

DR. A. HINOJOSA
PSIC. EMILIO LUCIO

ANÁLISIS DE LOS MÉTODOS DE EXAMEN

EL EXAMEN, que consiste en la acción de indagar y conocer las circunstancias y cualidades de un objeto, puede realizarse de varias maneras, una de las cuales consiste en el empleo de pruebas o tests, que sirven como instrumento.

En el sentido tradicional, una autoridad realiza el examen y decide aprobar o suspender al candidato según su criterio personal. En el examen de grado el alumno debe sustentar una tesis, replicada por cada uno de los sinodales, que obligan al candidato a defenderla y salir airoso del trance. En algunas facultades aún se emplea este método, que requiere de mucho tiempo, maestros y preparación del alumno para el examen mismo, más que adiestramiento para la profesión o el grado. El examen oral confronta directamente a las autoridades y al examinado, con todas las implicaciones psicológicas inherentes a tal situación. Cualquier jurado puede llevar el interrogatorio al campo que desee, no siempre por motivos racionales. La prueba suele hacerse en público, lo que agrava la situación y pone en juego los problemas de autoridad que pueden surgir de la relación autoridad y candidato. Por maduros que sean los examinadores, puede bastar una varian-

te de su ánimo producida por circunstancias ajenas al examen para que la situación se altere. Pueden influir además factores inconscientes.

La introducción de un temario fijo y la selección del temario por azar, redujo el capricho de los sinodales. Aunque estas innovaciones disminuyen la arbitrariedad, no excluye la mutua influencia que sinodal y examinado se ejercen mutuamente. Influye la edad de ambos, el sexo, la apariencia física, el tono de voz, forma de hablar y todos aquellos factores que pueden hacer que una persona sea atractiva o antipática.

Del examen oral se pasó al examen escrito. Este tiene las ventajas de que no implica el confrontamiento personal del alumno con el maestro sino la resolución de temas sobre el papel, con el tiempo necesario para pensar, a todos los alumnos les son aplicadas las mismas preguntas y la prueba puede ser colectiva, lo que significa ahorro de tiempo. Además, el examinador dispone de material escrito que puede leer, releer y juzgar tan detenidamente como quiera. Para evitar la influencia de los factores personales, la prueba puede firmarse con clave y ser calificada a ciegas. Pero algunas de las ventajas del

método escrito han resultado más ilusorias que reales. El tiempo ahorrado de prueba colectiva se consume en leer cada prueba, escrita con frecuencia en tipos de letra que requieren mucho esfuerzo. Después de leer algunas docenas de pruebas sobre las mismas preguntas, sobreviene un fastidio inevitable y deben revisarse varios cientos de cuartillas, lo que equivale a la lectura de un libro, en este caso repetido y fastidioso. En la calificación de las pruebas sigue influyendo el estado de ánimo del maestro al calificar, el rechazo que le produzca la letra ilegible, la mala ortografía, la redacción confusa y otros factores que aunque son importantes desde el punto de vista de la apreciación global del alumno, no lo son para apreciar su preparación en una materia específica. Cualquier profesor de experiencia sabe que si deja un lote de exámenes reposar algunos meses, al reelerlos después, suele rectificar calificaciones y no puede confiar plenamente en su capacidad de juicio apreciativo inmediato y no está a salvo de cometer alguna injusticia.

Otro problema del examen escrito es la valoración de la dificultad de los temas. Como éstos deben variarse, al menos cada año, aunque el profesor sea cuidadoso a veces pondrá exámenes difíciles y otras veces el examen resultará más fácil. Algunos profesores pasivos resuelven este problema preguntando siempre lo mismo, y los alumnos que heredan la experiencia de sus antepasados ya saben lo que tal maestro acostumbra preguntar. Otro problema consiste en que si se preguntase todo el contenido de un curso, la resolución de la prueba sería la repetición completa del texto y si se hacen unas cuantas preguntas se introduce en forma muy importante el elemento azar, puesto que los temas que forman una prueba de dimensiones aceptables suelen no representar más que el cinco o diez por ciento del contenido del curso.

El crecimiento explosivo de la población escolar ha mantenido la necesidad de pruebas que permitan evaluar a los alumnos y realizar de la mejor manera posible la tarea de seleccionar a los realmente capaces. El examen fraccionario o por escalones es un avance. El estudiante, mes por mes, puede ir comprobando de alguna manera que va progresando en su aprendizaje lo suficiente para proseguir. Un buen examen confronta al alumno consigo mismo, con su sentimiento de responsabilidad y su autoevaluación. El aspecto autoritario del antiguo maestro desaparece y quedan a solas el estudiante y su autocritica. Las pruebas adquieren otra dimensión y se convierten tanto en elementos didácticos como en expe-

riencias formativas que pueden favorecer la maduración. La observación de las deficiencias, de los errores cometidos y de sus causas, pueden contribuir al desarrollo de la personalidad del estudiante y dejar de lado el miedo, resentimiento y dependencia del profesor, para transformarse en una experiencia personal de la que el maestro es un simple vector. La prueba se convierte en un instrumento que puede ser manejado por el propio estudiante y para su experiencia.

Las condiciones que una buena prueba debe reunir son:

- 1.—Cubrir toda el área de conocimientos por examinar.
- 2.—Permitir una valoración objetiva y sin prejuicios.
- 3.—Ser aplicada colectivamente y calificada con rapidez.
- 4.—Explorar realmente las áreas de conocimiento que se desea.
- 5.—Excluir los elementos perturbadores en la relación de autoridad.
- 6.—Ofrecer un grado de dificultad uniforme.
- 7.—Eliminar el elemento azar.
- 8.—Eliminar los aciertos sin información, por inducción o inferencia.
- 9.—Eliminar los aciertos por sabiduría popular y sin información científica.
- 10.—Explorar no solamente la información memorizada o automática sino también la asimilación de los conocimientos y su integración.
- 11.—Explorar el manejo mental de la información.
- 12.—Explorar el conocimiento del lenguaje técnico como parte de la información.
- 13.—Producir interés y atractivo, en vez de miedo y fastidio.

