

Estudio de la confiabilidad y validez del examen diagnóstico de Química del XI curso premédico de la Escuela Latinoamericana de Medicina

Escuela Latinoamericana de Medicina

Ing. Yamira Gui-Sing Mendoza, Lic. Abilio A Valdés de la Cruz, MsC. Mónica Valdés Naranjo, Lic. Greter Zangróniz Sánchez, Lic. Daylén Sosa Herrero

E-mail: yamira@elacm.sld.cu.

RESUMEN

El presente trabajo tuvo como objetivo perfeccionar las pruebas diagnóstico de Química y validar la calidad de las aplicadas en el XI Curso Premédico. Como métodos se utilizaron el histórico-lógico, análisis documental, enfoque sistémico y métodos estadísticos. Se elaboraron dos nuevos diagnósticos, Baterías A y B, que de acuerdo al análisis estadístico realizado tiene mayor calidad que los empleados anteriormente. Se obtuvieron pruebas diagnósticas de mayor confiabilidad y validez, teniendo en cuenta las nuevas formas de enseñanza y los cambios realizados al Programa de la asignatura que permiten conocer las deficiencias y conocimientos de los estudiantes en los temas de la disciplina.

Palabras clave: Confiabilidad, validez de las pruebas, evaluación

INTRODUCCIÓN

En el curso 2002-2003 el examen diagnóstico (Baterías A y B) de Química que se venía aplicando en la asignatura fue perfeccionado y sometido a un estudio de confiabilidad y validez. Como resultado de ese trabajo se elaboró un instrumento evaluativo con las características que se reportan para este tipo de exámenes y se diseñó una nueva hoja de respuesta con las facilidades necesarias para las pruebas diagnósticas (1).

En el curso 2007-2008, en el departamento de Química, se introdujeron nuevas formas de organización de la enseñanza, se rediseñó el programa analítico de la asignatura y se propusieron nuevos objetivos generales para una mayor correspondencia con las formas empleadas en las Ciencias Básicas y buscando mejorar la interdisciplinariedad entre las asignaturas del propio Curso Premédico y la Morfofisiología. Los objetivos de este trabajo son perfeccionar el examen diagnóstico de la asignatura y validar su confiabilidad y validez.

La investigación realizada es del tipo exploratoria y constó de dos etapas: la primera, un trabajo metodológico intenso, el cual incluyó una revisión profunda del examen diagnóstico que se venía aplicando en la asignatura, hasta este momento y la segunda, validación del nuevo instrumento evaluativo.

MATERIALES Y MÉTODOS

En la revisión y corrección de los diagnósticos se emplearon los métodos de análisis documental, el histórico lógico en el análisis de los diferentes enfoques teórico-metodológicos para la elaboración de este tipo de pruebas de forma que cumplieran con las características y exigencias de los test objetivos. También se utilizó el enfoque sistémico para establecer las interrelaciones entre todos los temas de la asignatura.

Para validar los nuevos diagnósticos se tuvo en cuenta la confiabilidad y la validez ya que muchos autores consideran estos criterios de suma importancia en estos tipos de valoraciones.

La validación de este examen se realizó en el XI Curso Premédico, 2008-2009, aplicándose a un total de 1344 estudiantes, para ello se determinó la consistencia interna del mismo; la cual está compuesta por los índices de

facilidad (IF) y de discriminación (ID). El índice de dificultad determina en que medida una pregunta del examen es fácil o difícil y varía de 0 a 100. Consideramos que es más conveniente nombrarlo como índice de facilidad debido a que mientras más elevado es el índice más fácil es la pregunta.

