Descargado el: 8-11-2026 ISSN 1727-897X

INVESTIGACIÓN PEDAGÓGICA

Consideraciones acerca de la valoración de instrumentos evaluativos y la interpretación de resultados.

Some considerations about the assessment of evaluative instruments and the interpretation of their results.

Luis Alberto Corona Martínez[1]

[1] Hospital Universitario "Dr Gustavo Aldereguía Lima", Cienfuegos, Cienfuegos, Cuba, CP: 55100

Cómo citar este artículo:
Corona-Martínez L. Consideraciones acerca de la valoración de instrumentos evaluativos y la interpretación de resultados.. **Medisur** [revista en Internet]. 2007 [citado 2026 Nov 8]; 1(2):[aprox. 3 p.]. Disponible en: https://medisur.sld.cu/index.php/medisur/article/view/18

Resumen

Se analiza la utilidad y conveniencia del empleo de los índices de dificultad y de discriminación en la valoración y elaboración de instrumentos evaluativos. Estos índices han sido diseñados para su aplicación a las llamadas pruebas de criterios relativos, que no son las utilizadas en nuestro contexto, donde son aplicadas las pruebas con referencia a un criterio. Se insiste en la importancia de otras cualidades de los exámenes como son la pertinencia, la validez de contenido, la objetividad y la amplitud. También es cuestionada la valoración de la calidad con que es ejecutada la evaluación, a partir del análisis de la “curva” de distribución de las calificaciones en las distintas categorías de símbolos.

Palabras clave: métodos de estudio de materia médica

Abstract

The usefulness and conveniences of the usage of the difficulty and discrimination indexes in the appraisement and elaboration of evaluative instruments are analysed in this paper. These indexes have been designed to be applied in the so called tests of relative criteria that are not used in our context in which tests with reference to a criteria are applied. The importance of other qualities in tests such ad pertinence, validity of the content, objectivity/ objectiveness , and amplitude are also insisted on and the quality of the assessment starting from a curve of distribution of qualifications in different categories of symbols is a questionable aspect too.

Key words: methods of studying medical content

Aprobado:

Correspondencia: Luis Alberto Corona Martínez. editorial@jagua.cfg.sld.cu

 ISSN 1727-897X

INTRODUCCIÓN

La evaluación del aprendizaje es concebida como la parte del proceso de enseñanza que permite el control y valoración de lo aprendido, así como el proceso que da viabilidad a este propósito.

Es una gestión básica del proceso docente mediante la cual se logra la retroalimentación necesaria para programar las medidas a tomar para alcanzar de forma óptima los objetivos trazados, al permitir detectar deficiencias que pueden ser inherentes tanto al alumno como al profesor o el plan de clases.[1]

La comprobación y evaluación de los conocimientos, que en cierta medida contribuyen a las tareas generales de la enseñanza, tienen su misión principal y específica en el establecimiento del volumen, la profundidad y la solidez con que han sido asimilados los contenidos y la calificación de éstos; al mismo tiempo, contribuyen a despertar en los alumnos el deseo de estudiar mejor y de perfeccionar los conocimientos. Pero también deben promover y medir las transformaciones que en la personalidad de los educandos se han logrado, así como descubrir los motivos de una débil asimilación.[2]

Para lograr lo anterior, en el proceso docente son utilizados una gran variedad de instrumentos de evaluación o *pruebas*, como más frecuentemente se les conoce, a través de los cuales son obtenidos los *resultados* como expresión del cumplimiento de los objetivos.[3] El propósito de este artículo es exponer nuestras consideraciones sobre las formas más generalizadas de valorar la calidad de los instrumentos evaluativos y junto con ello, de juzgar la calidad con que se ha realizado la evaluación, en sentido general.

DESARROLLO

Existen indicadores cuantitativos para evaluar la calidad de las preguntas que conforman un examen; los más utilizados son el *índice de dificultad* y el *índice de discriminación*, los que son calculados mediante fórmulas que tienen en cuenta las calificaciones obtenidas por los alumnos.[4]

El primero permite determinar en qué medida una pregunta es fácil o difícil, y representa el porcentaje de los estudiantes que han respondido correctamente a la pregunta. El segundo permite determinar en qué medida una pregunta es bastante selectiva para distinguir un grupo fuerte de un grupo débil de estudiantes; representa la correspondencia entre el comportamiento del alumno en una pregunta y el comportamiento en el examen como un todo. La pregunta es más discriminativa en la medida en que sea correctamente contestada en los exámenes mejores, y errada en el grupo de los peores.

