

EVALUACIÓN DE UN EXAMEN PARCIAL DE BIOQUÍMICA*

Yolanda Saldaña Balmori¹, Héctor Javier Delgadillo Gutiérrez², Ignacio Méndez Ramírez³

¹Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México (UNAM). Av. Universidad 3000, Col. Universidad Nacional Autónoma de México, Delegación Coyoacán, México, D. F. CP 04510. Correo E: balmori@bq.unam.mx.

²Departamento de Sistemas Biológicos. Universidad Autónoma Metropolitana, Unidad Xochimilco. Calzada del Hueso 1100. Col. Villa Quietud, 04960, Del. Coyoacán D. F. Correo E: hjdelga@correo.xoc.uam.mx

³Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, UNAM, Av. Universidad 3000, Col. Universidad Nacional Autónoma de México, Delegación Coyoacán, México, D. F. CP 04510. Correo E: imendez@servidor.unam.mx

RESUMEN

Se evaluó un examen parcial de bioquímica aplicado recientemente utilizando las siguientes técnicas para su validación: 1) Índice de dificultad (π), 2) Índice de discriminación (D_i), 3) Coeficiente de discriminación ($rpbis$) y 4) Alfa de Cronbach. El estudio fue retrospectivo, descriptivo, transversal y observacional por lo que pertenece a una encuesta descriptiva. Se evaluaron las respuestas dadas por 1321 estudiantes y el examen constó de cinco temas. El total de reactivos de evaluación fue de 70, de los cuales 5 para termodinámica, 25 para agua pH e iones, 15 para química de proteínas, 15 para enzimas y con 10 se evaluaron las prácticas de laboratorio. Con el estudio de índice de dificultad se encontró que hay 17 reactivos o muy fáciles o muy difíciles, con el índice de discriminación se identificó que hay que revisar 11 reactivos y finalmente con el coeficiente de discriminación, el número a revisar fue de 35. Con el alfa de Cronbach se encontró que 28 reactivos deberán modificarse o eliminarse. El alfa de Cronbach fue obtenida con el paquete estadístico SPSS; el índice de dificultad, el índice y el coeficiente de discriminación se obtuvieron con el programa Excel. Se concluye que es factible evaluar mediante las técnicas aquí propuestas los exámenes de la asignatura de Bioquímica del Departamento de Medicina de la UNAM.

ABSTRACT

We assessed a recently applied partial evaluation in the subject of Biochemistry, using this techniques for its validation: 1) Index of difficulty (π), 2) Index of discrimination (D_i), 3) Coefficient of discrimination ($rpbis$) and 4) Cronbach's alpha. The study was retrospective, descriptive, transversal, and observational, that is, it corresponds to a descriptive survey. The examination was administered to 1321 students and consisted of five themes with a total of 70 items, distributed 5 for thermodynamics, 25 for water, pH, and ions, 15 for proteins, 15 for enzymes, and 10 for laboratory sessions. The index of difficulty indicated 17 items to be either very easy or very difficult, the index of discrimination yielded 11 items that have to be revised, and, finally, the coefficient of discrimination yielded 35 items to be revised. Cronbach's alpha revealed that 28 items have to be removed. Cronbach's alpha was obtained with the SPSS statistics software, and the difficulty and discrimination index, as well as the coefficient of discrimination, were obtained with Excel. We conclude that it is feasible to assess, by means of the proposed techniques, the examinations for the subject of Biochemistry at the School of Medicine (UNAM).

PALABRAS CLAVE:

Evaluación del aprendizaje, enseñanza de la bioquímica.

KEY WORDS:

Evaluation of the learning, education of the biochemistry

INTRODUCCIÓN

Históricamente el proceso de evaluación educativa ha sido un evento complejo pues en él intervienen muy diversos aspectos, sólo por mencionar algunos, se tiene que en ocasiones se ha priorizado a la enseñanza *vs.* el aprendizaje y es éste último, el que realmente marca la pauta de las metas de enseñanza alcanzadas; otro aspecto de la complejidad es la formación previa del aprendiz, pues para que adquiera el conocimiento en un tópico determinado se tuvieron que haber adquirido conocimientos previos para hacer viables los nuevos y finalmente entre otros muchos más factores participantes, se debe señalar que la afinidad entre enseñante y aprendiz es determinante pues de ello depende -en muchas ocasiones- el éxito de la adquisición de conocimiento.