Los clásicos exámenes orales tienen casi todos los inconvenientes contrarios a las condiciones enumeradas y aunque el examen escrito por temas sigue usándose abundantemente en nuestra Universidad, resulta anticuado y cumple pocos de los postulados propuestos. Desde luego, una buena técnica requiere conocimientos técnicos y pedagógicos, iniciativa, y renuncia al individualismo de tipo irracional.

Revisaremos someramente los principales tipos de pruebas ideadas que fueron tomadas en cuenta para hacer nuestra selección. Se les suele llamar reactivos para pruebas objetivas.

I.—Reactivos de verdadero y falso. Consiste en presentar dos afirmaciones, una de las cuales es falsa, para que el alumno señale la verdadera. Son los reactivos de elaboración más fácil, pero tienen grandes desventajas. El azar se introduce en fuerte porcentaje, induce al alumno a aprender de memoria y al profesor a presentar estímulos estereotipados, produciendo efectos psicológicos desagradables en los alumnos, que la terminan cansados por su inevitable longitud y con la sensación de ser apreciados en menos de los que saben y pueden.

II.—Reactivos por completar.—Consisten en preguntas que se complementan con una respuesta corta, que puede ser hasta de una sola palabra o en afirmaciones con espacios intermedios o finales en blanco. Este tipo de prueba no es tan objetivo y preciso como otros, porque puede haber varias formas de respuesta correcta, con variantes difíciles de controlar y las afirmaciones propuestas pueden tener un carácter ambiguo, que propicia tipos diferentes de respuesta e introduce de nuevo el criterio personal al calificar cada caso en particular.

III.—Reactivos de jerarquización.—Este método requiere el análisis de las preguntas y el empleo del criterio y del juicio. Consiste en presentar al alumno una serie de afirmaciones que contienen, en diverso grado, la respuesta que se considera correcta. El alumno deberá señalar la mejor respuesta, aunque las demás no sean totalmente erróneas. O podrá agruparlas señalando su orden jerárquico, ya sea lógico, cronológico o de otro tipo. A pesar de lo sugestivo de este método, su preparación ofrece grandes dificultades y la valoración de las respuestas resulta difícil, dado que todas contienen elementos verdaderos. El empleo de estas pruebas ha dado lugar a inconformidad y discusiones por parte de los alumnos, que se consideran tratados con injusticia al tener la convicción de que eligen con acierto y sin embargo la calificación obtenida resulta sorprendentemente baja.

IV.—Reactivos de aparejamiento.—Consiste en colocar por pequeños grupos, series de estímulos en dos columnas, derecha e izquierda y pedir al alumno que marque con una línea o un número cuáles están relacionados entre sí. Esta técnica es sólo una variante de la prueba de selección múltiple y puede resultar muy conveniente, sobre todo si se combina con ésta, siempre y cuando se toman en cuenta tres

condiciones: Que el número de reactivos de la derecha sea mayor que el de la izquierda, para dejar abierta la elección de un mayor número de posibilidades, que una alternativa de la izquierda se corresponda con varias de la derecha, o viceversa. Y también, que las listas de reactivos no sean demasiado extensas para que el alumno no emplee mucho tiempo buscando entre largas listas de estímulos.

V.—Reactivos de selección múltiple.—Consiste en presentar una lista de afirmaciones, habitualmente cinco, de las cuales una debe señalarse como correcta. A las formulaciones incorrectas se les llama distractores y su elaboración y características son muy importantes para la confiabilidad de la prueba. Ya que esta prueba parece ser la que se acerca más al ideal postulado para el examen, nos detendremos un poco en su descripción y luego ofreceremos los resultados y observaciones que hemos hecho durante su aplicación. Eligiendo la respuesta al azar, tendríamos un 20% de probabilidades de acierto y la corrección de este inconveniente se traduce en un aumento exagerado de los reactivos. Para obviar este defecto y hacerla eficaz, debe cumplirse con una serie de requisitos, tanto en la elaboración de la afirmación correcta como de los distractores. Los textos que tratan de la elaboración de pruebas no dan suficiente explicación técnica ni analizan algunos problemas de la prueba ni su manera de resolverlos. Por tal motivo nos decidimos a realizar este estudio, que aborda los problemas de este tipo de pruebas desde un ángulo distinto. Los reactivos de selección múltiple tienen la ventaja de permitir la presentación de problemas de alto grado de complejidad, ya que no quedan limitados a la presentación de hechos específicos.

Ejemplo para una prueba de física elemental:

I.—El término usado en conexión con el sonido, que corresponde al color en relación a la luz, es:

- 1.—Diapasón.
- 2.—Timbre.
- 3.—Tono.
- 4.—Intensidad.
- 5.—Armonía.

Si no se poseen conceptos claros y adecuados de los elementos de física no puede descubrirse la respuesta correcta, que fue presentada con mucha bre-

vedad, a pesar de representar un problema complejo.

O en este reactivo para la prueba de Introducción a la Medicina Humanística:

I.—El médico actual supera cada vez más:

1.—La necesidad de explicación metafísica de las enfermedades.

2.—La necesidad de recurrir a ciencias ajenas a la medicina.

3.—La necesidad de utilizar la lógica para el diagnóstico.

4.—La necesidad de expresar sentimientos y afecto hacia sus pacientes.

5.—La necesidad de compadecerse de los enfermos incurables.

A pesar de tratarse de un problema que abarca aspectos históricos de la medicina y de la personalidad del médico, la persona que sabe lo necesario acerca del problema, descubre en seguida cuál es la respuesta válida. El ignorante, en cambio, se pierde en los distractores, porque no sabe reconocer la verdadera. Se trata de una prueba de reconocimiento en que cada reactivo contiene la norma que señala la respuesta correcta. Este tipo de reactivos podría compararse a las pruebas que en óptica se usan para el reconocimiento de los colores: Se presentan círculos llenos de puntos de colores diversos, pero de igual intensidad lumínica. El sujeto normal descubre en seguida un número o una letra porque están formados con puntos de color diferente, que pasa desapercibido para el daltónico.

