

## Propuesta de indicadores de calidad para la evaluación de la calidad de las preguntas en los exámenes de opción múltiple

### Presentación:

Uno de los puntos más complicados del proceso de enseñanza – aprendizaje, sea bajo la modalidad presencial o bien con el uso de las Nuevas Tecnologías de Información y Comunicación, es la evaluación. Dicha parte del proceso resulta traumática y ansiógena en los educandos, aunque las nuevas corrientes educativas ven en este proceso el único elemento adecuado y eficiente para la evaluación de la efectividad y validez del método utilizado para la transferencia de conocimientos.

El pedagogo José Gimeno Sacristán, en su libro "Teoría de la Enseñanza y Desarrollo del Currículum" [1], propone considerar a la evaluación como una "comprobación de la validez de las estrategias didácticas", es decir, como el momento de buscar los datos que nos ayuden a decidir si la estrategia metodológica desarrollada fue o no adecuada, o en qué medida lo fue, para guiar un proceso de enseñanza que desemboque en la obtención de los resultados de aprendizaje previamente propuestos.

Desde este enfoque, uno de los grandes valores de la evaluación es el de ser un instrumento de investigación en la didáctica, engranada con el propio desarrollo científico de la teoría y técnica didácticas.

Teniendo en cuenta estos supuestos teóricos, estamos en condiciones de profundizar sobre ciertos aspectos prácticos de la evaluación:

- ❑ Cada instancia de construcción del conocimiento se da al interior de un ESPACIO CURRICULAR que se define por su FORMATO. Entendemos aquí por espacio curricular a un conjunto de contenidos educativos, seleccionados para ser enseñados y aprendidos en un tiempo determinado, y articulados en función de criterios (epistemológicos, psicológicos, pedagógicos, etc.) que le dan coherencia interna.
- ❑ Normalmente se denomina a este espacio curricular con el nombre de “programa” de un curso. Por otra parte, el *formato* es la forma de organización de un espacio curricular que define el tratamiento acorde con sus características y criterios de organización. Por ej. Taller, Módulo, Seminario, Proyecto, Asignatura etc.
- ❑ Los modos de abordaje del objeto de conocimiento (formatos) se particularizan como MATERIAS, ASIGNATURAS, MÓDULOS, PROYECTOS, SEMINARIOS, TALLERES, LABORATORIOS, y esta particularidad plantea también exigencias a las prácticas evaluativas. Las exigencias en la gestión de los espacios curriculares desde el punto de vista de las situaciones de enseñanza plantean condiciones a las prácticas. En otras palabras: Los distintos contenidos deben transmitirse en su formato más adecuado, y dicho formato exige un método de evaluación distinto.

Es por ello que en este trabajo nos proponemos desarrollar el análisis de los distintos formatos con sus características pedagógico - educativas con sus exigencias para las prácticas evaluativas y de manera específica pretendemos establecer un conjunto de parámetros que nos permitan evaluar la calidad de una pregunta de múltiple opción que incluimos en un examen.

## ANÁLISIS DE FORMATOS

### **Formato: ASIGNATURA - MATERIA**

El formato de asignatura – materia, es una **estructura disciplinar** propiamente dicha, y por lo tanto tiene un predominio marcado de la teoría. Sus principales características pedagógico - didácticas son las siguientes:

- ❑ Por una parte la transferencia del conocimiento se plantea desde la transposición didáctica, por la otra, el contenido será separado por el docente en núcleos temáticos definidos en los que primará la unicidad.
- ❑ Dichos núcleos temáticos serán organizados en relaciones jerárquicas, tanto a nivel interno del núcleo como en la presentación de los núcleos entre sí.
- ❑ Las prácticas evaluativas para el formato de asignatura deberán estar orientadas hacia la identificación de conceptos y problemas genuinos de la disciplina, y hacia la construcción y deconstrucción de conceptos y relaciones disciplinares.
- ❑ Es importante la formación no sólo del dominio de la sintaxis entre los distintos conceptos desarrollados, sino en la generación del educando en generar dicha capacidad de sintaxis.
- ❑ Debemos mencionar también la importancia de evaluar el manejo de las herramientas metodológicas propias del sistema epistemológico. Sin dicho manejo mal puede realizarse la transferencia del conocimiento.
- ❑ Deberá también formarse al educando en la toma de posición frente a los marcos teóricos desarrollados, en la fundamentación epistemológica de dicha postura y en la capacidad de defensa de la misma con argumentos convincentes, bien fundados y expuestos de manera clara, coherente y lógica.

