

Análisis estadístico de examen diagnóstico para valorar la capacidad de transferir conocimientos abstractos a sus aplicaciones en diversos contextos



L. L. Alfaro Avena, J. E. Chávez Pierce, J. Estrada Cabral, S. Flores García

*Departamento de Física y Matemáticas, Instituto de Ingeniería y Tecnología,
Universidad Autónoma de Ciudad Juárez, Avenida del Charro 610 Norte.,
CP 32310 Ciudad Juárez, Chih.*

E-mail: lalfaro@uacj.mx

(Recibido el 10 de Junio de 2014, aceptado el 24 de Marzo de 2015)

Resumen

Diversas evaluaciones realizadas a estudiantes del Instituto de Ingeniería y Tecnología de la Universidad Autónoma de Ciudad Juárez (IIT) indican que la mayoría de estos no desarrollan un aprendizaje funcional, pues no logran transferir habilidades y conceptos desde las ciencias básicas hacia su aplicación en ingeniería. Se han estado elaborando estrategias de aprendizaje de corte inductivo para contextualizar los conocimientos de los cursos básicos de física y matemáticas a problemas específicos propios de las carreras de ingeniería que se ofertan. Sin embargo, se requiere un estudio que verifique la eficacia de estas estrategias para propiciar un verdadero aprendizaje significativo. Con ese fin, se ha trabajado en el diseño y valoración de la utilidad de un examen diagnóstico que determine la capacidad de transferir conocimientos abstractos a sus aplicaciones en diversos contextos. El instrumento se aplicó a una muestra de 140 estudiantes del IIT de niveles intermedio y avanzado, y en el presente trabajo se describen sus características y se muestran los resultados obtenidos del análisis estadístico de dichos reactivos. La confiabilidad interna del examen resultó alta, los reactivos en promedio fueron medianamente difíciles para la muestra, mientras que su poder de discriminación resultó regular, y el 87% de los reactivos utilizados tiene una proporción de nivel de dificultad y poder de discriminación aceptable.

Palabras clave: Evaluación diagnóstica, Teoría y técnicas de evaluación, Confiabilidad interna.

Abstract

Several assessments to students of the Institute of Engineering and Technology at the University of Ciudad Juarez (IIT, by its Spanish acronym) indicate that most of them do not develop a functional learning, because they fail to transfer skill and concepts from basic science to its application in engineering. It has been developing learning strategies based on inductive reasoning to contextualize the knowledge of basic courses in physics and mathematics to specific problems of engineering programs that are offered. However, a study is required to verify the efficacy of these strategies to promote a real meaningful learning. To that end, we have worked on the design and evaluation of the usefulness of a diagnostic test that determines the ability to transfer abstract knowledge to their applications in various contexts. This tool was applied to a sample of 140 students from the IIT in intermediate and advanced levels, and this paper describes the characteristics of the instrument and displays the results obtained from the statistical analysis of its items. The internal reliability of the assessment was high, the items were moderately difficult on average for the sample, while its power of discriminations was regular, and almost 90% of the items used have an acceptable ratio of difficulty to discrimination power.

Keywords: Diagnostic evaluation, Theory of testing and techniques, Internal reliability.

PACS: 01.40.G, 01.40.gf

ISSN 1870-9095

I. INTRODUCCIÓN

El currículo tradicional de enseñanza en ingeniería promueve pocas oportunidades de relacionar de manera satisfactoria los cursos de ciencias con las aplicaciones propias de la disciplina [1]. Por lo cual impera la necesidad del diseño de secuencias de aprendizaje de corte inductivo, como proyectos o experimentos, que complementen dicha deficiencia [2]. El propósito de éstas es generar habilidades de razonamiento que permita al estudiante aplicar conceptos matemáticos y físicos a diversos contextos de ingeniería.

Sin embargo, se debe evaluar la utilidad de dichas estrategias. De inicio, se desea tener una visión clara del estado que prevalece en la educación actual y ver su evolución en el tiempo. Por ello se ha diseñado un instrumento de evaluación diagnóstica. Se desea que dicho instrumento sea válido y confiable. Es decir, que mida lo que se desea medir, y que al ser aplicado en circunstancias y con muestras semejantes arroje resultados semejantes.