La manera en la que se puede comprobar que se está ejerciendo la enseñanza es mediante la adquisición del aprendizaje obtenido por los estudiantes y las evaluaciones realizadas dan evidencia de la evolución del proceso enseñanza-aprendizaje. Además la evaluación es un componente fundamental dentro del proceso educativo ya que ayuda a valorar si se logran los objetivos propuestos. Jaap y cols. en 2003 señalan que en una evaluación se requiere: 1) que haya validez de contenido, 2) que haya validez de constructo teórico para identificar si el instrumento de evaluación está a la altura de las pretensiones, es una medida de qué tan bien se mide al constructo con ese instrumento, 3) que se apoye en una estructura lógica y congruente de los conocimientos, 4) que sea confiable, 5) que sea apropiada y viable en términos de tiempos y 6) que sea transparente en cuanto a los contenidos y a la forma de calificar (1).

Por otro lado la evaluación educativa cumple con el objetivo fundamental -que ya se había mencionado- de medir el aprendizaje adquirido por el estudiante, pero en primera instancia ésta debe servir para que el estudiante se dé cuenta, en función de las calificaciones obtenidas, si su técnica de estudio es eficiente o debe cambiar por otra que presume le proporcionará mejores resultados, si deberá intensificar el estudio, apoyarse en otros materiales que le ayuden a comprender la temática básica, etc., en segundo lugar para que el evaluador conozca, en función de las respuestas dadas por los estudiantes, si lo que está evaluando coincide con los objetivos planteados, con la dinámica de enseñanza utilizada, con los contenidos, etc. y en tercer término la evaluación tiene la finalidad de que el sistema escolarizado se entere y de ahí se tomen decisiones de la promoción o no del estudiante.

Los encargados de la evaluación son los que toman las decisiones, los que definen qué es normal, relevante, adecuado o bueno, en relación al comportamiento de los estudiantes así como a los contenidos que deben cubrir los resultados de su aprendizaje, además de los tiempos en los que se debe adquirir el mismo (2).

El resultado terminal conduce a emitir una calificación, misma que debe ser objetiva, que exprese una correcta correspondencia con la calidad de lo asimilado y que disminuya o elimine la influencia de un factor subjetivo mediante el cual el profesor pueda otorgar calificaciones diferentes a dos estudiantes ante resultados iguales. Los instrumentos de evaluación deben cumplir con el requisito de que su contenido esté directamente relacionado con los objetivos educacionales para que al aplicarlos demuestren validez y confiabilidad (3).

Es por ello que en la elaboración de los instrumentos de evaluación hay una gran responsabilidad social ya que entre otros, involucra el éxito o fracaso escolar, una diversidad en el rendimiento educativo, buenos o malos estudiantes y profesores, diferente calidad de la enseñanza, apreciación o no de excelencia institucional. El Centro Nacional de Evaluación de la Educación Superior (CENEVAL) fue creado con el propósito de regular el ingreso a las escuelas preparatorias y a las universidades. Es conveniente que todas las instituciones de educación tengan una evaluación validada con estándares de calidad y en caso de que esto se lleve a cabo, ayudaría a mejorar la enseñanza, pues con los resultados obtenidos, se puede indicar qué parte de los exámenes están bien hechos y en cuál hay que hacer mejoras para alcanzar la excelencia que tanto se busca en la educación.

