primer lugar, decidir la correcta operacionalización de los indicadores de comportamiento que se pretende medir. Si se busca medir el nivel de "extrapolación" en la conducta Comprensión, entonces deben diseñarse ítemes que consideren ejecuciones congruentes con la conducta;
- en segundo lugar, y asociado a lo anterior, está la dificultad de construir distractores apropiados, plausibles y eficientes. Deben ser apropiados en cuanto a que apunten directamente al comportamiento y contenido involucrados en el objetivo. Deben ser plausibles en cuanto que los distractores tengan una probabilidad de ser elegidos como respuesta correcta. Finalmente, deben ser eficientes en cuanto a su capacidad para discriminar correctamente entre los sujetos que dominan y los que no dominan los resultados de la instrucción.

Su construcción demanda mucho tiempo. Tomar en cuenta lo indicado en el punto anterior implica que el constructor de ítemes debe dedicar una cantidad de tiempo significativa-

mente mayor que si se tratara de ítemes de otros tipo. Esto, en cuanto a la construcción. Ahora, si se tratase de la validación de los ISM, es decir, de obtener la evidencia empírica que sustente la calidad de los ítemes utilizados, solamente se podrá saber una vez aplicada la prueba ya través de análisis estadísticos apropiados.

Le llevan más tiempo al alumno porque demandan más lectura y también exigen que adopte una decisión con más opciones de respuesta. Como consecuencia, la cantidad de ítemes ISM que se puede incluir en una prueba es mucho menor que si se utilizaran ítemes VF, por ejemplo. No obstante, si bien se evalúa menos material, la calidad de los ítemes permite medir conductas complejas, a las que no se puede acceder mediante otro tipo de ítemes, al menos entre los de respuesta cerrada.

Representa un problema adicional para los estudiantes de baja capacidad lectora o sin experiencia con este tipo de ítem. Esto es especialmente importante de considerar cuando se desea construir ISM para alumnos menores o de poco desarrollo lector (por ejemplo, primeros niveles de la educación básica). Podrá suceder que los errores que se adviertan no se deban a falta de dominio por parte del sujeto, sino a que no tuvo tiempo suficiente para leer, a que su capacidad de comprensión lectora se lo impidió, a que la novedad del ítem lo desorientó, etc.

#### **4. Construcción y diseño de los ISM.**

Las principales recomendaciones para elaborar buenos ISM pueden resumirse en los siguientes puntos:

(1) El principal factor para escribir buenos ISM es usar un lenguaje preciso v no ambiguo. Los términos empleados, por tanto, deben ser términos de manejo del estudiante, de forma que el ISM no se convierta en un ítem que mida la capacidad de vocabulario del alumno en lugar de conocimientos o destrezas intelectuales. Como es obvio, esta objeción no corresponde si lo que se evalúa es precisamente el manejo de vocabulario. La mejor manera de plantear directa e inequívocamente las cosas, es hacer una pregunta directa, en lugar de utilizar un enunciado incompleto. Además de centrar con precisión el problema, no obliga al alumno a reconstruir la proposición cada vez que lee una de las opciones.

EJEMPLO: ítem planteado como PREGUNTA.

*¿Qué NO buscan los arqueólogos?*

*(a) nuevas tierras*

*(b) obras de arte*

*(c) testimonios escritos*

*(d) ruinas antiguas.*

(2) Debe evitarse dar claves irrelevantes en los distractores, tales como:

- inconsistencias lógicas y gramaticales;
- haciendo la respuesta correcta consistentemente diferente a los distractores (más larga, más corta, más detallada, siguiendo una pauta, etc)
- asociaciones verbales entre el cuerpo y las opciones distractores Que se superponen: el mismo elemento aparece en todas menos una de las opciones.

(3) La redacción del cuerpo del ítem y de sus opciones debe hacerse en forma positiva. La redacción del cuerpo en términos negativos o doble-negativos solamente viene a entorpecer la lectura y comprensión por parte de los alumnos. Un ítem con esta redacción sería válido si precisamente se tratara de medir la comprensión de lectura en el alumno. El uso de un ítem planteado negativamente se justifica solamente cuando es crucial saber qué NO es algo, o qué NO hacer. En todo caso, debe destacarse la negación, por ejemplo, subrayándola.