Para ello el educando deberá ser capaz de tomar decisiones epistemológicamente bien fundadas.

En resumen, las exigencias para las prácticas evaluativas de los formatos de asignatura son:

- Identificar conceptos y problemas genuinos de la disciplina.
- Construir- deconstruir conceptos y relaciones disciplinares: sintaxis entre conceptos.
- Manejar herramientas metodológicas propias del entorno epistemológico.
- Revisar y tomar posición frente a los marcos teóricos.
- Tomar decisiones epistemológicamente fundadas

Para la acreditación de las asignaturas (aprobación del educando) deberá plantearse por lo tanto evaluaciones de proceso y de resultado.

Es conveniente el plantear una evaluación por cada unidad de aprendizaje. Si ello no fuera posible por la extensión de la unidad, deberá sopesarse la relación costo – beneficio de modificar el planteamiento de las divisiones en unidades realizada de acuerdo a la lógica y a la unicidad de tipo de conocimiento versus la posibilidad de evaluación de unidades completas. En algunos casos será preferible el plantear varias evaluaciones para una misma unidad temática y en otros se optará por la división de la unidad temática en dos o más partes.

### **Formato: MÓDULO**

Para el caso de los formatos del tipo “*módulo*”, las características pedagógico – didácticas de los espacios curriculares, contemplan un problema o tema central que da unidad a los contenidos y a las actividades.

Existirá entonces un hilo conductor que dará unidad a los múltiples abordajes que se realicen del tema. El tema constituirá una unidad de planteo y acreditación en sí mismo.

Las prácticas evaluativas de los módulos exigen el identificar y definir el problema, como así también la línea de abordaje del objeto de estudio.

Deberá evaluarse también la identificación de hilo conductor que se ha utilizado durante el desarrollo del módulo.

Dado que normalmente se realizarán múltiples abordajes al módulo, deberá el educando ser capaz de identificar cada uno de dichos abordajes, elegir el que considere más adecuado tomando posición respecto de ellos, y establecer las relaciones que el abordaje por él seleccionado, tiene con respecto a las otras alternativas posibles.

Deberá también el educando realizar la articulación interna del planteo, y ser capaz de proponer las líneas que se derivan del tratamiento de cada uno de ellos.

En resumen, las prácticas evaluativas para los módulos exigirán:

- Identificar problema o línea de abordaje del objeto de estudio.
- Identificar hilo conductor.
- Revisar múltiples abordajes y tomar posición por una línea estableciendo relaciones con las otras alternativas.
- Identificar articulación interna del planteo.
- Proponer líneas derivadas de tratamiento.

Es conveniente evaluar los módulos con la presentación de un trabajo final individual en el que se propondrá el análisis de alguna de las situaciones desarrolladas en el módulo respectivo, contextualizada en el medio.

Se deberá exigir que los alumnos realicen trabajos de campo a partir de problemáticas reales y no de planteos imaginario o situaciones utópicas o irreales, de manera tal de profundizar la exigencia de soluciones simples, estructuradas y posibles.

### **Formato: PROYECTO**

Para el caso del formato de proyecto, las características pedagógico - didácticas del espacio curricular estarán orientadas hacia la globalización del conocimiento. Será dicha globalización la que permitirá la incorporación de un conjunto de disciplinas que permitan desarrollar dicho proyecto.

Es menester tener en cuenta que la multidisciplinariedad debe realizarse sin pérdida de la identidad disciplinar. Existirá una disciplina a la que el proyecto pertenece, y un conjunto más o menos amplio de otras disciplinas que actuarán como apoyo para el desarrollo del proyecto en la medida que éste lo necesite.

Es importante que el proyecto se realice abordando un problema real que servirá como eje articulador.

El formato de proyecto deberán definirse claramente en sus distintos niveles a saber: **diseño, puesta en práctica y evaluación.**

Las prácticas evaluativas de los proyectos plantearán entonces como exigencia el que el educando sea capaz de identificar claramente el problema y la solución que al mismo planteará a través del proyecto.