Es preocupante observar que no pocos profesores y personas vinculadas a la docencia hacen énfasis en estos índices (y en otros que miden confiabilidad), sin tener en cuenta que éstos, muy en boga en otros modelos de enseñanza basados en principios pedagógicos diferentes a los nuestros, donde la evaluación tiene un fin más discriminativo y la calificación un enfoque más cuantitativo, están diseñados precisamente para ser aplicados a las llamadas *pruebas de criterios relativos*, y que no son las utilizadas en nuestro contexto, donde son aplicadas en la certificación de la calidad del aprendizaje, las *pruebas con referencia a un criterio*.

Las pruebas de criterios relativos se fundan en la referencia a la “distribución estadística normal”, es decir, a la curva de distribución de los resultados de todos los estudiantes que han realizado la misma prueba. Mediante estas pruebas siempre la clasificación de los alumnos identificará al menos dos categorías: los que aprueban y los que no aprueban, sin que se tenga la certeza de que los que aprueban sean, además, competentes. Es útil para concursos, competiciones y oposiciones.

Las pruebas de criterios absolutos son aquellas que se fundan en la referencia a un criterio aceptable de actuación consistente en el dominio de un objetivo específico fijado de antemano. Su objetivo es pues, determinar si una persona posee el dominio o maestría de una tarea dada, y no comparar un individuo con otro o con un grupo de individuos. Sólo este tipo de prueba permite **certificar** la competencia de un determinado personal; sólo éstas informan sobre lo que los individuos son o no capaces de hacer.

Las llamadas pruebas objetivas, cuyo uso se ha generalizado de manera extraordinaria en todo el mundo, habitualmente se enmarcan, por su sistema de calificación, dentro de las pruebas de criterios relativos. Hay autores que han propuesto fórmulas para el cálculo del índice de

discriminación en exámenes que no son pruebas objetivas,[5] y han enfatizado incluso, en la importancia de construir instrumentos con preguntas de determinado grado de dificultad y discriminación, pero ¿qué utilidad puede tener una pregunta que logre un alto grado de dificultad si el contenido que mide es realmente irrelevante?. Además, si analizamos cómo se calcula este índice (expresa el porcentaje de alumnos que no contestaron correctamente la pregunta), más que alegrarnos porque una pregunta muestre un índice de dificultad alto debemos preocuparnos por el hecho de que un número elevado de alumnos no haya podido mostrar dominio de un contenido o alcanzar un objetivo.

En cuanto a la utilidad de una pregunta para discriminar entre grupos de alumnos "fuertes" y alumnos "débiles", en la enseñanza basada en objetivos lo verdaderamente relevante es constatar en qué medida los estudiantes logran alcanzar el objetivo de forma óptima, y no en qué medida difieren unos estudiantes de otros, por lo que también es cuestionable la utilidad del índice de discriminación para la elaboración de instrumentos evaluativos, con excepción de aquellos que son confeccionados para ser utilizados en exámenes de oposición, concursos o exámenes de premio.

Por lo tanto, más que pretender confeccionar instrumentos conformados por preguntas que posean altos índices de dificultad o de discriminación, debemos esforzarnos en garantizar otras cualidades como la pertinencia, la validez de contenido, la objetividad y la amplitud, que consideramos son de mayor importancia para lograr un buen instrumento.

Otro aspecto a debatir radica en que, con cierta frecuencia, escuchamos o leemos en algunos informes cómo es valorada la calidad con que la evaluación ha sido realizada teniendo en cuenta la "curva" de distribución de las calificaciones en las distintas categorías, partiendo de supuestos como que, una evaluación de calidad debe corresponderse con una distribución de las calificaciones predeterminada; que a cada categoría de calificación le corresponde un determinado porcentaje de éstas y por tanto, que la magnitud en que la distribución real se aleja de la distribución hipotética le resta "certeza" a la evaluación realizada.