En el presente documento se reportan el resultado del análisis de confiabilidad del instrumento y las características de los reactivos al ser aplicado a una muestra considerable.

Estos parámetros por sí solos no determinan la utilidad de un examen; sin embargo, es necesario que dicho instrumento, así como los reactivos individuales, cumplan con ciertos parámetros para que los resultados que se obtengan de estos tengan un significado y valor heurísticos.

II. METODOLOGÍA

Se diseñó un examen con 15 pares de reactivos en dos contextos que denominamos el matemático y el físico. Los problemas en el primer contexto fueron dos de funciones, uno de sistema de ecuaciones lineales, uno de razón de cambio, uno de diferencias finitas, cinco de variación, dos de cálculo integral, uno de vectores y dos de estadística. En el contexto físico ocho problemas fueron de dinámica, dos de termodinámica, dos de electromagnetismo, uno de química y dos de situaciones probabilísticas.

La Figura 1 muestra un ejemplo de reactivo en el contexto matemático con su par en contexto físico.

El objetivo principal de la evaluación es observar la habilidad del estudiante de transferir los recursos matemáticos adquiridos en diversos cursos a problemas en el contexto físico. Los reactivos mismos también pretendieron desarrollar habilidades intelectuales como

- La comparación
- El orden
- La solución de contingencias
- El análisis de enunciados
- La selección de argumentos para la solución de problemas.

El examen se aplicó primeramente a un grupo piloto de 32 estudiantes para realizar adecuaciones en su redacción, y posteriormente a una muestra de 140 estudiantes de nivel avanzado de diversas carreras del Instituto de Ingeniería y Tecnología de la Universidad Autónoma de Ciudad Juárez.

A. Confiabilidad de la prueba y análisis de reactivos

Se buscó que la evaluación fuese objetiva y contara con rigor técnico, siendo esta una prueba desconocida por los evaluados, aplicándola en condiciones idénticas y controladas y utilizando técnicas estadísticas rigurosas.

Se le realizó un análisis estadístico para encontrar la confiabilidad interna y las características psicométricas de los reactivos utilizados, como son: el grado de dificultad (p), el poder de discriminación y el criterio de calidad ambos (relación discriminativa RD) [3, 4, 5, 6].

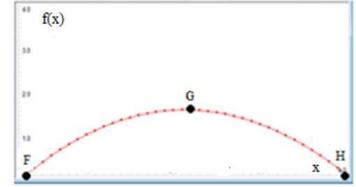
En lo relacionado al poder de discriminación, además del índice de discriminación (D) se utilizó el coeficiente de discriminación biserial. Este último, tiene la ventaja de que se basa en la totalidad de la muestra, en lugar de sólo el 57%.

Este coeficiente se calcula de la siguiente manera [5, 7, 8]

$$r_{pbis} = \frac{\bar{x}_1 - \bar{x}_0}{S_x} \sqrt{\frac{n_1 n_0}{n(n-1)}} \quad (1)$$

donde n_1 es el total de aciertos del reactivo, n_0 el total de errores, $n = n_1 + n_0$, x_1 es la media de las puntuaciones totales de los que contestaron correctamente el reactivo, x_0 la media de los que contestaron incorrectamente, y S_x es la desviación estándar de las puntuaciones totales.

23.- En la figura de la derecha se muestra una gráfica de una función parabólica $f(x)$. Elige la opción que contenga el orden correcto con respecto a la magnitud de la razón de cambio de la función en los puntos F, G y H.

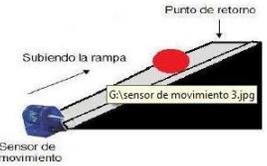


- 1) $F > G > H$
- 2) $G > F > H$
- 3) $F = H > G$
- 4) $F = G = H$

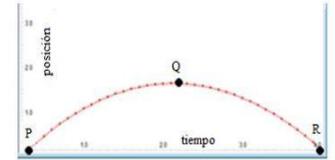
Explica brevemente tu razonamiento.