DESARROLLO

En la Facultad de Medicina de la Universidad Nacional Autónoma de México (UNAM), desde hace poco más de 50 años se han venido aplicando en las materias de los ciclos básicos, exámenes de selección múltiple para evaluar el conocimiento de los estudiantes. Este tipo de evaluaciones presenta ventajas cuando se necesita saber el conocimiento que han adquirido los estudiantes de grandes poblaciones ya que agiliza los tiempos para la obtención de resultados, además de que es eficaz para valorar si en todos los grupos se ha cubierto los objetivos, pero, al mismo tiempo presenta desventajas ya que con algunas preguntas no siempre se puede valorar el nivel de conocimiento obtenido, la posibilidad de que por azar se obtengan respuestas correctas, la facilidad de copiar respuestas, la posibilidad de que haya preguntas que sean conocidas y divulgadas

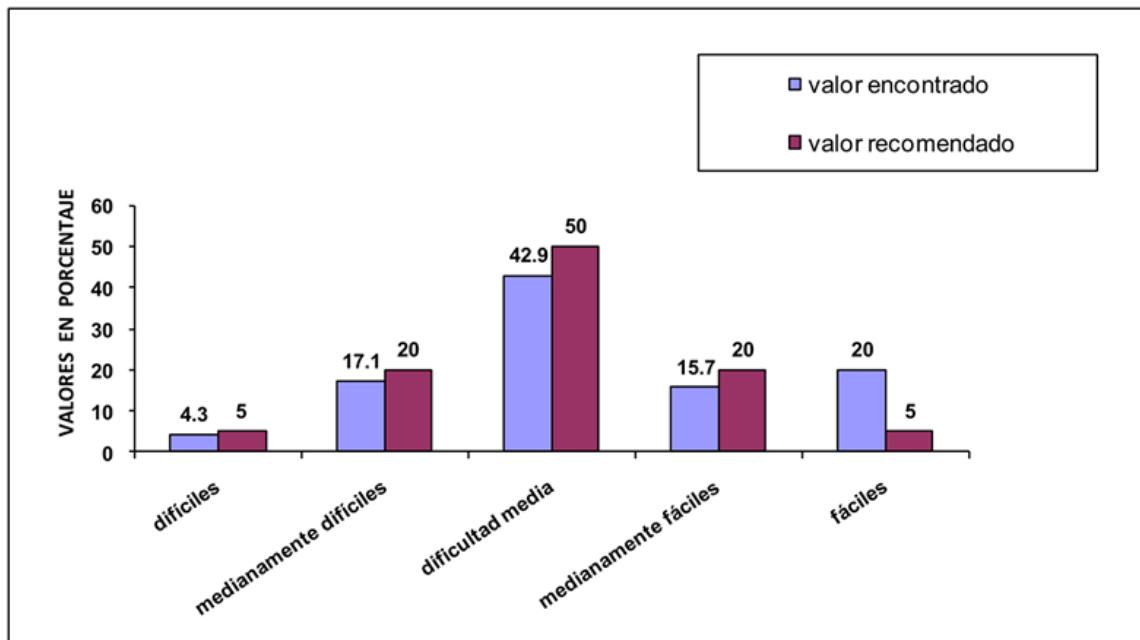


Figura 1. Índice de dificultad (Pi) de los reactivos de evaluación. El Pi se calcula dividiendo el número de estudiantes que contestaron correctamente un reactivo entre el número total de los que lo contestaron. Los valores obtenidos del índice de dificultad (Pi) son: difíciles < 0.3; medianamente difíciles de 0.31 a 0.5; de dificultad media de 0.51 a 0.7; medianamente fáciles 0.71 a 0.8; fáciles > 0.8. El número sobre las barras indica el número de reactivos. Los valores recomendados son sugeridos por (4).

entre los estudiantes antes del examen, etc.

Dentro del proyecto de investigación educativa relacionado con el rendimiento de aprendizaje adquirido por los estudiantes de Bioquímica y Biología Molecular en la Facultad de Medicina de la UNAM, se presenta en este trabajo, el análisis realizado a las respuestas dadas por los estudiantes en un primer examen departamental en el que se cubrieron los siguientes temas: Termodinámica, agua pH e iones, química de proteínas, enzimas y prácticas de laboratorio relacionadas con los temas estudiados. Este trabajo se llevó a cabo basándonos en el Examen de Habilidades y Conocimientos Básicos (EXHCOBA) que se usa en gran escala en México para conocer los estándares de calidad (4).

