

Un paso hacia la evaluación formativa

Gabriel Carrillo*

Magdalena Carrillo*

Introducción

“La evaluación del proceso enseñanza-aprendizaje es el acopio sistemático de datos cuantitativos y cualitativos, que sirve para determinar si los objetivos de aprendizaje se están realizando en los alumnos”.¹

Este juicio de valor se puede hacer de manera más objetiva midiendo aquellos ámbitos del educando en los que se espera un cambio de actitud o un aprendizaje.

Para efectuar estas mediciones se cuenta con varios instrumentos como son: entrevistas, cuestionarios, pruebas, trabajos, escalas, técnicas proyectivas, técnicas sociomédicas y la observación directa entre otras.

De todos estos instrumentos, en el presente estudio se tratará sobre aquellas pruebas denominadas como de “respuesta cerrada” debido a que han demostrado ser las más prácticas para evaluar a grandes grupos de población como es el caso de los 1,200 alumnos por semestre que se atienden en la Facultad de Medicina de la Universidad Nacional Autónoma de México.

Desde hace más de 10 años los exámenes departamentales de bioquímica en esta Facultad han sido del tipo de respuesta cerrada, ya sea opción múltiple, correlación u ordenamiento; y los datos son procesados por computadora, de donde es obtenido el análisis estadístico básico de cada grupo y de la población en general (media, desviación típica, etc.).

Además se obtiene un análisis de cada pregunta en cuanto a su grado de dificultad e índice de discriminación.

En términos generales, hay consenso entre los profesores y autoridades, en relación a que con estos datos se puede tener una imagen estadística del examen aplicado y de la población a la que se le aplicó.

Es pertinente indicar que durante el curso los maestros suelen aplicar exámenes parciales utilizando además otros instrumentos de evaluación como: exámenes orales, por temas, o semiestructurados, que junto con el examen departamental se utilizan para hacer la evaluación final.

Planteamiento del problema

Con el empleo de pruebas de falso y verdadero durante los exámenes parciales (en las cuales se deben descontar los errores de los aciertos para darle validez), se detectó que los alumnos tienen la tendencia a contestar todas las preguntas de una prueba de respuesta cerrada independientemente de que sepan o no las respuestas correctas.

Se pretendió entonces que en el examen departamental sólo se diera respuesta a aquellas preguntas que el alumno domina, evitando contestaciones al azar, para lo cual se aplicó el criterio de descontar los errores de los aciertos.

Esto provocó descontento entre los alumnos, a quienes pareció injusto por parte del maestro, que se les tomara en cuenta los errores cometidos.

Lo anterior indica que no tienen conciencia de que para fines de evaluación del rendimiento escolar (entendida en su cabal significado) es tan importante acertar como no equivocarse, especialmente tratándose de estudiantes de medicina.

Nos encontramos ante un problema doble:

* Departamento de Bioquímica, Facultad de Medicina, UNAM.

uno de información, en el que los alumnos deben tener dominio real sobre los conceptos y datos de los contenidos programáticos; como de formación, ya que los alumnos deben tener una actitud responsable en el momento que se someten a una prueba.

Si tales condiciones no se cumplen, es muy probable que el sistema de evaluación, tal y como se maneja actualmente, incida negativamente en la formación del estudiantado.

La sola posibilidad de que esto suceda, nos enfrenta a la siguiente interrogante: ¿Es posible que los exámenes tradicionales de rendimiento escolar, puedan además construirse en instrumentos formativos?

Objetivos

Medir el grado de certeza que tienen los estudiantes al contestar un examen.

Buscar si la puntuación corregida nos ayuda a mejorar el sistema de medición del rendimiento escolar.

Averiguar la magnitud del aprendizaje equivocado que se puede generar durante el proceso enseñanza-aprendizaje.

Definición de términos

Puntuación

a) total = número de aciertos obtenidos en la prueba.

b) corregida = número de aciertos obtenidos en la prueba menos el número de errores cometidos.

Respuestas seguras

Respuestas dadas por el alumno con la certidumbre de haber elegido la alternativa correcta, que puede haber sido contestada acertada o erróneamente.

Respuestas dudosas

Preguntas en las cuales el alumno expresa su inseguridad de que la alternativa elegida haya sido la correcta, que también pueden haber sido contestadas acertada o erróneamente.

Sin contestar

Preguntas que el alumno no relaciona con

alguna de las alternativas de respuesta dadas en la prueba.

Aprobados

Alumnos que tuvieron 42 o más aciertos en la prueba.

Reprobados

Alumnos que obtuvieron 41 o menos aciertos en la prueba.

Sujetos

Fueron estudiados los cuestionarios resueltos por 53 alumnos que cursan el 2o. semestre de Médico Cirujano en la Facultad de Medicina de UNAM; estos estudiantes forman parte de dos grupos, con diferente profesor y horario (uno matutino y otro vespertino).