Sin llegar a la jerarquización, algunos de los distractores que deben ser desechados pueden contener parte de la verdad, pero ser tan vagos o generales, que la respuesta correcta se destaque como la de elección. Todas estas cualidades dan al reactivo de selección múltiple una flexibilidad que le convierte en una de las mejores pruebas modernas. Además, resulta interesante y rápida para el alumno preparado. El alumno impreparado podrá emplear mucho más tiempo tratando de deducir o inferir la respuesta adecuada, puesto que no la reconoce, pero por esto mismo, no se sentirá tratado con injusticia y la prueba servirá como experiencia didáctica que le lleve a la convicción de su ignorancia. Al revisar la prueba con los alumnos, pronto aceptaron la evidencia y mostraron cierta repugnancia por seguir viendo el resto de sus errores, cosa completa-

mente distinta a lo que acontece con las pruebas de tema libre o de jerarquización, que pueden dar lugar a discusión con grandes argumentaciones, racionalizaciones e inconformidad.

Si se tiene a la vista el programa del curso, se incluyen reactivos que cubran todo el campo de la enseñanza en una misma prueba y además se pueden poner varios reactivos para un mismo conocimiento, reducir su número para los temas menos importantes y establecer un equilibrio adecuado. Regulando estos porcentajes se puede mantener homogeneidad a través de los años, variándolos según la experiencia. También puede ayudar a la apreciación de la unidad de enseñanza de un departamento y a regularizar la actividad de los profesores.

Para este trabajo, tomamos un grupo del curso de Introducción a la Medicina Humanística y otro del curso de Psicología Médica. En ambos grupos realizamos exámenes parciales y final por temas. En las mismas ocasiones, aunque en diferente día aplicamos los reactivos de selección múltiple. Otro profesor que imparte las mismas materias aplicó en su grupo y en las mismas condiciones, las mismas pruebas de selección múltiple y proporcionó los resultados de las calificaciones que obtuvo con el examen por temas, con objeto de establecer las correlaciones. Además, realizamos el estudio de cada pregunta y de los reactivos de la siguiente manera.

1.—Calificar el grado de dificultad de cada pregunta, lo cual se logra observando simplemente el porcentaje de calificaciones acertadas.

2.—Medir la capacidad discriminativa de cada pregunta. Para esto se toma el grupo de alumnos de calificación más alta en el total de la prueba y el grupo con calificación más baja y se compara el porcentaje de errores y aciertos de cada grupo en relación al reactivo. De este modo puede observarse que hay reactivos que marcan notables diferencias entre los alumnos preparados y los impreparados y otras que no establecen discriminación satisfactoria.

Con el objeto de establecer comparaciones, colocamos tanto el grado de dificultad como la capacidad discriminativa de cada pregunta en escala de uno a diez, de acuerdo con la altura de los porcentajes.

ANÁLISIS DE LOS MÉTODOS DE EXAMEN

Escala de Grado de Dificultad	Escala de Discriminación
1 a 10 = X	1 a 5 = I
11 a 20 = IX	6 a 10 = II
21 a 30 = VIII	11 a 15 = III
31 a 40 = VII	16 a 20 = IV
41 a 50 = VI	21 a 25 = V
51 a 60 = V	26 a 30 = VI
61 a 70 = IV	31 a 35 = VII
71 a 80 = III	36 a 40 = VIII
81 a 90 = II	41 a 45 = IX
91 a 100 = I	46 a o más = X

La escala de grado de dificultad se elaboró basándose en el porcentaje de alumnos que contestaban cada pregunta y la escala de discriminación se elaboró de acuerdo con las diferencias de porcentajes de alumnos de alto y bajo puntaje que contestaban cada pregunta. Si ambos grupos contestaban acertadamente a una pregunta en igual proporción, por ejemplo 70%, la discriminación es 0, o sea que la pregunta no permite distinguir entre los buenos y los malos alumnos. Si la diferencia es más de .46 (la máxima hallada fue de 36 a 40%), la discriminación alcanza el grado X.

3.—Una vez señalados los puntos 1 y 2, se hizo lo mismo con otro grupo del mismo año, y se correlacionaron los resultados.

4.—Obtenidos los datos anteriores se tomó aisladamente cada reactivo y se analizaron las causas de su éxito o fracaso, para normar su elaboración en las siguientes pruebas.

Dividimos las observaciones en dos grupos: factores que ayudan a la efectividad del reactivo y factores que merman o anulan su efectividad.

Debe recordarse que para que una prueba sea considerada como eficaz, conviene que no pueda ser resuelta en su totalidad por los alumnos, sino en un setenta por ciento, aproximadamente. De este modo, la dificultad de la prueba será tal, que tenga suficiente amplitud para medir a los alumnos más brillantes y a los menos eficientes. Una prueba medianamente difícil, discrimina a los alumnos muy inferiores de los medianos y buenos. Solamente una prueba suficientemente difícil discrimina a los alumnos buenos de los brillantes y distingue también a los grupos inferiores. El "diez", quedará colocado en las últimas porciones de la curva, de acuerdo con la media. La experiencia permite ir modulando la

apreciación de los grupos, ya que no todos son iguales.

Factores positivos:

1.—Se encontró conveniente usar en los distractores palabras y formulaciones técnicas que no difieran de los de la respuesta válida y que se acerquen a las formulaciones técnicas esperadas.

Ejemplo:

LOS TROPISMOS SE PRESENTAN EN:

- a) Las especies prehomínidas y los simios superiores.
- b) La conducta instintiva estereotipada.
- c) Las plantas y los animales inferiores.
- d) El recién nacido, antes de la mielinización.
- e) Determinada clase de enfermos mentales.

LOS RESULTADOS OBTENIDOS FUERON LOS SIGUIENTES:

	Mitad de alumnos con bajo puntaje	Mitad de alumnos con alto puntaje	TOTAL
N.C.	2%	2%	2%
a)	10%	2%	6%
b)	34%	14%	24%
c)	54%	82%	68%
d)	0%	0%	0%
e)	0%	0%	0%

Un 68% del grupo eligió la respuesta acertada, lo que da un grado de dificultad conveniente. El número de alumnos con bajo puntaje general que escogió la respuesta acertada, fue 54%; en cambio entre los de alto puntaje el porcentaje de aciertos alcanzó al 82%. Estos datos significan que la pregunta fue contestada con acierto en cifras diferentes entre los buenos y los alumnos deficientes y que por lo tanto establece adecuada discriminación entre ambos grupos. El grado de dificultad fue calificado como IV y el de poder discriminativo como VI. (En escala de I a X.) Solamente dos distractores funcionaron bien por los motivos señalados y otros.