Nivel de dificultad muy difícil ____ difícil ____ fácil ____ muy fácil ____!

26.- En el laboratorio de física se realizó un experimento, el cual consiste en arrojar una pelota que sube y baja a lo largo de una rampa. Los resultados del experimento se muestran en la siguiente gráfica de posición-tiempo.



A continuación se muestran los argumentos de una conversación entre cuatro estudiantes acerca de la magnitud de la rapidez de cambio de la posición de la pelota.



¿Cuál de los argumentos es correcto?

- a) ESTUDIANTE A: "Puedo decirles que los tiempos correspondientes a los puntos P y R, la magnitud de la rapidez de cambio de la posición es la misma".
- b) ESTUDIANTE B: "Yo estoy seguro que en el tiempo correspondiente al punto Q la rapidez de cambio es diferente de 0".
- c) ESTUDIANTE C: "Pues yo opino que en cualquier punto la rapidez de cambio de la posición es la misma".
- d) ESTUDIANTE D: "Todo lo que ustedes argumentan es correcto".

Explica brevemente tu razonamiento.

FIGURA 1. Par de reactivos 23 y 26 de la evaluación diagnóstica. El primero se encuentra en el contexto matemático, mientras el segundo en físico. Al final se anexa el examen completo.

La Tabla I muestra una regla de "dedo" para determinar la calidad de los reactivos en términos de su poder de discriminación y la acción sugerida.

TABLA I. Calidad de los reactivos en función del índice de discriminación biserial. Se indica acción recomendada.

D_i o r_{pbis}	Calidad	Recomendación
> 0.39	Excelente	Conservar
$0.30 - 0.39$	Buena	Puede mejorarse
$0.20 - 0.29$	Regular	Revisar
$0.00 - 0.19$	Pobre	Revisar exhaustivamente
< 0.00	Pésima	Descartar

Aunque el nivel de dificultad y el poder de discriminación dependen de las características de la muestra, como su

habilidad y trasfondo educativo, el estudio de estos parámetros puede revelar errores en el diseño del reactivo mismo, sobre todo cuando el poder de discriminación es muy bajo, mientras su grado de dificultad es intermedio.

IV. RESULTADOS

Aunque 30 reactivos parecían demasiados para una aplicación de menos de dos horas, la mayoría de los estudiantes respondieron el examen en su totalidad, y sólo un 9% de los reactivos quedaron sin responder. Sin embargo, el promedio de aciertos fue bajo, pues el grado de dificultad medio resultó de 0.36, con una desviación estándar de 0.21.

El coeficiente de confiabilidad interna fue de 0.61, que está en el límite inferior de lo aceptable para pruebas de rendimiento académico [4, 9]. El índice de discriminación promedio fue de 0.29, con una desviación estándar de 0.19, mientras que el coeficiente de discriminación biserial medio resultó ser 0.28, con una desviación estándar de 0.15. La relación discriminativa media fue 3.44, con 2.26 en su desviación estándar.

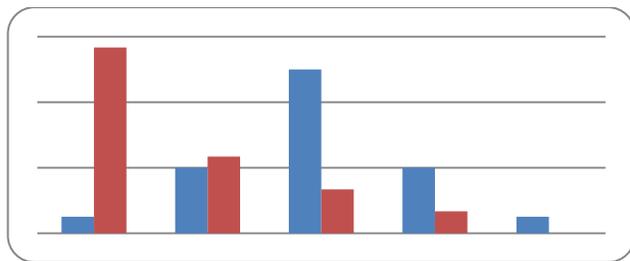


FIGURA 2. Porcentaje de reactivos de un examen equilibrado (azul) y porcentaje de reactivos como función de su grado de dificultad.

La Figura 2 muestra una gráfica de la distribución del porcentaje de reactivos en función de su grado de dificultad (en rojo), y la distribución deseable para una prueba equilibrada (en azul). En esta, como en la tabla II, AD significa “altamente difícil”, MD “medianamente difícil”, DM “dificultad media”, MF “medianamente fácil y AF, “altamente fácil.