Instrumento

Se utilizó el primer examen departamental de Bioquímica para medir el rendimiento escolar de los dos primeros meses del curso.

Este examen departamental:

- Consta de 70 preguntas.
- Se triplica simultáneamente a toda la población que cursa ese semestre.
- La puntuación bruta es transformada en calificación por el profesor tomando en cuenta el siguiente criterio general:

Calificación		Puntaje
No acreditada	(NA)	0-41
Suficiente	(S)	42-51
Bien	(B)	52-61
Muy bien	(MB)	62-70

- Los alumnos contestan en tarjetas perforadas para que éstas sean procesadas directamente por la computadora.

La computadora realiza un análisis estadístico del examen con base en las respuestas de toda la población a la que fue administrado, que en resumen son los siguientes:

Medía aritmética	37.93
Desviación típica	10.85
Confiabilidad	0.89
Error típico	3.59
Discriminación	0.98
Grado de dificultad	medio
Puntaje al azar esperado	14
No. de alumnos examinados	846

Tabla I

Grupo I									
	Puntuación		Respuestas seguras			Respuestas dudosas			Sin
	Total	Corregida	Total	Aciertos	Errores	Total	Aciertos	Errores	contestar
x Aprobados	53.40	42.40	63.0	52.0	11.0	2.40	1.40	1.0	4.8
y Reprobados	31.55	5.63	57.09	31.0	26.09	0.91	0.55	0.36	10.64
t	7.19	5.22	1.61	6.47	-4.15	4.69	1.93	2.46	-4.72
tablas t 0.05 = 2.093 t 0.01 = 2.861 gl = 19									
Diferencia significativa	+	+	-	+	+	+	-	+	+
Grupo II									
	Puntuación		Respuestas seguras			Respuestas dudosas			Sin
	Total	Corregida	Total	Aciertos	Errores	Total	Aciertos	Errores	contestar
x Aprobados	52.86	42.66	57.0	46.80	10.20	12.0	6.06	5.93	1.13
y Reprobados	32.0	9.23	48.7	25.94	22.76	15.0	6.05	8.94	6.88
t	8.14	6.46	1.96	8.44	-4.91	-1.88	1.22	-3.14	-3.19
tablas t 0.05 = 2.042 t 0.01 = 2.750 gl = 30									
Diferencia significativa	+	+	-	+	+	-	-	+	+
Columna	A	B	C	D	E	F	G	H	I

Indicaciones especiales

A los alumnos de los grupos I y II, además de las indicaciones propias del examen, se les dio la siguiente instrucción antes de resolver la prueba: Toda respuesta equivocada restará un punto del total de aciertos, a menos que sea manifestada como respuesta dudosa.

Presentación de datos

Se extrajeron los siguientes datos de cada uno de los exámenes presentados por los alumnos: (ver tabla I)

1. Puntuación total (columna A tabla I)
2. Las respuestas seguras erróneas (columna E tabla I)
3. Las respuestas dudosas acertadas (columna G) y erróneas (columna H) y la suma de estas dos (columna F)
4. Las respuestas sin contestar (columna I)
Y se determinaron:
 1. La puntuación corregida (columna B)
 2. Las respuestas seguras acertadas (puntuación total menos aciertos dudosos columna D)
 3. Respuestas seguras totales que son: respuestas seguras acertadas más respuestas seguras erróneas (columna C)

Dado que el alumno tenía la opción de marcar una respuesta como dudosa, o dejarla sin contestar; llamaremos a las respuestas erróneas como aprendizaje equivocado.

Para el estudio se subdividió cada grupo en 2 partes. El subgrupo de aprobados; o sea, con 42 aciertos o más y el subgrupo de reprobados con menos de 42 aciertos.

Una vez clasificados los datos de esta manera, se procedió a determinar la media de cada dato para cada subgrupo y el estadígrafo *t* de Student para ver si las diferencias que mostraban los subgrupos eran significativas (ver tabla I).

Se calcularon además la media y desviación típica de la puntuación total y corregida para los dos grupos mostrados en la tabla 2.

Análisis de las tablas de resultados

Para la columna de puntuación total (A) y corregida (B) se encontró como era lógico una diferencia significativa entre el grupo de aprobados y reprobados.

Pero en las puntuaciones corregidas, por el hecho de que prácticamente se duplique la desviación típica (tabla II) podríamos suponer que el uso de la puntuación corregida disminuye la tendencia central de las puntuaciones en este tipo de exámenes y por lo tanto permite distinguir más claramente a los que saben de los que no saben.

Para la columna de respuestas seguras totales (C) no se rechazó la hipótesis nula que dice: los alumnos reprobados tienen certeza de sus respuestas en el mismo número de preguntas que los aprobados.

Para la columna de respuestas seguras acertadas (D) existe una diferencia significativa entre el número de aciertos obtenidos entre acreditados y reprobados.