2.—Funcionan mejor los distractores que requieren buen análisis para su comprensión.

Hay distractores de formulación muy simple y definida, expresados de manera breve y de modo que su comprensión no requiere reflexión. Otros, en cambio, contienen implicaciones que deben observarse con cuidado antes de rechazar o aceptar las afirmaciones que contienen. En cierto tipo de materias, en que los conocimientos pueden ser memorizados y son de tipo concreto, probablemente este tipo de reactivos distractores no sería el más adecuado. Pero en materias como la psicología o la fisiología, que requieren de conceptos altamente elaborados e interrelacionados entre sí, pueden resultar muy convenientes reactivos que pongan en juego la capacidad crítica y analítica del alumno. Así en el espacio tan reducido que supone un reactivo de cinco afirmaciones, se explora un área muy considerable de su pensamiento. Es frecuente que las personas tomen los clisés sin observarlos cuidadosamente y acepten su contenido sin conocer sus implicaciones. En este punto, los distractores del tipo que estamos señalando tocan un rasgo peculiar del carácter: la pasividad receptiva. Desde el punto de vista teórico, es indeseable que los rasgos de carácter intervengan en una prueba, pero es inevitable, dado que una persona reacciona con lo que es y en su rendimiento intervienen las peculiaridades de su carácter. Esto es un asunto importante, pero que debe estudiarse en otra parte. Es muy probable que el estudiante que reacciona con determinada peculiaridad en una prueba, repetirá este tipo de conducta en otras situaciones de su vida.

3.—*Funcionan adecuadamente los distractores que contienen afirmaciones que van en la dirección de los prejuicios comunes*, ya que el alumno ignorante tiende a suplir su ignorancia con la sabiduría popular o convencional. El alumno irresponsable y oportunista cree fácilmente haber hallado una solución gratuita, que acepta cuando no está preparado.

Ejemplo de distractor:

LA EMBRIOLOGÍA HUMANA ENSEÑA

b) que el desarrollo convierte al hombre en un ser de naturaleza distinta a la del animal.

El distractor b, fue tan poderoso que absorbió casi todos los errores, por dos motivos. El primero es su relación con los prejuicios y la sabiduría popular, que acepta diferencias de naturaleza entre el hombre y el animal, lo que puede ser cierto sólo desde el punto de vista teológico. Para un biólogo

o un antropólogo las diferencias son de grado y dirección, pero no considera que se trata de un ser específicamente distinto en cuanto a su naturaleza.

4.—*Se encontró que funcionan mejor los distractores que van en la dirección de las imágenes convencionales del médico.*

En la materia que examinamos se tratan estos temas, pero probablemente sucede lo equivalente en otros campos científicos. Ejemplo:

EL MÉDICO ACTUAL TIENDE A CONVERTIRSE EN:

- a) el hombre más popular y deseado por la sociedad.
- b) un elemento político de gran importancia.
- c) el mejor representante de la conciencia social.
- d) un obrero automatizado de la ciencia.
- e) el sustituto moderno del sacerdote.

Resultados:

	Mitad de alumnos con bajo puntaje	Mitad de alumnos con alto puntaje	TOTAL
N.C.	2%	2%	2%
a)	2%	0%	1%
b)	9%	5%	7%
c)	24%	12%	18%
d)	54%	62%	48%
e)	9%	19%	14%

Un total de 58% de alumnos resolvieron adecuadamente la prueba lo que dio un grado de dificultad de V (o sea suficientemente amplio) y en cambio, la discriminación alcanzó un grado II, o sea inferior al de dificultad, por lo que se presta para analizar varios factores interesantes.

El distractor c, va en relación de una imagen convencional del médico, que lleva a los alumnos a considerarlo como el mejor representante de la conciencia social alcanzó un 22% entre los alumnos de bajo puntaje y hasta un 12% de los de alto puntaje. El segundo lugar lo ocupó el distractor e, siendo de observar que cayeron en el error más del doble de los alumnos de alto puntaje. Se trata de una imagen idealizada del médico que aunque defectuosa, está mejor orientada que la del reactivo c, y se prestó a equívoco. El distractor a, funcionó escasamente por ser demasiado obvio, sin embargo, el 2% de los alumnos de bajo puntaje lo seleccionó.

5.—Es conveniente elaborar distractores precisos y que se dirijan específicamente hacia lo preguntado, porque el estudiante mal preparado puede tener algunas nociones vagas, pero suficientes para descartar el reactivo equivocado.

Ejemplo:

31.—LOS TRABAJOS DE PASTEUR FUERON IMPORTANTES PORQUE:

e) permitieron curar la rabia en el hombre.

Este distractor fue elegido por el 45% de los alumnos de bajo puntaje y sólo por el 34% de los de alto puntaje. En total fue elegido por el 40% del grupo. En la expresión "curar" está la dificultad, ya que no se ha logrado curar a un enfermo de rabia. Este distractor muestra cómo el empleo de términos definidos y directos aumenta su eficacia, acentuada por su relación con una idea popular.

Factores negativos en la elaboración de los distractores:

1.—Se ha observado que el empleo de términos del lenguaje común y de conceptos que aparentan pertenecer a lo vulgar no funcionan adecuadamente como distractores.

2.—No funcionan adecuadamente los distractores demasiado desconectados del tema, porque su desconexión sirve de orientación al estudiante para desecharlos.

Ejemplo:

14.—EL COCIENTE INTELECTUAL CONSISTE EN:

- a) relacionar el desarrollo intelectual con la edad.
- b) multiplicar la edad cronológica por la edad mental.
- c) calcular el deterioro mental de un sujeto.

Este reactivo sirve de ejemplo para ilustrar cómo formulaciones demasiado alejadas del tema (distractor c) obtienen poco resultado.

3.—Es una desventaja formular los distractores de manera vaga porque esto desanima a elegirlos. Las proposiciones científicas suelen tener un carácter preciso y determinado que las hace destacarse fácilmente de distractores mal formulados.