El índice de discriminación nos permite una clasificación discriminativa de los estudiantes según su nivel de actuación y varía entre -1 y +1. Una pregunta con índice de discriminación mayor de 0,35 se evalúa de excelente, con este índice entre 0,25-0,34 es buena, de 0,15-0,24 está en el límite para revisar y menor de 0,15 es mala (2). Para el procesamiento de los datos se utilizó la aplicación Excel de la familia Office para Windows XP. En una hoja de cálculo del libro de Excel se creó la matriz de trabajo. Los valores de 0 y 1 son para preguntas incorrectas y correctas respectivamente. Este procedimiento se realizó de la misma forma para ambas baterías.

Para definir el tamaño real de la muestra se utilizó la expresión $N = No/(1 + No/n)$ donde n es el tamaño de la población (3).

Así se determinó cuantos estudiantes se debían de tomar para el estudio, 436 para cada batería, los que se escogieron por muestreo simple aleatorio por exclusión.

A partir de la matriz de trabajo se ordenaron los diagnosticados en orden decreciente de puntuación, para ambas baterías, y se tomó el 27% de los de mayor puntuación denominados fuertes y menor calificación denominados débiles, respectivamente. Esto nos permitió hacer los dos grupos tan grandes y diferentes como fue posible de 117 exámenes cada uno.

Después de seleccionados los grupos fuerte y débil se determinaron los IF e ID, para cada pregunta, según las siguientes fórmulas (donde F es el número de respuestas exactas en el grupo fuerte, D es el número de respuestas exactas en el grupo débil y N es el número total de estudiantes en estos dos grupos):

Índice de dificultad (IF) = $(F+D)/N * 100$

Índice de discriminación (ID) = $2 * (F-D)/N$

También se determinó la confiabilidad a través del coeficiente alfa de Cronbach (α). Este indicador puede tomar valores entre 0 y 1, donde mientras más cercano a uno mayor es la confiabilidad del instrumento. El valor mínimo aceptado es 0,7.

Para el cálculo de este coeficiente se realizó la matriz de correlación entre ítems utilizando el programa estadístico SPS ($\alpha = N*pr / (1 + pr(N+1))$ donde N es el número de ítems y Pr es el promedio de correlaciones entre ítems). Luego de tener las correlaciones entre ítems se calculó el promedio de estas incluyendo el coeficiente de correlación una vez y excluyendo los coeficientes entre las mismas puntuaciones (4).

RESULTADOS

Corrección de los diagnósticos

Después de una exhaustiva revisión se detectaron las siguientes deficiencias en los exámenes que desde el curso 2002-2003 se estaban aplicando:

- En la Batería A los ítem 1.2 y 1.3, 1.4 y 1.5, 2c) y 2e), 3a) y 3b) responden al mismo objetivo relacionado con la estructura de las sustancias, clasificación de las disoluciones, factores que afectan la velocidad de la reacción y características del equilibrio químico respectivamente y el ítem 2c) está formulado de forma negativa.
- En la batería B las preguntas coinciden con las de la batería A, pero en un orden diferente (2).

Se confeccionaron dos nuevos exámenes diagnósticos (Baterías A y B) que mantiene la misma estructura que el anterior. Con los cambios realizados evaluamos los objetivos generales de la asignatura que debe dominar un estudiante para ingresar a la Educación Superior, lo que nos permite recoger, después de aplicado, la información por temas de las deficiencias y conocimientos de los estudiantes.

Validación de los nuevos diagnósticos

Según los resultados obtenidos para IF, ID y el estadígrafo alfa de Cronbach los nuevos diagnósticos, atendiendo a la metodología empleada, se evalúan de excelente.

Los resultados obtenidos para IF arrojan que en ambas baterías existe una pregunta difícil y una muy fácil y para ID que no hay preguntas malas ni para revisar en ninguna de las 2 baterías.

El estadígrafo alfa de Cronbach dio como resultado que ambas baterías tienen una alta confiabilidad (0,83 y 0,84). Se puede utilizar para cualquier cuestionario aplicado como mínimo a una cantidad de individuos cinco veces mayor que el número de ítems con el fin de evitar correlaciones entre ítems extremadamente altas (4).