Las técnicas utilizadas para llevar a cabo este análisis fueron cuatro, las tres primeras conforme lo plantean Backhoff y colaboradores (4): 1) Índice de dificultad (Pi) que sirve para identificar qué tan fácil o difícil es un reactivo en sus respuestas y se obtiene dividiendo el número de estudiantes que contestaron bien un reactivo entre el total de los que lo contestaron; si se obtiene un valor <0.3 son difíciles, de 0.31 a 0.5 son medianamente difíciles, de 0.51 a 0.7 son de dificultad media, de 0.71 a 0.8 son medianamente fáciles y arriba de este valor son fáciles (Fig.1); 2) Índice de discriminación (Di) donde se analizan el 27% de alumnos que tuvieron la

calificación más alta en contraste con el 27% de los que la obtuvieron más baja (4), se considera como aceptables los valores superiores a 0.3 (Fig. 2); 3) Coeficiente de discriminación (rpbis) el cual tiene la ventaja de incluir la dispersión y el promedio de cada uno de los reactivos en el 100% de los alumnos y de las respuestas por lo que se considera la técnica más adecuada, un valor satisfactorio es de 0.26, aunque lo ideal es a partir de 0.36 (4) (Fig.3) y 4) alfa de Cronbach que consiste en evaluar qué tan confiable es un examen en cada uno de sus temas e indica cuáles reactivos hay que eliminar para que se incremente su validez.

OBJETIVOS

Demostrar que existen formas mediante las cuales se puede calificar el instrumento con el que se evalúa el aprendizaje de los estudiantes validando los reactivos de los exámenes mediante diferentes técnicas.

Calcular el índice de dificultad (Pi), el índice de discriminación (Di), el coeficiente de discriminación (rpbis) así como el alfa de Cronbach para cada una de las 70 preguntas del primer examen departamental.

Enlistar los reactivos ordenados por temas e indicar en cuáles de ellos se recomienda una revisión.