En la Tabla II se resume el grado de dificultad obtenido para cada uno de los reactivos. Notamos que predominan los problemas altamente difíciles (casi el 60%), por lo que bajo este criterio, la evaluación no está equilibrada. Sin embargo, esto depende de la validez de los reactivos. Si se espera que el estudiante domine el tema, y el reactivo está bien planteado y redactado, se preserva aunque para la muestra resulte con un alto grado de dificultad.

TABLA II. Clasificación de los reactivos por su grado de dificultad (p) obtenido en la evaluación.

Item	p_i	Clasificación	Item	p_i	Clasificación
1	0.63	DM	27	0.69	DM
2	0.15	AD	30	0.21	AD
3	0.49	MD	6	0.11	AD
4	0.31	AD	18	0.29	AD
5	0.77	MF	16	0.35	MD
11	0.12	AD	28	0.19	AD
12	0.74	DM	14	0.58	DM
13	0.19	AD	7	0.26	AD
17	0.43	MD	29	0.18	AD
19	0.41	MD	9	0.26	AD
20	0.51	MD	15	0.86	MF
21	0.20	AD	25	0.14	AD
22	0.22	AD	8	0.31	AD
24	0.26	AD	10	0.20	AD
26	0.34	MD	23	0.45	MD

La Figura 3 muestra un acumulado de la frecuencia de los reactivos como función de la calidad, tanto para el índice de discriminación (en azul) como para el coeficiente de discriminación biserial (en rojo).

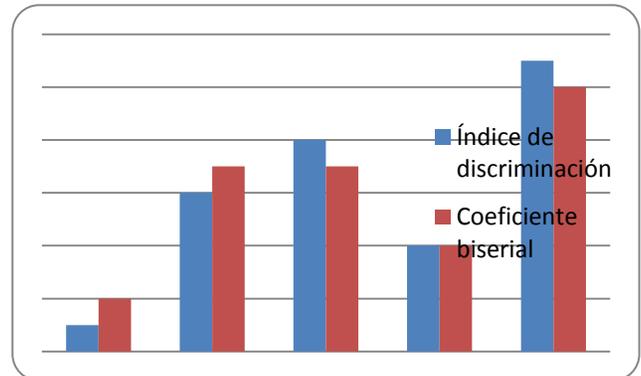


FIGURA 3. Frecuencia de reactivos en función de su índice de discriminación (azul) y su coeficiente de discriminación biserial (rojo).

TABLA III. Calidad de los reactivos por su poder de discriminación (D o r_{pbis}).

Item	D	Calidad	r_{pbis}	Calidad
1	0.42	Excelente	0.36	Buena
2	0.26	Regular	0.29	Regular
3	0.58	Excelente	0.45	Excelente
4	0.03	Pobre	0.03	Pobre
5	0.21	Regular	0.18	Pobre
11	0.05	Pobre	0.16	Pobre
12	0.34	Buena	0.26	Regular
13	0.03	Pobre	-0.01	Pésima
17	0.53	Excelente	0.45	Excelente

19	0.53	Excelente	0.46	Excelente
20	0.32	Buena	0.27	Regular
21	0.21	Regular	0.19	Pobre
22	0.40	Buena	0.46	Excelente
24	0.24	Regular	0.22	Regular
26	0.48	Excelente	0.42	Excelente

TABLA IV. Calidad de los reactivos por su poder de discriminación (D o r_{pbis}).

Item	D	Calidad	r_{pbis}	Calidad
27	0.50	Excelente	0.45	Excelente
30	0.21	Regular	0.21	Regular
6	0.19	Pobre	0.29	Regular
18	0.21	Regular	0.20	Pobre
16	0.40	Buena	0.30	Regular
28	0.24	Regular	0.33	Buena
14	0.03	Pobre	0.06	Pobre
7	-0.11	Pésima	-0.02	Pésima
29	0.45	Excelente	0.48	Excelente
9	0.56	Excelente	0.52	Excelente
15	0.24	Regular	0.30	Buena
25	0.24	Regular	0.32	Buena
8	0.42	Excelente	0.41	Excelente
10	0.11	Pobre	0.06	Pobre
23	0.53	Excelente	0.42	Excelente

Las Tablas III y IV muestran dichos coeficientes para cada reactivo en sus contextos físicos y matemáticos, respectivamente. Se indican los reactivos que tienen necesidad de ser revisados, y en particular el par 13 – 7, descartado.