DISCUSIÓN

Los exámenes diagnósticos propuestos en relación con los anteriores tienen en común que mantienen la misma estructura, pero se diferencian en que poseen mayor relación con los objetivos del Programa de Química y tienen un total de 27 respuestas posibles en vez de 25.

- En la Batería A se sustituyeron en la pregunta 1 los ítems que repetían objetivos por otros relacionados con las propiedades de las sustancias y con los grupos funcionales característicos de las funciones orgánicas. Se

incluyeron, en la pregunta 2 la espontaneidad de las reacciones químicas y en la 3 un ítem relacionado con la extensión de la reacción.

- En la Batería B, en la pregunta 1 se sustituye dos ítems por otros que evalúan las funciones químicas orgánicas y las formas de expresar la concentración. En las preguntas 2 y 3 se sustituyen las ecuaciones químicas por otras de manera que exista correspondencia con las de la Batería A y se incluyen las modificaciones propuestas para esa Batería.

De esta forma se obtienen dos temarios independientes en los que se evalúan los conocimientos que estos estudiantes deben poseer para ingresar a las Ciencias Básicas, es decir, objetivos de Pre-Universitario vencidos (de acuerdo con el Programa Cubano) (1).

En cuanto al índice de facilidad, excepto para dos ítems, se encontraban entre 30 % y 70 %. El IF medio para la Batería A es de 46,53% y en la B de 47,43%, para una prueba deben confeccionarse un conjunto de preguntas que tengan índices repartidos entre el 30% y el 70 % por lo que el índice medio se situaría entre 40% y 60% (6). Estos resultados no difieren del IF medio del diagnóstico anterior, 50%, por lo que no existen variaciones en cuanto al grado de dificultad de los diagnósticos nuevos con los viejos.

Los resultados del ID arrojan que hay 25 preguntas evaluadas de excelente y 2 de bien en ambas baterías, resultados superiores en relación con los diagnósticos anteriores en los que habían, en cada batería, 2 preguntas para revisar o cambiar.

De estos resultados comprobamos que los estudiantes del grupo débil no tuvieron ventaja sobre los del grupo fuerte. Los niveles medios de competencia de los estudiantes se corresponden con el grupo en que se encuentran ubicados (4).

El coeficiente alfa de Cronbach fue de 0,83 y 0,84 para la Batería A y B respectivamente resultados superiores a los del año anterior (0,78 y 0,79) lo que demuestra que los nuevos instrumentos evaluativos son más confiables.

CONCLUSIONES

Este examen diagnóstico (Batería A y B) constituye un instrumento que sirve en primer lugar para medir el cumplimiento de los objetivos generales de la enseñanza media por parte de los estudiantes en la asignatura Química y sobre la base de esto, trazar estrategias que sirvan para remediar las deficiencias encontradas y así lograr una mejor preparación de los estudiantes para su ingreso a las Ciencias Básicas y en segundo lugar un instrumento de elevada calidad dada por su confiabilidad y validez.

REFERENCIAS BIBLIOGRÁFICAS

1. *García Hernández I, Salas Perea R. La construcción de instrumentos evaluativos escritos. Elementos para mejorar su calidad. CENAPEM. 1997.*
2. *Gilbert J. Educational handbook for health personnel. Revised Ed. Geneva. Ed. No. 35/1981.*
3. *Gil-Coya R. Análisis multivariado aplicado a un experiment pedagógico sobre trabajo extradocente en Matemática. Tesis de Maestría. Universidad de la Habana. 1996.*
4. *Gui Sing Y y otros. Validación estadística de la efectividad del examen diagnóstico de Química en el Curso Premédico. ELAM. 2003.*
5. *Colectivo de asignatura. Pruebas diagnósticos. Departamento de Química. Noviembre 2008.*
6. *Gilbert J. Guía metodológica para el personal de salud. OMS. OPS. 1981.*