Si bien es cierto que es bastante improbable que todos los estudiantes (o gran parte de ellos) alcancen resultados con el mismo nivel de excelencia, debido fundamentalmente a que ese resultado es en buena medida, el fruto de la intensidad con que intervino cada alumno en el proceso, con que asumió la responsabilidad de su propio aprendizaje, muy relacionado con su dedicación, motivación y sus aptitudes y capacidades en determinado momento (factores que varían de un alumno a otro), también es cierto que un proceso correctamente organizado y ejecutado, con objetivos (y criterios de consecución) bien definidos; que sea creativo o al menos, productivo; que sea participativo y motivador; capaz de atender de forma diferenciada las particularidades individuales del alumnado; que en resumen, sea verdaderamente instructivo, educativo y desarrollador, podría lograr, y a su vez constatar, resultados bastante homogéneos y con un alto grado de consecución.

Ese es el proceso docente-educativo de excelencia al que debemos aspirar y por el cual debemos trabajar; es el proceso que permitirá que la gran mayoría de los alumnos alcancen los objetivos de su formación con un elevado nivel de calidad. Por lo tanto, nos es difícil asumir que, para aceptar que la evaluación haya sido correctamente realizada, necesariamente un grupo de alumnos tiene que alcanzar los objetivos con un nivel regular de desempeño, y que otro grupo menos numeroso, no logre alcanzar dichos objetivos.

Adicionalmente y no menos importante es el siguiente argumento: juzgar la manera en que han sido cumplidos los objetivos, la manera en que los estudiantes han aprendido y por tanto, la calidad con que el proceso docente ha sido ejecutado, a partir del análisis de las calificaciones solamente, sin tener en cuenta la manera en que se miden los resultados, en que se han constatado y definido esas calificaciones (situación desafortunadamente no infrecuente), no sólo es de un valor muy limitado, sino que, por lo engañoso que resulta, puede ser tremendamente peligroso para el propio proceso y para la sociedad en sí misma.

CONSIDERACIONES FINALES

En nuestro sistema de educación médica superior, la evaluación del aprendizaje implica el control y la valoración de conocimientos, habilidades y hábitos, así como los modos de actuación que los estudiantes van adquiriendo a través del proceso docente-educativo, de acuerdo con los objetivos de cada asignatura,

estancia o rotación en particular, y del plan de estudio en general.

La necesidad de una correcta evaluación del aprendizaje, basada en los más relevantes principios pedagógicos y psicológicos, que le permitan cumplir con las funciones a ella asignadas, no admite cuestionamientos. Pero a partir de lo hasta aquí expresado, nos parece injustificado el énfasis de la literatura pedagógica en el empleo de estos y otros índices para la valoración de instrumentos evaluativos, como también algunas formas a partir de las cuales son sacadas inferencias acerca de la calidad con que es ejecutada la evaluación.

El debate sobre estos u otros aspectos en un tema tan importante como es la evaluación del aprendizaje, sin dudas contribuirá al mejoramiento de ésta y, con ello, al perfeccionamiento del proceso docente-educativo que el desarrollo social nos exige.

REFERENCIAS BIBLIOGRÁFICAS

1. González Menéndez R. La evaluación de objetivos educativos. Rev Cubana Educ Med Sup. 1990 ; 4 (1): 42-49.

2. Rodríguez Mendoza A, Neyra Fernández M, Berra Socarrás M. Experiencias sobre el perfeccionamiento del sistema de evaluación en la carrera de Medicina. Rev Cubana Educ Med Sup. 1994 ; 8 (1-2): 55-60.

3. Alvarez de Zayas CM. La escuela en la vida. Didáctica. La Habana: Editorial Pueblo y Educación; 1999.

4. Pereda Rodríguez MA. Aspectos básicos en el diseño y análisis estadístico de los tests de opciones múltiples aplicables en los procesos evaluativos en la enseñanza. Aula Abierta. 1993 ; 61: 221-238.

5. Pérez Díaz AR. Medida del índice de dificultad y de discriminación por pregunta, en exámenes que no son de TEST. Rev Cubana Educ Med Sup. 1990 ; 4 (1): 50-56.