Para la columna E, respuestas dudosas erróneas también hay diferencia significativa entre los subgrupos.

Para la columna de respuestas dudosas totales (F) un grupo presenta diferencia significativa y el otro grupo no.

Para la columna G respuestas dudosas acertadas no hubo diferencia significativa.

La columna H respuestas dudosas erróneas sí hubo diferencia significativa, pero el grupo I son los aprobados quienes cometieron más errores mientras que en el grupo II fueron los reprobados.

Interpretación

Llama la atención la columna B de la tabla I, en donde la media de los reprobados baja desproporcionalmente en relación a la media de los aprobados que baja mucho menos.

En la tabla II el hecho de que prácticamente se dupliquen las desviaciones típicas, podríamos suponer que el uso de la puntuación corregida disminuye la tendencia central de las puntuaciones y por lo tanto permite distinguir más claramente a los dos subgrupos.

Por esta razón es conveniente hacer estudios más detallados sobre el uso de la puntuación corregida.

Son interesantes también los resultados obtenidos en la columna C de respuestas seguras totales de donde se desprende que no existe diferencia significativa entre la seguridad con que responden los aprobados y reprobados

TABLA II

	Puntuación total		Puntuación corregida	
	x	s	x	s
Grupo I	41.95	12.77	23.14	23.47
Grupo II	41.78	12.30	24.91	21.68

sino que la diferencia está en el número de aciertos y errores. Lo cual nos lleva al comentario de la columna E de respuestas seguras erróneas que es factible de asociar con aprendizaje equivocado o mal entendido de los contenidos programáticos.

Usualmente el aprendizaje equivocado no se puede distinguir de la inseguridad o ausencia de aprendizaje, es decir, el aprendizaje equivocado (E), el aprendizaje inseguro (F) y la ausencia de aprendizaje (I) son tomados como una misma entidad.

En este estudio resulta sorprendente el número de *Respuestas seguras erróneas* en el grupo de los aprobados que alcanza más del 15 por ciento y alarmante en el caso de los reprobados ya que es más del 40 por ciento, cuando la cifra debería tender a *cero*.

Proyección del trabajo

Además del conocimiento adquirido, este estudio nos deja una inquietud sobre la responsabilidad social que tiene el sistema educativo en relación a la demanda de profesionistas eficientes y eficaces que requiere actualmente la sociedad mexicana.

Así como el médico al recibirse adquiere un compromiso con la sociedad, la Facultad de Medicina tiene a su cargo la responsabilidad de habilitar profesionistas bien capacitados para mantener la salud de los ciudadanos.

Es indispensable que el médico delimite claramente los conocimientos que tiene del campo en el que se desenvuelve y perciba con claridad lo que no sabe.

Esta actitud que parece evidente y se considera como parte de la formación del individuo no siempre es alentada por el sistema educativo, o dicho de otra forma hay cierta incongruencia entre fines y procedimientos

del sistema que hacen prevalecer ciertos aspectos negativos que es necesario corregir.

Al parecer los sistemas de evaluación usados de la misma forma en que se ha venido haciendo durante mucho tiempo, desde la primaria, secundaria, preparatoria y hasta la escuela profesional, arraigan en el individuo la idea de que no importan los errores que se cometan siempre y cuando se acierte a un mínimo de 60 por ciento de los casos.

¿No resulta un poco ingenuo formar a individuos bajo la norma de un "aceptable" 60 por ciento de eficiencia y al cabo de 16 ó 17 años exigir o esperar que trabaje bajo una norma superior?

Esta situación además de desalentar la estricta autocrítica hacia el propio cumplimiento de las responsabilidades asumidas, causa un diferencial entre lo que se espera de un profesionista y lo que está dispuesto a hacer cuando se establece un compromiso de trabajo.

Si se logra hacer que el alumno tome conciencia de la importancia que tiene no cometer errores, reconocer y aceptar la que desconoce, para que sea precisamente en esa área en la que se capacite, se dará un paso muy importante en la formación de las futuras generaciones. □

Bibliografía

1. Paquete de autoenseñanza de evaluación del aprovechamiento escolar. Comisión de nuevos métodos de enseñanza UNAM 1977.
2. Surber, R.J., y Philip, L.S.: Testing for misunderstanding. *Educational psychologist*. Vol. 16, no. 3, 165-174, 1981.
3. Lemus, L.A.: Evaluación del rendimiento escolar. Ed. Kapelus, Buenos Aires, 1974.
4. Cuadernos del curso: Objetivos y evaluación, del centro de didáctica, UNAM, 1975.
5. Downie, N.M., y Heath, R.W.: Métodos estadísticos aplicados. Ed. Harla, México, 1973.
6. Ary, D.; Jacobs, L.Ch., y Razavieh, A.: Investigación pedagógica. Ed. Interamericana, México, 1982.