Deberá además especificar las distintas estrategias posibles, las características de la estrategia seleccionada y los pasos a seguir en la misma.

En la formulación del proyecto, debe exigirse la especificación de los marcos teóricos atinentes teniendo en cuenta que el correcto abordaje del problema deberá plantearse desde las distintas líneas disciplinares atinentes.

En resumen, las prácticas evaluativas para el formato de proyecto deberán:

- Identificar, formular, y plantear la solución de un problema.
- Manejar estrategias específicas para los objetivos antes propuestos.
- Revisar marcos teóricos para la resolución del problema.
- Articular abordajes desde líneas disciplinares diversas.

Las evaluaciones de proyecto deberán realizarse por medio de trabajo escrito en el que se desarrolle el mismo. Debe exigirse de ser posible, las prácticas de campo o resultados de la aplicación de las soluciones siempre que sea posible.

### **Formato: SEMINARIO**

Los seminarios son investigaciones que se realizan en contextos escolares, cuyo objetivo final puede contemplar desde la producción de un conocimiento nuevo, la realización de un planteo epistemológico actualizado o la simple comunicación de resultados de proyectos realizados.

Las exigencias para las prácticas evaluativas de los formatos de seminarios deberán exigir la definición del tipo de investigación a realizar, y la construcción clara del objeto de estudio, dependiendo de la decisión epistemológica tomada.

No es necesario que la investigación comprenda todos los niveles posibles. Puede limitarse al diseño, ampliarse a la puesta en práctica o desarrollar también la etapa de evaluación.

Deberá el educando en todos los casos construir el marco teórico para la investigación, a través de un planteo epistemológico actualizado, y producir, desde la perspectiva de alumno un conocimiento nuevo. La evaluación deberá realizarse también sobre la forma y claridad de la comunicación de los resultados y la planificación de la transferencia de los mismos.

En resumen, las exigencias para las prácticas evaluativas del formato seminario son:

- Elección del tipo de investigación.
- Construcción del objeto de estudio (según decisión epistemológica).
- Decidir nivel de resolución de la investigación: diseño, puesta en práctica, evaluación).
- Construir el marco teórico para la investigación: planteos epistemológicos actuales.
- Producir un conocimiento nuevo (desde la perspectiva del alumno).
- Comunicar resultados.
- Planificar o ejecutar acciones de transferencia.

Serán evaluados a través de la elaboración de trabajos de investigación sustentados en el desarrollo de problemáticas que surgirán de los marcos teóricos que se aborden.

#### **Formato: TALLER**

Los talleres plantean, como ningún otro formato, una tensión inevitable entre teoría y práctica. Un taller deberá incluir necesaria e imprescindiblemente actividades prácticas claras y completas en su unidad conceptual, pero no podrá desprenderse el educador que realice la evaluación de un taller, de medir y sopesar el nivel de manejo que el educando tiene de la teoría necesaria para realizar la práctica que dicho taller plantea.

Debe tenerse en cuenta también que todo taller debe dar por resultado una producción tangible.

Las prácticas evaluativas de los talleres, deberán concentrarse en la determinación del grado de articulación entre la teoría y la práctica que se realizó para esta instancia del tratamiento de la realidad. No existe punto de tratamiento del conocimiento con mayor aplicación práctica a la realidad que el formato pedagógico de taller.

La articulación entre la teoría y la práctica puede plantearse en cualquiera de las dos direcciones posibles: o bien la teoría se aplica para la resolución práctica de un problema, o bien la resolución de un problema preciso sirve para demostrar los principios involucrados y arribar a conclusiones teóricas.

Exigencias para las prácticas evaluativas de LOS TALLERES:

- Articular teoría y práctica como instancia de tratamiento de la realidad que pueden plantearse en forma bidireccional.
- Alcanzar una producción que puede revestir formas distintas: resolución de un problema, producción de materiales u objetos, diseño de un proyecto.
- Evaluar los procesos y resultados de producción.
- Reflexionar sobre los procesos realizados.

Estos espacios de integración de contenidos conceptuales, procedimentales y actitudinales que apuntan a la resolución de problemas, se acreditarán a través de la producción de materiales y/o la elaboración de proyectos que serán defendidos oralmente. Se tendrá en cuenta además, para la evaluación el proceso de producción y reflexión de los alumnos.