Ejemplo:

3.—LA INTUICIÓN ES:

- a) el presentimiento de que algo nefasto va a ocurrir.
- b) la sospecha de que una idea es errónea.
- c) Alcanzar un conocimiento inmediato sin razonamiento previo.
- d) integrar un pensamiento con una emoción.
- e) la interpretación adecuada de los recuerdos.

Los dos primeros distractores, a y b, resultaron demasiado vagos y contrastantes con la afirmación precisa y técnica que les sigue. Los dos reactivos finales, aunque tuvieron escaso funcionamiento, lograron algún éxito gracias a su formulación precisa, que recuerda las afirmaciones científicas.

4.—Hay que evitar distractores que contengan una verdad parcial en relación con lo preguntado.

Ejemplo:

42.—UNA CARACTERÍSTICA DEL CEREBRO HUMANO ES:

- b) el desarrollo predominante de los lóbulos frontales.
- c) el predominio funcional de uno de los hemisferios sobre el otro.

Resultados:

	Grupo de alumnos con bajo puntaje	Grupo de alumnos con alto puntaje	TOTAL
a)	16%	3%	10%
b)	24%	30%	27%
c)	17%	38%	27%
d)	30%	22%	26%
e)	11%	2%	6%
NC	2%	3%	3%

Las afirmaciones contenidas pueden ser discutibles y llevan a confusión a los alumnos, aunque tengan una buena información.

5.—Se encontró inconveniente el uso de distractores que tengan elementos que les den alguna apariencia de absurdos, porque esto permite eliminarlos fácilmente, aun con poca información.

Ejemplo:

La imaginación permite: a) "Que un sujeto lleve a enfermar de esquizofrenia." A veces se recibe la sorpresa de encontrar que un distractor que parecía absurdo es aceptado por un alto porcentaje de alumnos. Los motivos se analizaron ya: el ir en la dirección de un prejuicio, halagar la vanidad, etc. y así resultó en el ejemplo citado.

6.—Finalmente, los distractores deben requerir también determinado grado de análisis, los que no lo requieren resultan fáciles de eliminar.

Ejemplo:

La afirmación establece: *Al médico actual le conviene* y el reactivo distractor a completa: "Leer exclusivamente literatura médica". Resultó obvia la incorrección del distractor y solamente unos cuantos alumnos muy prejuiciados o de carácter mercantilista aceptaron este distractor.

Características que aumentan la efectividad de un reactivo sólido:

1.—El uso de palabras técnicas y científicas. Una de las funciones de las pruebas es explorar, además de la información, el uso y la comprensión del lenguaje técnico de la materia que se examina. El lenguaje técnico significa una barrera para el estudiante listo pero mal preparado y no lo es para un estudiante bien informado. Para que funcione bien es necesario que también se encuentren términos técnicos semejantes en los distractores.

Ejemplo:

1.—LA SENSACIÓN ES:

- a) la percepción de la configuración de las cosas.
- b) la recepción a nivel cerebral de un estímulo, a través del órgano sensorial específico.
- c) el resultado de la integración de los órganos de los sentidos entre sí.
- d) el pensamiento de que algo inminente ocurrirá.
- e) discriminar por el tacto la forma de un objeto.

	Grupo de alumnos con bajo puntaje	Grupo de alumnos con alto puntaje	TOTAL
a)	18%	10%	14%
b)	43%	52%	47%
c)	39%	32%	36%
d)	0%	0%	0%
e)	0%	5%	3%

El reactivo c alcanzó un 36% de éxito a pesar de ser erróneo, debido a su formulación en términos técnicos, que impresionaron al estudiante mal informado y los distractores d y c que no se expresan en formulación técnica fracasaron.

2.—Es preferible hacer afirmaciones que requieran una información exacta del punto examinado. De este modo, los alumnos mal preparados son fácilmente eliminados, al aumentar la necesidad de discriminación.

Ejemplo:

32.—LOS TRABAJOS DE KOCH TIENEN IMPORTANCIA PORQUE:

b) indican la etiología específica de la tuberculosis.

d) descubrieron las reacciones para diagnosticarla.

El distractor d que tiene una formulación muy precisa alcanzó un alto puntaje. El reactivo verdadero (b) también obtuvo un alto puntaje, pero su formulación precisa eliminó a los estudiantes de conocimientos vagos.

3.—También se encontró que resultan más discriminativos y precisos los reactivos que requieren el uso de pensamiento ordenado y científico, de fuentes de conocimiento adecuadas y materias técnicas auxiliares.

Ejemplo:

Los huesos de la cadera permiten a) suponer la posición en que caminaba el ser a que pertenecieron. Si un estudiante no tiene buena información de la mecánica de la cadera además de conocimientos antropológicos, tendrá mayor dificultad para resolver este reactivo. Si se acompaña de distractores formulados en condiciones semejantes conserva la homogeneidad y aumenta en la capacidad discriminativa del reactivo.

Aspectos negativos que deben evitarse en las formulaciones válidas.

1.—Deben evitarse las palabras comunes y poco técnicas. Aunque su empleo pudiera ser útil para "despistar", en realidad no resulta conveniente usar tales recursos, sino elevar el nivel técnico de las formulaciones en lo general.

Ejemplo:

3.—LA ANATOMÍA COMPARADA ENSEÑA:

- a) que el cerebro humano es específicamente distinto al del animal.
- b) que el aparato digestivo de un rumiante es menos complicado que el humano.
- c) que el ojo humano se parece más al de una gacela que al de un lemúrido.
- d) que la boca con labios gruesos del negro es un rasgo regresivo.
- e) que el hombre actual tiene restos de cola.

El distractor *d*, aunque falso, tiene una formulación más técnica que el *e*, cuya formulación recuerda lo popular y produce desconcierto en un estudiante poco seguro de sus conocimientos. La consecuencia fue que estudiantes preparados la descartaron y eligieron otro reactivo de apariencia más técnica. El grado de dificultad del reactivo se aumentó sin mejorar su capacidad discriminativa.

2.—Debe evitarse la referencia a conocimientos que puedan adquirirse o inferirse de información popular, que aunque pueden ser válidos no permiten discriminar entre los alumnos bien y los mal preparados, además de que los conocimientos de sabiduría popular no tienen la calidad de los académicos.