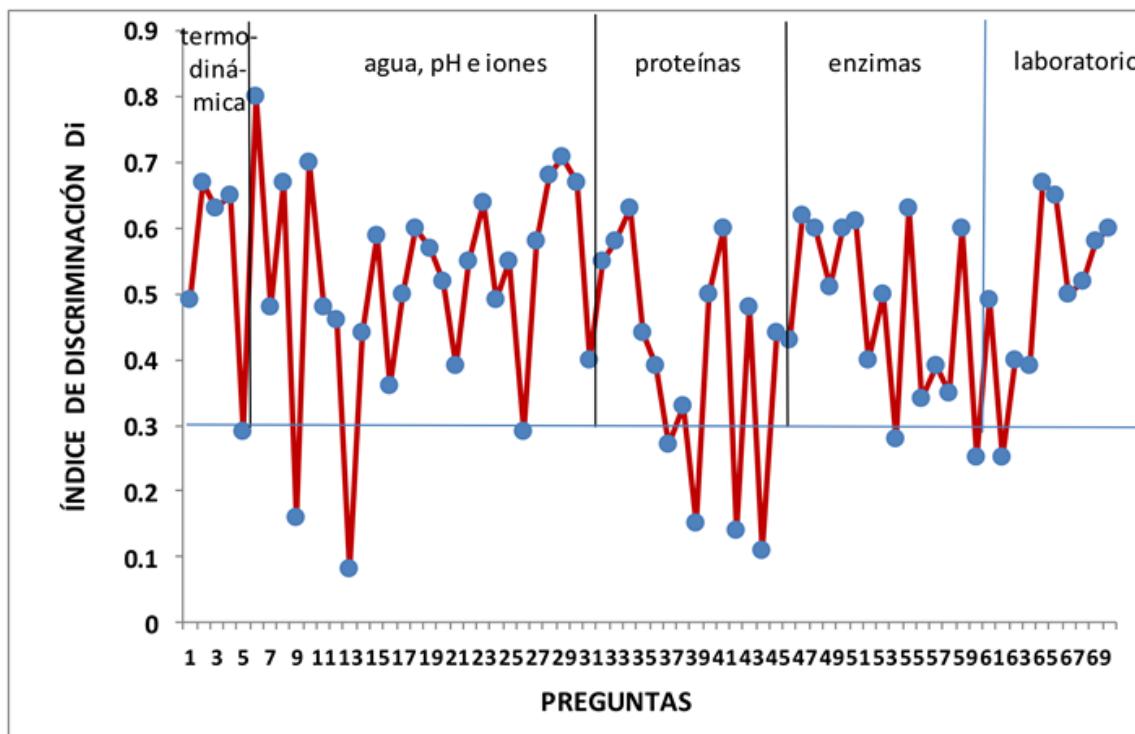


Figura 2. Índice de discriminación (Di). Índice de la discriminación (Di) de los reactivos , el recomendado es a partir de 0.3, según lo indicado en la figura la línea horizontal.

MÉTODO

El examen parcial de Bioquímica y Biología Molecular recientemente aplicado lo presentaron 1321 alumnos. La unidad de estudio fue el promedio de la calificación que los alumnos obtuvieron en cada uno de los 70 reactivos, por lo mismo fueron 70 unidades de estudio las cuales estuvieron distribuidas de la siguiente manera: 5 reactivos de termodinámica, 25 de pH, agua e iones, 15 de proteínas, 15 de enzimas, y 10 de práctica de laboratorio. El tipo de estudio fue retrospectivo, observacional, transversal y descriptivo por lo que se encuadra en una encuesta descriptiva (5)

El índice de dificultad (P_i), el de discriminación (Di) y el coeficiente de discriminación ($rpbis$) se obtuvieron de la técnica para medir el examen de conocimientos básicos (EXHCOBA) de Backhoff y col. (4). Los valores recomendados por estos y otros autores para el índice de dificultad son del 5% para los reactivos fáciles, 20% para los medianamente fáciles, 50% con una dificultad media, 20% medianamente difíciles y 5% difíciles; los valores recomendados para el índice de discriminación son con un valor de $Di > 0.30$ y los recomendados para el coeficiente de discriminación fueron con un $rpbis > 0.26$. El alfa de Cronbach se obtuvo de acuerdo al manual de SPSS de Camacho

(6) donde se recomienda quitar los reactivos que bajan el valor del alfa.

RESULTADOS

Índice de dificultad. Respecto al índice de dificultad un 4% de los reactivos fueron difíciles, 17% medianamente difíciles, 43% de dificultad media, 16% medianamente fáciles y 20% fáciles. Al contrastarlos con los valores recomendados, los reactivos fáciles fueron los que más discreparon, con respecto a los demás índices de dificultad en donde los valores encontrados son semejantes a los recomendados. Según este análisis 17 reactivos habrá que revisarlos para modificarlos o eliminarlos. (Fig. 1 y Tabla 1).

Índice de discriminación. En lo que concierne al índice de discriminación se encontró un 9% de reactivos deficientes, 7% regulares, 11% buenos y 73% excelentes. (Fig. 2). Según este análisis 11 reactivos deben ser eliminados o modificados.

Coeficiente de discriminación. En este coeficiente se obtuvieron el mayor número de reactivos con una discriminación negativa, esta fue del 22%, un 11% fue pobre, 16% regular, 11% buena y un 40% tuvieron un excelente poder de discriminación,

TABLA 1 CLASIFICACIÓN DE LAS PREGUNTAS SEGÚN SU ÍNDICE DE DIFICULTAD.

TEMAS	DIFÍCILES	MEDIANAMENTE DIFÍCILES	DIFICULTAD MEDIA	MEDIANAMENTE FÁCILES	FÁCILES	TOTAL
TERMODINÁMICA	0	5	1,2	0	3,4	5
AGUA, pH E IONES	13	12,16,21	7,8,9,11,14, 15,18,19,20, 23,24,27	6,17,22,25,26	10,28,29,30	25
PROTEÍNAS	38,43	36,41	31,35,37, 39,42,44	33,45,	32,34,40	15
ENZIMAS	0	55,57,59	48,50,51,53, 54,56,58,60	0	46,47,49,52	15
LABORATORIO	0	61,62,66	65,70	63,67,68,69	64	10
TOTAL	3	12	30	11	14	70

Los números incluidos son los asignados a cada una de las 70 preguntas.