### **Formato: LABORATORIO**

Las características pedagógico- didácticas de los espacios curriculares según el formato de laboratorio incluyen aprendizajes determinados en instalaciones especiales, con equipamientos apropiados e instrumentos tecnológicos y o materiales específicos.

Exigencias para las prácticas evaluativas de LOS LABORATORIOS:

- Articular teoría y práctica como instancia de tratamiento de la realidad que pueden plantearse en forma bidireccional.
- Formular, revisar, comprobar hipótesis formuladas en diversos marcos teóricos.
- Utilizar instalaciones, equipos, y materiales específicos.
- Evaluar los procesos y resultados.
- Reflexionar sobre procesos y resultados desde el punto de vista científico y ético.

A la luz de dicho análisis, es evidente que postular cualquier método evaluativo como universalmente válido, o como el más adecuado para todos los formatos es cuando menos una postura errada.

El presente trabajo entonces se restringirá solamente a los *test* de opción múltiple como parte de los *test* objetivos, considerados como una de las herramientas más adecuadas para la evaluación en los formatos de asignatura – materia y de módulo.

## EVALUACIÓN DE LA CALIDAD DEL SET DE PREGUNTAS:

Una de las tareas más difíciles para el educador que se decide por el *test* de opción múltiple, es la parametrización del examen en sí. Para medir la calidad del examen, deberemos asignar a cada una de las preguntas ciertos indicadores según el resultado que producen cada vez que son utilizadas en un examen. Por lo tanto, la calidad del examen será consecuencia directa del promedio de los guarismos arrojados por los parámetros de cada una de las preguntas que lo componen.

Por lo tanto, consideramos cuatro indicadores matemáticos que se constituyen en parámetros de cada pregunta de opción múltiple. Estos parámetros son: **confiabilidad, objetividad, dificultad y discriminación.**

### 1. CONFIABILIDAD:

Se dice que las puntuaciones de una prueba son confiables cuando aplicadas en diversas oportunidades producen resultados aproximadamente similares. Define a la confiabilidad como *una estimación del grado de consistencia o constancia entre repetidas mediciones efectuadas a los sujetos con el mismo instrumento* (Lafourcade, 1971).

Dado que en el campo de las ciencias de la conducta, es imposible obtener el mismo resultado utilizando el mismo instrumento (someter a un mismo grupo de alumnos a una misma prueba producirá de seguro resultados diferentes, puesto que los individuos no se comportan dos veces de manera idéntica), se expondrán aquí sólo los principios en los que, para Lafourcade, se apoya el concepto de confiabilidad de una prueba. Ellos son:

#### 1.1 El número de ítems incluidos en la prueba

Si el número de reactivos es bajo, hay mayores posibilidades de que los resultados no sean confiables. Un mayor número de reactivos en general posee una muestra mayor de lo que se intenta medir, y tiende a cancelar los efectos del azar. La tabla de Ebel ilustra esta relación entre la longitud del *test* y su confiabilidad.

#### TABLA DE EBEL

REACTIVOS	CONFIABILIDAD
5	0,2
10	0,33
20	0,5
40	0,66
80	0,8
160	0,89
320	0,94
640	0,97
	0,99

Gráfico 1.

El gráfico 1 muestra la curva de confiabilidad de Ebel. En él se ve claramente por qué ningún autor recomienda un examen de opción múltiple de menos de 40 reactivos (confiabilidad 0.667). Por caso, si un educador opta por un examen de 10 preguntas, debe hacerlo a sabiendas de que la confiabilidad de los resultados es de sólo 33%.

Debemos considerar también que, dado que solo un número infinito de reactivos produciría una confiabilidad del 100%, y el aumento indiscriminado del número de éstos influirá en la adecuación de la prueba al tiempo disponible, la elección final del número de reactivos recaerá generalmente en el rango de 40 a 80.

### **1.2 Grado de homogeneidad de los elementos:**

Las pruebas que traten de comprobar conocimientos o información sobre una sola temática específica, tendrán a ser más confiables que las pruebas globales. Si la decisión del educador es la de incluir distintos tipos de tópicos de diversa índole, porque esto conviene a los fines de la evaluación (un examen promocional por ejemplo), deberá tener en cuenta que la confiabilidad de los resultados se verá necesariamente disminuida, y por lo tanto deberá trabajar sobre las otras características que tienden a aumentar el grado de confiabilidad.