3.—Es un inconveniente formular el reactivo válido de manera amplia y complicada, pues en su cuidada elaboración el alumno puede advertir que se trata de una formulación mejor preparada que los distractores y llamar su atención sobre ella, aun sin reconocer su validez por carencia de información. En una serie de distractores presentados en forma técnica pero sencilla, se presentó la afirmación válida de la siguiente forma:

7.—EL PENSAMIENTO CIENTÍFICO PROGRESA POR:

- a) la necesidad que el hombre tiene de explicarse y resolver de la manera más adecuada posible los problemas y las necesidades de su existencia.

El reactivo fue resuelto adecuadamente por el 86% de los alumnos (grado II), resultando más fácil de lo deseable. Su poder discriminativo alcanzó solamente el grado I, sirviendo escasamente para distinguir entre los alumnos de alto y bajo puntaje. La cuidada y extensa elaboración de la proposición válida llamó la atención y condujo directamente a su análisis y selección, tanto a los alumnos preparados como a los ignorantes.

4.—Una afirmación válida resulta perjudicada cuando no se formula de manera clara o se dificulta inútilmente su comprensión.

33.—LISTER SE OCUPÓ PRINCIPALMENTE DE

- a) los fenómenos de inmunidad en las intervenciones quirúrgicas.
- b) destruir los gérmenes que infectaban las heridas quirúrgicas.
- c) difundir la aplicación de los anestésicos.
- d) desarrollar la técnica de la asepsia.
- e) desarrollar la técnica de ligaduras para evitar la hemorragia.

En el reactivo válido se quiso evitar el término antisepsia para no facilitar demasiado el reconocimiento y exigir su análisis, lo que no era indispensable pues los términos técnicos de asepsia y antisepsia exigen el análisis de por sí. El resultado fue que el reactivo se dificultó hasta dar sólo un 55% de aciertos y alcanzó un grado VII de dificultad. La capacidad discriminativa resultó de grado VIII o sea muy buena, pero este resultado pudo haberse alcanzado sin introducir demasiada dificultad para los alumnos preparados.

5.—Deben evitarse los reactivos cuya respuesta válida puede estar sujeta a interpretaciones que varían con el diverso criterio científico, y se prestan a confusión al elegir reactivos de validez dudosa.

Reacción de los alumnos ante este tipo de pruebas:

Resulta útil aplicar una prueba corta de adiestramiento para que el alumno aprenda a resolverla y adquiera confianza. Es favorable explicar que en este tipo de pruebas se trata de eliminar tanto el azar como la irregularidad en la calificación por parte del profesor, factores no siempre advertidos y que causan temor. También es útil explicarles que casi nadie resuelve el 100% de la prueba y que este grado de dificultad tiene por objeto proporcionar una escala de medida suficientemente amplia y además que el grado de dificultad de la prueba es medido por los alumnos mismos, que con su capacidad para resolverla indican en dónde deberían colocarse las calificaciones mejores, medianas y bajas. Esta última explicación tuvo además la ventaja de disminuir la tendencia a ayudar clandestinamente a los alumnos ignorantes, haciéndoles notar que si el grado de dificultad de la prueba se determina por el rendimiento global del grupo, cada alumno mejorado por ayuda clandestina rompe el equilibrio

de la curva y no se establece una discriminación adecuada entre los alumnos, perjudicando a los que ayudan.

Después de la primera aplicación los alumnos recibieron con agrado las calificaciones, porque en lo general, tenían la impresión de haber obtenido notas más bajas. Se les permitió revisar sus pruebas y conocer la manera de calificarlas, observando que los alumnos inconformes por su baja calificación quedaron convencidos al revisar y examinar sus respuestas, evitándose las discusiones. Esto quizá parezca extraño en otros ambientes, pero en nuestra escuela es frecuente que el alumno inconforme, de manera informal, discuta con el profesor, le pida que revise su prueba y la defienda con alegatos y discusiones muy incómodas. Después de la primera aplicación, los alumnos adquirieron confianza excesiva y en la siguiente, las puntuaciones bajaron. Para el examen final preguntaron con ansiedad si se les pondría examen por temas y pidieron unánimemente la prueba de selección múltiple. Se procuró que ésta fuera algo más difícil que las anteriores y se advirtió a los alumnos su exceso de confianza. Se observó también, que algunos alumnos que habían obtenido notas brillantes en otras materias resultaron bajos en nuestras pruebas. Intrigados por estos resultados, ya que los alumnos afirmaban haberse preparado bien, se les aplicaron otras pruebas y los resultados bajos se repitieron a pesar de una preparación responsable. Esto fue particularmente claro en un alumno, que estudiado directamente demostró ser aplicado, pero poco inteligente, con poca penetración y mala capacidad analítica, simple repetidor y memorista. Otro de los alumnos hubiera obtenido la nota más alta en examen por temas, pues en una prueba que se le hizo, anotó algunos párrafos al pie de la letra, al grado de hacer pensar que los había copiado, y en cambio sólo alcanzó resultados medianos en las pruebas de selección múltiple. No se encontró el caso de alumnos que por ansiedad o inhibición hubieran fallado, ya que es más favorable ponerse a meditar sobre material dado que recordar componer y redactar el tema que se pide. Además, la prueba fue tomada con interés y en cierto modo como si fuera un pasatiempo o rompecabezas, con interés en sí mismo. Las notas para copiar o "acordeones" resultan prácticamente inútiles y solamente hay que vigilar que los alumnos no se comuniquen entre sí. Es posible imprimir dos pruebas diferentes variando la colocación de los mismos reactivos, de modo que se dis-

tribuyan en distinta numeración y sea más difícil comunicarse la respuesta adecuada.

Para cerciorarnos de las correspondencias en la calificación de las pruebas por temas y las de reactivos de selección múltiple, se aplicaron ambos tipos de pruebas a dos grupos, uno a alumnos de nuestra clase y otro de otro profesor ajeno a este estudio y que proporcionó la oportunidad de su clase. Se aplicó primero el examen por temas, explicando a los alumnos que la calificación obtenida en este examen sería la tomada cuenta y al día siguiente el de selección múltiple. En el momento de distribuir las pruebas se les informó que si algunos de ellos alcanzaban mejor puntuación en esta segunda prueba que en la primera, ésta sería su calificación definitiva, no así si resultaba inferior a la del examen por temas. Esto se hizo con el objeto de estimularles a la resolución de la prueba y obtener un esfuerzo mayor.