(4) Hay que usar mesuradamente los recursos de "Ninguno de los anteriores" o "Todos los anteriores", El "Ninguno de los anteriores" debe utilizarse como opción válida solamente cuando es necesario que el alumno conozca que todos los distractores son erróneos. Por su parte, el "Todos los anteriores" estimula la adivinación, En efecto, basta que un sujeto sepa que hay dos opciones correctas para que infiera que la respuesta correcta del ítem es "Todos los anteriores", sin importar el dominio o conocimiento que tenga de las restantes opciones,

#### **IV. ÍTEMES DE APAREAMIENTO.**

##### **1. Descripción.**

Estos ítemes son miembros de la familia de preguntas objetivas. Consisten en dos conjuntos de elementos de información, normalmente ubicados en columnas en paralelo. El trabajo del estudiante consiste en establecer relaciones de apareamiento entre los elementos de un conjunto con los elementos del otro, sobre la base de instrucciones específicas acerca de los criterios para hacer los apareamientos.

Formalmente, puede describirse el ítem de apareamiento como un conjunto de estímulos para el cual existe un conjunto mayor de respuestas posibles, las que deben ser asignadas a los estímulos de tal manera que para cada estímulo haya a lo menos una respuesta. La acción solicitada al respondiente consiste en establecer una relación de pertenencia (bajo cierto respecto) de a lo menos una respuesta por cada estímulo.

De esta forma, si se definen los siguientes conjuntos y sus elementos:

SISTEMA CIRCULATORIO = {corazón, venas, arterias}

SISTEMA DIGESTIVO = {estómago, hígado, intestino, páncreas}

SISTEMA ENDOCRINO = {hipófisis, suprarrenales, tiroides, gónadas}

puede generarse un ítem de apareamiento como el siguiente.

ITEM: II ASIGNE CADA UNO DE LOS ORGANOS DE LA COLUMNA "A"

<u>-A- SISTEMA</u>	<u>-B- ÓRGANO</u>
	<input type="checkbox"/> GLÁNDULAS SUPRARRENALES
	<input type="checkbox"/> PÁNCREAS
DIGESTIVO	<input type="checkbox"/> ARTERIAS
	<input type="checkbox"/> INTESTINO
	<input type="checkbox"/> HIPÓFISIS
CIRCULATORIO	<input type="checkbox"/> TIROIDES
	<input type="checkbox"/> ESTÓMAGO
	<input type="checkbox"/> HÍGADO
ENDOCRINO	<input type="checkbox"/> GÓNADAS
	<input type="checkbox"/> VENAS
	<input type="checkbox"/> CORAZÓN

El campo de aplicación de estos ítems es bastante restringido. Su uso más frecuente se da en relación al conocimiento simple. Por ejemplo, aparear autores con sus obras; personajes con sus situaciones históricas; países y capitales, etc. No obstante, un examen más cuidadoso puede llevar a plantear una gama de campos de conducta un poco más amplia como, por ejemplo: términos y definiciones; relaciones de causa-efecto; asociación de términos (sinónimos, antónimos) .

Algunos ejemplos...

EJEMPLO 1: El siguiente ejemplo apunta a medir "conocimiento de hechos": *Indique qué obra literaria corresponde a cada uno de los autores de la columna -A-, marcando en el paréntesis al lado de cada nombre en la columna -B- la letra que corresponde.*

*A: AUTORES*

*B: OBRAS*

*a) Cervantes*

*Trafalgar*

*b) Azorín*

*Amor y Pedagogía*

*c) Unamuno*

*El Quijote de la Mancha*

*d) Baroja*

*Camino de Perfección*

**EJEMPLO 2:** El ítem a continuación intenta determinar "relaciones de causa-efecto": *Indique cuáles de los efectos de la columna -B- se deben a cuál causa en la columna -A-, marcando en el paréntesis la letra que corresponde a la causa.*

*A: Causas*

*B: Efectos*

*a) Aumenta la demanda*

*Inflación*

*b) Reemplazo de oro por plata*

*Deterioro de la industria nacional*

*c) Emisión inorgánica de dinero*

*Aumento de los precios*

*d) Liberación de las tasas arancelarias*

*Desaparece elemento de mayor valor*

*Déficit fiscal*

**EJEMPLO 3:** Asociación de términos o sinonimia: *Indique para cada término en la columna -A- su sinónimo entre los términos de la columna -B-, marcando en el paréntesis la letra que corresponde al término de la columna -A-*

*A: Término original*

*B: Sinónimo*

*(a) Grande*

*Extenso*

*(b) Bello*

*Inmenso*

*(c) Amplio*

*Breve*

*(d) Corto*

*Hermoso*

## **2. Ventajas relativas de los ítemes de apareamiento.**

Las principales ventajas que representan este tipo de ítemes son las siguientes:

Son relativamente fáciles de construir, ya que básicamente suponen (a) esquemas clasificatorios o determinar la pertenencia de los elementos Respuesta a los conjuntos indicados en los elementos Premisa; o (b) relaciones de causa-efecto como, por ejemplo, asignar los efectos estipulados en el conjunto Respuestas a las causas detalladas en el conjunto Premisas. De esta manera, se trata de elaborar en primer lugar, el conjunto de los elementos Premisa y, a continuación, establecer un conjunto de elementos Respuesta que sean consistentes con las Premisas establecidas.