### **1.3 Grado de discriminación de los ítems:**

Si bien el grado de discriminación (capacidad de separar a los alumnos entre los bien preparados y los mal preparados para el examen) es un parámetro de fundamental importancia, y será analizado exhaustivamente por separado, debemos mencionar que Lafourcade lo incluye entre los parámetros que mayor influencia tienen en la confiabilidad final del examen.

Una prueba que contenga reactivos con un promedio alto de grado de discriminación, logrará resultados más confiables que otra que incluya reactivos con menor promedio de dicho grado. Si el docente cuenta con una herramienta informática que evalúe el grado de discriminación de cada pregunta almacenada en su set, podrá recurrir a aquellas con mayor grado de discriminación en las pruebas finales, de manera de contrarrestar el efecto de la alta homogeneidad de los reactivos que dicho tipo de exámenes debe incluir.

### **1.4 Grado de dificultad de la prueba:**

Una prueba que posea dificultad media es superior en confiabilidad de resultados tanto a las que tienen mucha dificultad como a las que resultan muy fáciles. Para Lafourcade, un buen examen será aquel que sea superado por algo más de la mitad de los educandos.

Otra vez vemos aquí la importancia de la parametrización de las preguntas del *set*, puesto que podemos independizar, hasta cierto punto, el grado de dificultad de la prueba en sí, del grupo de educandos a los que finalmente se somete a la prueba.

En otras palabras, la sucesiva corrección del grado de dificultad de cada pregunta según el porcentaje de alumnos que la respondieron bien en cada uno de los exámenes en los que se la incluyó, hará que el educador cuente con un *parámetro de dificultad histórico de cada pregunta en sí*, y por lo tanto incluirá en un *test* el conjunto de preguntas que ya ha mostrado tener dificultad media, independientemente del resultado que arrojará el examen conformado en el grupo particular de educandos a los que se les examinará.

## 2. OBJETIVIDAD:

Si a una prueba se le asigna un puntaje que sea independiente del juicio del docente que deba evaluar, se puede afirmar que se está calificando de modo objetivo. La ausencia de esta característica afectará naturalmente a los otros factores de calidad del examen, dado que mal podría inspirar confianza la nota de un maestro afectada por el concepto o juicio previo que el educador tenga del educando.

Por último, incluiremos la fórmula de confiabilidad de Kuder - Richardson denominada KR21, que puede ser aplicada con facilidad, ya que sólo requiere conocer la media y la desviación estándar de los resultados de una prueba. Cabe destacar que, dado que el índice calculado ( $r$ ) es un parámetro de *cada examen* y no del *set* de preguntas ni de una pregunta en particular.

La fórmula propuesta es:

$$r = \left( \frac{k}{k-1} \right) * \left( 1 - \frac{m(k-m)}{ks^2} \right)$$

Donde:

r = coeficiente de confiabilidad k = número de ítems de la prueba m = media s = desviación estándar.
---

Según Lafourcade (1971), para pruebas construidas por un docente (no estandarizadas) se considera suficiente un coeficiente de 0.60 o superior.

## 3. METODO DE CÁLCULO RECURSIVO DEL GRADO DE DIFICULTAD:

Proponemos para el cálculo del grado de dificultad un sistema recursivo que describimos a continuación: En principio, se asignará a una pregunta recién creada y que no ha sido evaluada nunca, el grado de dificultad 0.5. Luego dicho valor se irá corrigiendo sucesivamente tomando el resultado de cada uno de los exámenes. El educador tomará en cuenta que la confiabilidad de dicho parámetro aumentará con el número de alumnos que, a lo largo de las distintas pruebas hayan dado respuesta a dicho reactivo.

Para que rápidamente se adecue a los resultados obtenidos, el parámetro del número de veces que se evaluó la pregunta será inicialmente el número 10. Luego, al ingresar los resultados de los exámenes sucesivos, se calculará el grado de dificultad como el número de veces que el reactivo fue contestado seleccionando la opción correcta, dividido por el número total de alumnos a los que se propuso el reactivo en un examen.