Se amplió una prueba empleada de cincuenta reactivos de selección múltiple, compuesto cada uno de una afirmación válida y cuatro distractores, 250 proposiciones en total. El estudiante bien preparado la resolvió en un tiempo variable entre media y tres cuartos de hora, porque fue capaz de reconocer la respuesta adecuada casi de inmediato y sólo tuvo que leer rápidamente las demás para asegurar su elección. Los alumnos mal preparados tardaron hasta el doble, ya que tienen que pensar y repensar, dando vueltas a todas las posibilidades de acertar. Transcurrido cierto tiempo, suelen aceptar como realidad su ignorancia y acaban señalando la que les parece más probable. Casi nunca renuncian a la posibilidad de acertar, aunque sea por azar y son pocos los que se abstienen de responder algo que no saben y dejan en blanco el reactivo. Son el pequeño número de N.C. (no contestó) que se puede ver en algunos de los ejemplos citados.

El grado de dificultad de la prueba dio un promedio de 3.4, que corresponde a un grado III de dificultad e indica que solamente un promedio del 21% al 30% de los alumnos fracasó. Esto hizo la prueba un poco más fácil de lo que recomiendan los técnicos en la elaboración de test, tratándose de estudiantes universitarios.

Analizando el contenido de los reactivos, encontramos que de acuerdo con su implicación podían quedar divididos en seis grupos de la manera siguiente:

I.—Atributo de un objeto ("La especie humana

se caracteriza por") 18 preguntas. Índice de dificultad 3.5. Grado de Dif. III.

II.—Condición ("La vejez se inicia cuando") 13 preguntas. Índice de Dif. 3.2. Grado de Dif. III

III.—Causa efecto ("La cultura produce") 13 preguntas. Índice de Dif. 3.9. Grado de Dif. IV.

IV.—Acción ("Las religiones se producen") 2 preguntas, índice de dificultad 2 Grado de Dif. II

V.—Definición ("El instinto es") 1 pregunta, índice de dificultad 7 Grado de Dif. VII

VI.—Afirmación ("La salida de los primeros dientes coincide con") 1 pregunta, índice de Dif. 2 Grado de Dif. II.

En el grupo principal el puntaje promedio fue 75. Desviación estándar 20.1.

En el grupo testigo el puntaje promedio fue de 54 Desviación estándar 24

De acuerdo con estos puntajes promedio y las desviaciones, se estableció la tabla de calificaciones en cada uno de los grupos, obteniéndose los siguientes resultados:

<i>En el grupo principal:</i>		
<i>% de alumnos</i>	<i>No. de alumnos</i>	<i>Calificación</i>
3.7	4	3
3.7	4	4
10.1	11	5
27.7	30	6
14.8	16	7
20.3	22	8
9.2	10	9
10.1	11	10
TOTAL: 99.6		108

<i>En el grupo testigo:</i>		
<i>% de alumnos</i>	<i>No. de alumnos</i>	<i>Calificación</i>
10	6	3
12	7	4
19	11	5
33	19	6
12	7	8
13	8	9
TOTAL: 99		58

Como ya se dijo, a ambos grupos se aplicó también una prueba por temas, con objeto de ver si existía correlación entre las calificaciones. Para obtener esta correlación se usó el coeficiente de correlación de Spearman, con los siguientes resultados:

En el grupo principal, la correlación obtenida fue de 7.5, significativa al 5%.

En el grupo testigo la correlación fue de .73, también significativa al 5%.

Estas correlaciones indican que tanto en el grupo principal como en el testigo, las calificaciones coincidían altamente en los dos tipos de pruebas empleadas, no siendo de esperar discrepancias que excedan de un 5%, cifra insignificante en el tipo de material que manejamos. O sea, que en las pruebas de reactivos múltiples se obtuvieron casi las mismas calificaciones que con las pruebas por temas. Después de lo expuesto al principio, los resultados parecerían contradictorios, pues era de esperarse una mayor discrepancia. Sin embargo, debe tomarse en cuenta que ambas pruebas por temas fueron proyectadas con mucho cuidado y calificadas después de manera escrupulosa en ambos grupos, en condiciones de mayor tiempo y atención que lo que es ordinario emplear y de aquí su mayor precisión.

CONCLUSIONES

Habiendo hecho un estudio detallado por las pruebas por temas y las de selección múltiple, escogimos ésta por las ventajas que señalamos al principio y podemos concluir que el examen por temas sólo debe emplearse en grupos pequeños y en los que se quiera apreciar a la vez que la información, también la capacidad expresiva, gramatical, el orden y otras cualidades que sobrepasan el margen de los conocimientos que la materia examinada exige. Requieren mucho tiempo y dedicación por parte del profesor, experiencia e imparcialidad, pero pueden funcionar adecuadamente cuando se dispone de estas posibilidades. No resultan recomendables para grupos que excedan de quince alumnos ni para exámenes departamentales, tanto por el tiempo requerido como por la dificultad para regular el criterio de calificación. Inspiran inseguridad al alumno y pueden dar lugar al sentimiento de ser tratado con injusticia, rechazo de la calificación y argumentaciones difíciles de evitar, dejando insatisfecho al alumno del trato que se le dio.

La prueba de reactivos de selección múltiple se presta a un tratamiento regular, da confianza al estudiante y le deja satisfecho en la apreciación de sus conocimientos, obviando problemas derivados del sentimiento de injusticia y argumentaciones. Permite la unificación del criterio de calificación y el establecimiento de exámenes departamentales. Su califi-

cación se realiza con rapidez y evita cansancio al profesor.

En cambio, su preparación es laboriosa y requiere no sólo de adiestramiento técnico, sino también de cierta cantidad de experiencia, ingenio e imaginación. Su calificación requiere de conocimientos de estadística elemental. Sin embargo, el tratamiento estadístico es relativamente sencillo y rápido, y en todo caso, un ayudante puede establecer en poco tiempo las curvas de calificación de varios grupos y entregar al profesor la escala de distribución de calificaciones.