Según este análisis 36 reactivos sería conveniente eliminarlos o modificarlos. (Fig. 3).

Alfa de Cronbach. El alfa de Cronbach tiene una escala entre 0 y 1. Entre más cercano esté el valor a la unidad tendrá más confiabilidad. En la Tabla 2 se señalan los 28 reactivos del total de 70 que deberán modificarse o eliminarse para que dé un valor más alto y por lo tanto más confiable.

En la figura 4 se observa que de los 70 reactivos

del examen 16 de ellos (23%) estuvieron bien elaborados ya que no se les encontró ninguna falla en las cuatro técnicas aquí revisadas; 31 (44%) tuvieron una sola recomendación para su mejoramiento; 13 reactivos (19%) tuvieron dos recomendaciones y finalmente 10 de ellos (14%) alcanzaron 3 recomendaciones para su mejoramiento.

De los resultados obtenidos en este estudio se concluye que los 16 reactivos de evaluación que no tienen sugerencias de corrección, sumadas a las

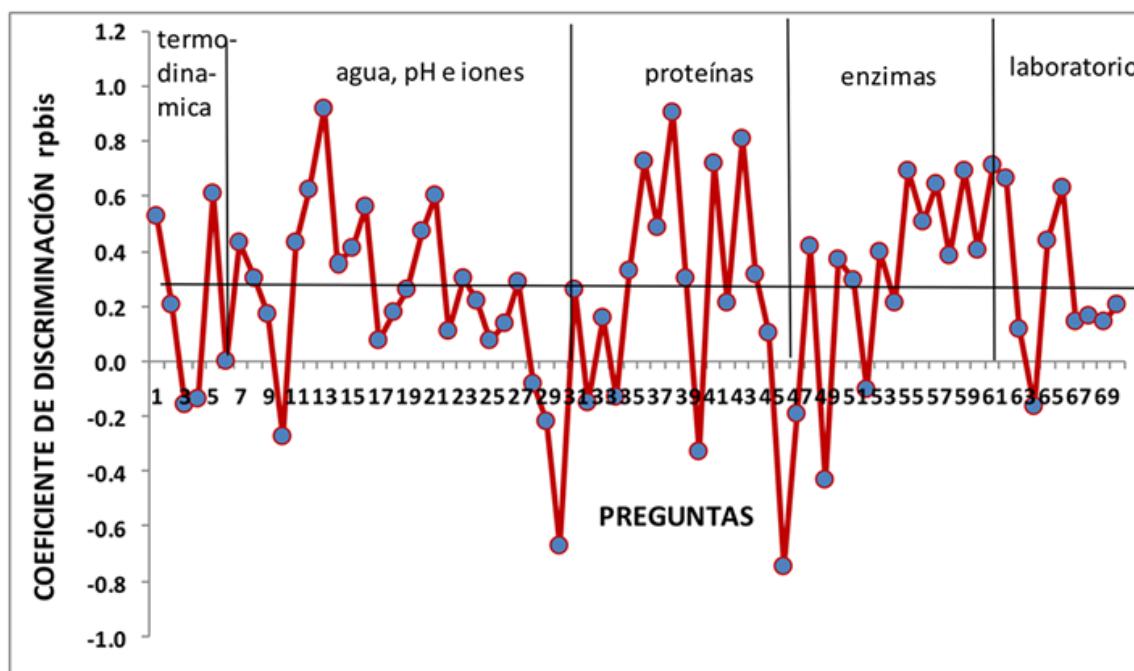


Figura 3. Coeficiente de discriminación (rp bis). El coeficiente de discriminación satisfactorio es a partir de 0.26, tal como lo indica la raya horizontal en la figura, el coeficiente de discriminación ideal es a partir de 0.36.