El método que se desarrollará en el presente trabajo, apunta a ser una herramienta que permita al educador el identificar rápidamente las preguntas que siempre son respondidas bien por la mayoría de los alumnos a los que se le propone el reactivo (demasiado fáciles) o aquellas en los que sistemáticamente los educandos seleccionan los distractores (demasiado difíciles) para su corrección.

Es de esperar que con la utilización sistemática del programa, el docente cuente con un conjunto amplio de preguntas sobre cada tema que tenga un grado de dificultad media que ayude a aumentar la confiabilidad de los resultados del *test*.

Se ha generalizado la idea de incluir en una prueba, ítems fáciles al comienzo con el objeto de motivar al alumno, y difíciles al final para desafiar la capacidad de los alumnos más brillantes. Si bien algunos expertos aprueban este punto de vista, la mayoría de las investigaciones presentadas en simposios desde 1932 concluyen lo contrario.

Por último, Lafourcade (1971) menciona una técnica desarrollada por Noll en 1965 para evaluar si el grado de dificultad ha sido adecuado. La misma consiste en detectar si la media de las puntuaciones coincide o se halla muy próxima a la mediana de la muestra y si no existe un número importante de aciertos o desaciertos totales. En tal caso se estará ante una prueba de dificultad adecuada *para ese grupo de educandos*.

Un párrafo aparte merece la corrección por azar. De la investigación bibliográfica realizada, **no se ha podido detectar un sólo autor partidario de descontar de los porcentajes alcanzados en una prueba, los aciertos que pudieran realizarse por azar.**

Los argumentos esgrimidos al respecto son:

- a) La amenaza de disminución de puntaje actúa como un elemento distorsivo y condicionante al momento de seleccionar la opción correcta.
- b) Contribuye a aumentar el grado de stress del educando, que ya se encuentra sometido a una circunstancia ansiógena.
- c) Orienta al alumno hacia el análisis de sí, desde algún punto de vista la opción seleccionada pudiera ser considerada incorrecta y por lo tanto disminuir sus chances de aprobación en vez de orientarlo hacia la identificación de la respuesta.
- d) Es francamente incompatible con los reactivos en los que se requiere seleccionar la opción más adecuada (en lugar de la opción correcta).
- e) No representa ninguna ventaja que no pueda implementarse aumentando el número de respuestas correctas para la aprobación del examen.

Para el caso de elaboración de estadísticas en las que quiera neutralizarse los efectos del azar, es obvio que *deberá descontarse sólo la proporción de posibilidad y no un punto por cada respuesta incorrecta*. Toda vez que la chance está indicada por la simple fórmula "favorable sobre posible", sólo se debe descontar un punto por respuesta incorrecta en los *test* de verdadero / falso en los que la chance es del 50%.

Para un *test* general de cuatro opciones, deberá descontarse con fines estadísticos, un punto por cada cuatro respuestas incorrectas, puesto que la chance será 1/4.

Siempre con fines estadísticos, Lafourcade propone, para la corrección de puntajes, la fórmula de Hors, que es la siguiente:

$$Pc = R - \left( \frac{W}{(n-1)} \right)$$

Donde:

Pc = puntaje corregido  
R = número de respuestas correctas  
W = número de respuestas incorrectas  
n = número de opciones totales.

#### 4. CALCULO MATEMÁTICO DEL GRADO DE DISCRIMINACIÓN:

El grado de discriminación de una pregunta es la capacidad del reactivo de discriminar entre los alumnos que dominan la materia sujeta a examen y los que no la dominan. Resulta entonces en el parámetro más importante de un reactivo.

Un índice sencillo de discriminación que es de rápida obtención y por lo tanto resulta práctico es el *Índice de Pemberton*. El mismo consiste en dividir el grupo de examinados en tres, el 27% de mayor puntaje, el 27% de menor puntaje y un tercer grupo intermedio. Para calcular el índice, se descarta el grupo de resultados intermedios y se resta el número de respuestas correctas de ambos grupos, dividiendo el resultado por el número total de alumnos sometidos al examen.

$$IP = \left( \frac{R - 1}{n} \right)$$

Dónde:

R2 = respuestas correctas del grupo superior  
R1 = respuestas correctas del grupo inferior  
n = número total de examinados

El *índice de Pemberton* tiene su justificación teórica en lo siguiente: El hecho de que un alumno de respuesta correcta o incorrecta a un reactivo, sólo provee información para la determinación del grado de dificultad que el mismo encontró en la pregunta, pero nada nos dice del reactivo en sí. El procedimiento correcto para la determinación del grado de discriminación de una pregunta consistirá en relacionar la respuesta de un reactivo en particular (correcta o incorrecta) con el resultado del examen de cada educando. Dicho en otras palabras, la pregunta discriminará bien si es contestada bien por aquellos alumnos que pasaron satisfactoriamente el examen y mal por aquellos que fracasaron en el examen.