No se pusieron a prueba otras técnicas porque ya han sido ensayadas en diversas partes y se tiene evidencia de las ventajas de la que empleamos. A pesar de que en los textos se dan observaciones útiles para su elaboración, casi no se toman en cuenta los factores que se analizaron en este trabajo. Habiendo observado la irregularidad en el funcionamiento de los reactivos, nos propusimos analizar en cada uno de ellos las causas de su éxito o fracaso. De aquí derivamos la serie de observaciones que se registraron y en las que intervienen elementos que tienen que ver no sólo con el carácter de la prueba, sino también con el del estudiante. Es útil conocer bien estos factores para elaborar pruebas eficaces y regulares. Creemos que las dificultades y problemas que surgen con las pruebas departamentales se derivan de la elección defectuosa de los métodos de examen, de la irregularidad de los cursos y de la falta de preparación técnica en los encargados de la elaboración de las pruebas.

Si recordamos las cualidades que adscribimos a una buena prueba, veremos que las cumple satisfactoriamente en su mayoría y mejor que otras pruebas conocidas.

1.—Cubre en una sola prueba y proporcionalmente a la importancia que quiera darse, toda el área de conocimientos del curso.

2.—Permite una valoración objetiva y libre de prejuicios, ya que se califica automáticamente y sin tomar en cuenta la redacción, el tipo y la dificultad de la letra, la capacidad de exposición, y otros factores de inexactitud en las pruebas por temas, "distractores" para el profesor.

3.—Puede ser aplicada colectivamente a grupos numerosos y calificada con rapidez, ahorrando al profesor no solamente mucho tiempo, sino fastidio y el sentimiento de hacer un trabajo rutinario y estéril.

4.—Explora realmente las áreas de conoci-

to deseadas, al eliminar factores de contaminación y ser diseñada con toda la precisión que se desee.

5.—Excluye, hasta donde es posible, los elementos perturbadores de la relación de autoridad, al ser prácticamente anónima y automática.

6.—Ofrece un grado de dificultad uniforme, ya que es medido por el propio grupo al trazar las curvas de calificación y también según la experiencia del profesor, al comparar los resultados obtenidos en los diversos grupos y en los años consecutivos.

7.—Elimina, hasta donde es posible por ahora, el elemento azar, por cubrir toda el área de conocimientos, constar de un número suficiente de distractores y utilizar éstos de modo adecuado, así como poner todos los reactivos que se desee sobre un mismo tema, para cubrirlo mejor en forma proporcionada.

8.—Elimina en grado suficiente los aciertos por deducción o inferencia cuando se carece de la información adecuada.

Este es quizá uno de los logros más difíciles de obtener, ya que no es fácil plantear una pregunta que no lleve de algún modo implícita la respuesta o al menos una parte de ella. Todo el que pregunta sabe algo de la respuesta, si no es que tiene una respuesta completa ya en su mente.

9.—Se puede lograr eliminar o atenuar el efecto de la sabiduría popular, o del "sentido común", en la respuesta a preguntas para las que no se contaba con la suficiente preparación académica.

10.—Elimina la información memorizada o simplemente automática, ya que para resolver la prueba es necesario encontrar las soluciones contenidas en material que requiere necesariamente del análisis y la comprensión para encontrar la respuesta correcta.

11.—Este tipo de prueba explora el manejo mental de la información, que no puede emplearse pasivamente sino de manera dinámica, al encontrar en los reactivos una oportunidad de aplicación y no solamente de exposición, para lo que se requiere determinado interés y actividad.

12.—Explora el lenguaje técnico, que forma parte importante de la información sobre una materia, porque en la formulación de los reactivos se emplea dicho lenguaje, que sólo será comprendido adecuadamente por los alumnos que tengan la información al respecto. Esta fue una de las observaciones derivadas del análisis de nuestro material.

13.—La prueba resultó interesante y agradable para los alumnos, que se aplicaron a ella con interés.

y olvidándose al poco tiempo de su relación con el examen de tipo tradicional.

Pensamos que la alta correlación obtenida entre la escala de calificación de la prueba por temas y la de reactivos de selección múltiple, se debe principalmente a que las pruebas por temas, en ambos grupos, principal y testigo, fueron seleccionadas y preparadas con mucho cuidado, revisando las notas calificativas, enmendando las inexactitudes y procurando no tomar en cuenta los elementos perturbadores.

Finalmente, añadiremos que las observaciones realizadas por nosotros añaden elementos prácticos para su elaboración, de significación mayor cuando se trata del examen en materias de tipo de ciencias no exactas. Ponen de relieve una vez más la continua influencia de los procesos caracterológicos en cualquier actividad y su influencia benéfica o perturbadora, tanto en la mente del alumno como en la del profesor, cuando no se tiene su advertencia. Por otra parte, la medición de su efectividad constituye un esfuerzo para sacar los exámenes de su tradicional subjetividad e inexactitud y proveerlos de pruebas ponderadas, que coloquen el acto de examinar a un alumno en el mismo nivel didáctico con que se

procura enseñarle. Resulta un anacronismo pretender ser didácticos en un aspecto y sigamos actuando en otro de manera anticuada y paracientífica, desvirtuando con nuestros métodos de examen lo mismo que pretendemos enseñar. El hecho mismo de no poder prescindir aún de los exámenes está señalando una falla en los procedimientos de enseñanza. En un buen sistema, el aprendizaje debe efectuarse gradualmente y de modo tal que un paso conduzca al siguiente y estar en determinado lugar del curso signifique la capacidad para haberlo alcanzado, con la posibilidad y el derecho de proseguir. Pero esto, escapa por ahora a nuestras posibilidades, al menos en los centros de estudio de nivel profesional. Pero deberemos seguir pensando en el modo de conseguir que los exámenes resulten inútiles y el alumno pueda cursar, aprobar o detenerse sin necesidad de introducir un elemento extraño de carácter amenazante y perturbador, logrando que el elemento prueba se amalgame a la enseñanza, sin que puedan distinguirse el uno del otro y que un estudiante se pueda reconocer como situado en una determinada parte del camino, que podrá seguir recorriendo o no, según la capacidad que logre desarrollar con su esfuerzo y orientado por su profesor.