Son relativamente fáciles de calificar por cuanto no exigen a quien califica un conocimiento previo que vaya más allá de verificar la corrección o incorrección de la respuesta, sobre la base de una pauta de calificación previamente elaborada y disponible. Es un trabajo que incluso puede hacerse por medios mecánicos o electrónicos.

Son compactos y eficientes, por cuanto permiten evaluar una gran cantidad de material: utilizando poco espacio impreso y, a la vez, exigiendo un tiempo menor al estudiante que si se utilizara otro tipo de ítem. Esta ventaja permite, por tanto, evaluar grandes cantidades de material de aprendizaje.

No es fácil adivinar en este tipo de ítemes, al contrario de los ítemes de V-F. La existencia de dos o más respuestas (entre 8 y 15 según el tipo de alumnos) disminuye fuertemente las probabilidades de acertar por azar. No obstante, la probabilidad de acierto por azar va creciendo a medida que el sujeto va despejando respuestas al asignarlas a las premisas.

Su mejor uso se da cuando la materia a evaluar es homogénea, a fin de medir la habilidad de los estudiantes para aparear términos, hechos, definiciones, eventos, relaciones causales, y otros materiales que impliquen relaciones simples.

## **3. Desventajas relativas de los ítemes de apareamiento.**

Entre las principales desventajas que presentan estos ítemes se cuentan:

(1) Estimulan la memorización mecánica simple por parte del estudiante. Esto no es un problema del estudiante sino que proviene de la misma naturaleza del ítem y de la forma en que se lo utiliza más frecuentemente. Si bien la memorización no es condenable en sí misma como estrategia cognoscitiva, sí lo es un esquema evaluativo que recurrentemente haga alusión al reconocimiento o evocación de contenidos de memoria.

(2) Su uso está limitado a material homogéneo. Este inconveniente deriva del efecto de

respuesta por descarte que implicaría la incorporación de material disímil en las Premisas y Respuestas del ítem. Incluso si se trata de material homogéneo, debe tenerse cuidado de no dar claves, por ejemplo, de carácter gramatical.

(3) Su capacidad de medición no alcanza fácilmente a procesos cognitivos complejos o de orden superior. Básicamente, su campo de acción está referido al ámbito del conocimiento simple, según la acepción de B.S. Bloom.

#### **4. Aspectos relacionados con la construcción.**

Algunas recomendaciones básicas para la construcción y presentación de los ítemes de apareamiento:

Para cada ejercicio, seleccionar solamente material homogéneo y relacionado de alguna forma (lógicamente, estructuralmente, cronológicamente, etc) de tal manera que cada respuesta que se propone sea plausible en relación a cada una de las preguntas. Esto exige tener cuidado en relación al género (femenino, masculino), número (singular, plural), el uso de mayúsculas (todos los elementos inicializados con mayúscula o ninguno de ellos).

Al generar los ítemes debe tenerse cuidado de que el número de elementos Respuesta sea mayor que el de las Premisas. Esto contribuye a eliminar o, más bien, a atenuar los efectos de descarte y adivinación.

Debe haber instrucciones que indiquen claramente al estudiante la base sobre la cual realizar los apareamientos. Debe señalarse claramente si una Respuesta (o una Premisa) puede ser utilizada más de una vez. Obviamente, esto último eleva significativamente la complejidad del ítem, por lo que es recomendable utilizarlo solamente con alumnos mayores.

Debe establecerse claramente la forma de responder: si trazando una línea, que es razonable cuando son pocos elementos pero complica la calificación cuando éstos son numerosos; o bien, escribiendo al lado de la respuesta el nombre de la premisa con que se asocia, usando por ejemplo, la inicial o letra identificadora.

El material del ejercicio debe ser organizado de manera que el estudiante pueda responder eficientemente y, a la vez, obviar el problema del patrón de respuestas. Por ejemplo:

- ordenar las respuestas en orden lógico (alfabético, numérico, cronológico, de mayor a menor, etc)
- ubicar todo el ejercicio en una sola página a fin de evitar que el sujeto pierda tiempo volviendo la hoja y se distraiga

#### Referencias.

Bloom, B., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives. The classification of educational goals. Handbook I: Cognitive Domain*. New York: David McKay.