El índice de discriminación de un reactivo deberá disminuir si el mismo fue contestado bien por un alumno del grupo inferior. También deberá disminuir si fue contestado mal por un alumno del grupo superior.

Consecuentemente, el índice aumentará si el reactivo es contestado bien por un alumno del grupo superior, y mal por un alumno del grupo inferior.

Se propone en el siguiente trabajo, **realizar una relación entre el índice de discriminación y el grado de dificultad de cada pregunta.**

La idea básica consiste en aumentar la sensibilidad del indicador para los casos en que, un alumno del grupo superior conteste mal un reactivo o que un alumno del grupo inferior lo conteste bien.

Esto permitirá rápidamente (con la sola evaluación del reactivo en uno o dos exámenes) hacer notar aquellas preguntas que presentan problemas a la hora de discriminar.

La ecuación propuesta será distinta para el caso de una pregunta que fue contestada bien de la ecuación a utilizar cuando la respuesta al reactivo es incorrecta.

En el caso de una respuesta acertada al reactivo se propone la siguiente ecuación:

$$D = \left[ \frac{(99N) + \frac{NX}{100}}{100} \right]$$

Donde:

D = Índice de Discriminación. N = Nota obtenida por el educando. X = Grado de dificultad del reactivo.
--

En el caso de que la respuesta sea incorrecta, la ecuación propuesta es:

$$D = \left[ \left( 1 - \frac{N}{100} \right) * \left( 1 - \frac{N - X}{100} \right) \right] * 100$$

### **Veamos cómo funcionan las fórmulas propuestas:**

Supongamos el **caso (1)**:

Un alumno del 27% superior obtuvo 80% como resultado del examen. Dicho alumno responde acertadamente una pregunta de grado de dificultad 10. Dado que se trata de una pregunta fácil, el índice de discriminación no deberá modificarse mucho de la media, toda vez que ya resultaba probable que el educando contestara bien la pregunta por ser ésta de bajo índice de dificultad y por estar éste bien preparado.

La ecuación deberá castigar el grado de discriminación en el caso de que el alumno del caso (1) conteste mal la pregunta, puesto que es una pregunta fácil y fue mal contestada por un alumno del grupo superior.

En el caso (1) el índice de discriminación arroja un valor de 79.28 si se contestó bien y de 6 si se contestó mal, con lo que la ecuación se comporta correctamente.

**Veamos el caso (2)**

El mismo alumno del grupo superior, que obtuvo 80% en el examen, enfrenta una pregunta difícil, de grado de dificultad 80. La ecuación deberá disminuir el grado de discriminación para el caso de que el alumno la conteste mal (puesto que el reactivo estaría discriminando incorrectamente).

El resultado arrojado es de 79.84 si contesta bien y 20 si contesta mal. Otra vez aquí la ecuación se comporta como es esperable.

**Caso (3):**

Un alumno que fracasa en el examen por obtener 20% de nota, se encuentra incluido en el grupo de 27% inferior calificación. Dicho alumno enfrenta una pregunta fácil de grado de dificultad 10. La ecuación deberá arrojar un índice de discriminación bajo si el alumno contesta la pregunta bien, (19.82) y aumentar su grado de discriminación si el alumno del grupo inferior contestó mal. (72).

Otra vez la ecuación arroja los resultados esperados.

**Por último veamos el caso (4):**

Un alumno del grupo inferior, que obtuvo 20% de aciertos en el examen, enfrenta una pregunta de grado de dificultad alto (80). En el caso de que la conteste bien, el índice de discriminación deberá ser muy bajo puesto que el alumno es del grupo inferior y está contestando bien una pregunta difícil. La ecuación arroja un resultado de 19.96 para esta circunstancia. En el caso de que la conteste mal, el grado de discriminación será de 100, lo cual también es correcto toda vez que se trata de un alumno mal preparado que contesta mal una pregunta difícil.