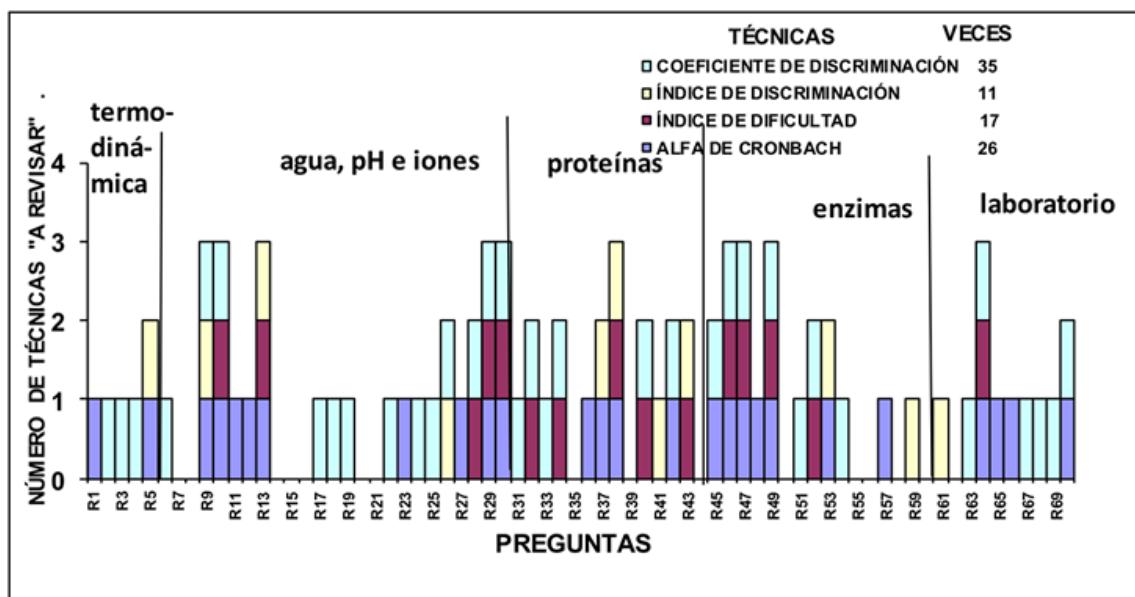


Figura 4. Técnicas con las que fueron evaluados los reactivos. Los resultados obtenidos después de haber realizado el estudio indican que los reactivos bien elaborados se muestran en la figura con espacios en blanco, las que tienen 1, 2 ó 3 sugerencias de corrección se muestran con una, dos o tres barras encimadas. Los datos que se incluyen se refieren al número de veces que las 4 diferentes técnicas arrojan datos para que los reactivos deban ser modificados.

31 que tienen una sola, se pueden incorporar a un banco de reactivos para su futura utilización. Por ejemplo el siguiente reactivo no tiene sugerencia de corrección:

14.- Identifica el espacio líquido que tiene como amortiguador principal al par bicarbonato/ácido carbónico. La respuesta correcta es Plasma, la cual se eligió de un grupo de opciones que son: a) Orina b) Plasma c) Agua total d) Líquido intersticial y e) Líquido intracelular.

Por otro lado, hay 10 reactivos que tienen 3 sugerencias de correcciones, razón por la cual no se consideran aptos en la manera en la que están redactados, ya sea porque los resultados indican que son fáciles o difíciles (Tabla 1).

Dentro del grupo de los reactivos fáciles, entre otros, están los siguientes: 10, 29, 46, 47.

46.- Molécula sobre la que actúa la enzima. La respuesta correcta es Sustrato que se eligió del grupo de respuestas a) Enzima b) Sustrato c) Inhibidor competitiva d) Inhibidor no competitiva y e) pH y temperatura.

Mientras que los reactivos 38 y 43 son difíciles.

38.- Reconoce la estructura de las proteínas que tienen como único enlace covalente al puente disulfuro entre cadenas. La respuesta correcta es Estructura cuaternaria y se eligió del grupo de opciones a) Estructura primaria b) Estructura secundaria c) Estructura terciaria d) Estructura cuaternaria y e) Estructura nativa.

ria d) Estructura cuaternaria y e) Estructura nativa.

DISCUSIÓN

Para el análisis de este examen se utilizó el Examen de Habilidades y Conocimientos Básicos (EXHCOBA) (4) debido a que está validado, es confiable y evalúa tres formas para calificar cada uno de los reactivos del examen, además se incrementó con el alfa de Cronbach, que es un parámetro de confiabilidad.

El objetivo principal de este trabajo es demostrar que existen formas para identificar la calidad de los reactivos de evaluación y para medir el aprendizaje de los estudiantes.

Es importante notar que estas técnicas dan diferentes resultados (Figs. 2, 3 y Tabla 1), pero la más confiable de las cuatro es el coeficiente de discriminación (Fig. 3) porque evalúa el 100% de los reactivos así como al 100% de los alumnos. Además, arrojó mayor información ya que permite identificar aquellos ítems o reactivos que requieren modificaciones, pues no obtuvieron valores aceptables. Lo conveniente de acuerdo a esta técnica de análisis es que en un examen el coeficiente y el índice de discriminación -que como ya se mencionó- analiza el 27% de alumnos que tuvieron la calificación más alta y el 27% de los que tuvieron la calificación más baja, siguieran una ruta paralela puesto que así se puede ver que en lo relativo al

TABLA 2 REACTIVOS QUE SEGÚN EL ÍNDICE DE CONFIABILIDAD DEL ALFA DE CRONBACH DEBEN SER REVISADOS PARA SU FUTURA UTILIZACIÓN.

TEMAS	REACTIVOS	ALFA DE CRONBACH	REACTIVOS A REVISAR
TERMODINÁMICA	del 1 AL 5	0.552	1 y 5
AGUA, pH E IONES	del 6 al 30	0.761	9 a 13, 23, 27, 29 y 30
PROTEÍNAS	31 al 45	0.697	36, 37, 38, 42 a 45
ENZIMAS	46 al 60	0.645	46 a 49, 53 y 57
LABORATORIO	61 al 70	0.625	64 a 66 y 70

54% de los reactivos analizados tendría semejanza al 100% de ellos. Esa situación no sucedió en este trabajo y por ende es necesario revisar el examen aquí referido. El alfa de Cronbach indica cuales de los reactivos se sugiere eliminarlos o modificarlos para tener un valor más alto y entre más alto sea el valor del alfa, se tendrá mayor confiabilidad. Esto sería conveniente trabajarla para cada uno de los

temas que conforman el examen y ahí ver cuáles de los reactivos se deben eliminar o modificar. Lo más relevante de este trabajo es la factibilidad de hacer uso de técnicas ya validadas que existen en la literatura para así incrementar la confianza en los exámenes de Bioquímica y Biología Molecular que se aplican en la Facultad de Medicina de la UNAM.



REFERENCIAS

- Jaap S, Cees G, Sall MT (2003) Educational evaluation, assessment, and monitoring: a systemic approach. The Netherlands, Lisse, Swets and Zeitlinger Publishers
- González PM (2000) Evaluación del aprendizaje en la enseñanza universitaria. Revista Pedagógica Universitaria Vol. 5(2):31-55
- Días Rojas PA y Leyva SE (2013) Metodología para determinar la calidad de los instrumentos de evaluación. Educación Médica Superior. Vol. 27(2):269-286
- Backhoff E, Larrazolo N, Rosas M (2000) Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). Revista Electrónica de Investigación Educativa 2 (1). Consultado el 25 de septiembre de 2014. Disponible en. <http://redie.uabc.mx/vol2no1/contenido-backhoff.html>
- Méndez RI, Namihira GD, Moreno AL, Sosa de MC (1990) El protocolo de investigación. Lineamientos para su elaboración y análisis. Editorial Trillas. México
- Camacho RJ (2003) Estadística con SPSS para Windows, versión 11. Editorial Alfaomega Ra-Ma.