### **CONCLUSIÓN:**

Los procesos de evaluación, son absolutamente necesarios para comprobar la validez de las estrategias metodológicas utilizadas, y el aprendizaje realizado por los estudiantes.

Las instancias de evaluación suelen derivar en circunstancias de alta ansiedad, que repercuten en los resultados. Los test objetivos, frecuentemente disminuyen el trauma ansiógeno del estudiante en dichos eventos.

No todas las estrategias evaluativas son igualmente adecuadas o eficaces a todos los formatos utilizados en los distintos espacios curriculares.

En particular, los test de opción múltiple son especialmente adecuados como herramienta de evaluación en los formatos de *asignatura – materia* y también de *módulo*.

Resulta entonces fundamental el fijar un mecanismo que nos permita establecer parámetros de calidad para cada una de las preguntas que empleamos en nuestros exámenes de opción múltiple. Los parámetros analizados en el trabajo han sido: CONFIABILIDAD, OBJETIVIDAD, DIFICULTAD Y DISCRIMINACIÓN.

La confiabilidad es un factor complejo, dado que depende del número de ítems incluidos en la prueba, del grado de homogeneidad de los elementos, del grado de discriminación de los ítems, y del grado de dificultad de la prueba.

La objetividad, quedará garantizada cuando el puntaje esté establecido de antemano, independientemente del juicio posterior del docente que realice la corrección. Esto es, que dependa matemáticamente de las opciones elegidas por el estudiante.

Se propone en esta ponencia un método recursivo para realizar el cálculo del grado de dificultad de los ítems de un examen, mediante una fórmula y un procedimiento que se va ajustando a través de los sucesivos resultados obtenidos por cada grupo de alumnos.

Finalmente, se propone realizar una relación entre el índice de discriminación y el grado de dificultad de cada pregunta, a través de una ecuación matemática que, aplicada recursivamente, pone en evidencia muy rápidamente los ítems o preguntas que presentan dificultades y deben ser mejoradas.

## BIBLIOGRAFÍA

Abortes Hugo, Mecanismo para restaurar la jerarquía “natural entre géneros. Exámenes de opción múltiple: una evaluación discriminatoria.

<http://www.jornada.unam.mx/1999/ene99/990105/exámenes.htm>

Carreño H., Instrumentos de Medición del Rendimiento Escolar.

[http://www.universidadabierta.edu.mx/SerEst/Apuntes/AlmendarezMaria\\_instrumentos\\_de\\_medicion.htm](http://www.universidadabierta.edu.mx/SerEst/Apuntes/AlmendarezMaria_instrumentos_de_medicion.htm)

Educarchile, Algunas consideraciones sobre preguntas de opción múltiple.

[http://www.educarchile.cl/modulos/psu/articulos.asp?id\\_noticia=9281&esc=estudiante](http://www.educarchile.cl/modulos/psu/articulos.asp?id_noticia=9281&esc=estudiante)

Elola Nydia, V. Toranzos Lilia, Evaluación educativa: Una aproximación conceptual, Buenos Aires 2000.

<http://www.campus-oei.org/calidad/luis2.pdf>

Lafourcade, P, (1971), Evaluación de los Aprendizajes, Kapelusz, Buenos Aires, Argentina.

Lavié Martínez, La evaluación del aprendizaje a través de Internet

<http://tecnologiaedu.us.es/edutec/paginas/92.html>

Otilia Mentruyt, Acabar con la tragedia de la corrección. Pruebas objetivas de rendimiento. Tests de multiples choice.

<http://www.monografias.com/trabajos11/test/test.shtml>

Porras, L. 2002. Evaluación.

<http://mailweb.udlap.mx/~lporras/ciedd/evaluacion.html>

Sains Leyva Lourdes, Apuntes y reflexiones en torno a los problemas de la evaluación del aprendizaje, La Habana, Cuba.

[http://www.nuestraldea.com/aseycap/apuntes\\_reflex\\_eva.htm](http://www.nuestraldea.com/aseycap/apuntes_reflex_eva.htm)