La Tabla V muestra la relación discriminativa, en la cual sólo los reactivos 4, 5, 14 y 7 resultaron con parámetros no deseables.

TABLA V. Calidad de los reactivos por su poder de discriminación.

Item	RD	Calidad
1	2.12	AA
2	5.71	AA
3	3.73	AA
4	0.31	AR
5	0.75	AR
11	3.14	AA
12	1.55	AA
13	1.03	AA
17	4.67	AA
19	4.68	AA
20	2.25	AA
21	3.33	AA
22	6.45	AA
24	2.96	AA
26	5.00	AA

Item	RD	Calidad
27	2.64	AA
30	4.44	AA
6	5.83	AA
18	2.33	AA
16	3.54	AA
28	5.64	AA
14	0.16	AR
7	-1.48	AR
29	8.00	AA
9	7.57	AA
15	1.60	AA
25	4.67	AA
8	4.96	AA
10	1.43	AA
23	4.23	AA

V. CONCLUSIONES

El instrumento de evaluación resultó con un coeficiente de confiabilidad alto, apenas apropiado para una evaluación de desempeño académico. El grado de dificultad promedio de los reactivos fue medianamente difícil, lo cual no es necesariamente negativo, pero es deseable disminuir el grado de dificultad para poder contar con un mayor número de respuestas, con sus respectivas justificaciones, que permitan hacer un mejor análisis de la transferencia de conocimientos y habilidades de los estudiantes en diversos contextos.

El nivel de discriminación basado en los dos parámetros utilizados en este trabajo resultó congruente para la mayoría de los reactivos, con 6 centésimas de variación en promedio, que modifica en poco la acción a realizar recomendada por dicho criterio. A este respecto, se recomienda conservar 10 reactivos, mejorar 3, revisar 16 y descartar 1, el número 7.

Cuando se consideran los parámetros de dificultad y discriminación, la relación discriminativa resultante sugiere la revisión exhaustiva de 4 reactivos (4, 5, 14 y 7).

En resumen, el instrumento arrojó buenos parámetros para la mayoría de los reactivos utilizados, aunque algunos requieren de mejora. De manera que las conclusiones que se obtengan de la aplicación de este instrumento pueden considerarse válidas.

REFERENCIAS

- [1] Terenzini, P. T., Cabrera, A. F., Colbeck, C. L., Parente, J. M. & Bjorklund, S. A., *Collaborative Learning vs Lecture/Discussion: Students' Reported Learning Gains*, Journal of Engineering Education **90**, 123-130 (2006).
- [2] Ávila, M. S., Luna, J., López, C. & Flores, S., *La experimentación con situaciones físicas como generadoras de significados previos al cálculo*, IV Encuentro Internacional sobre la Enseñanza del Cálculo, Puebla, México (2010).
- [3] Alfaro, L. L., Chávez, J. E. & Salazar, M. C., *Evaluación diagnóstica en Matemáticas básicas a estudiantes de primer ingreso del Instituto de Ingeniería y Tecnología de la UACJ*, IV Congreso de investigación CIPITECH, Nuevo Casas Grandes (2011).
- [4] Ruiz, C., *Confiabilidad*, <<http://www.carlosruizbolivar.com/articulos/archivos/Curso%20CII%20%20UCLA%20Art.%20Confiabilidad.pdf>>, Consultado el 6 de noviembre de 2010.
- [5] Backhoff, E., Larrazolo, N. & Rosas, M., *Nivel de dificultad y poder de discriminación del examen de habilidades y conocimientos básicos (EXHCOBA)*, REDIE Revista Electrónica de Investigación Educativa **2**, 1-16 (2000).
- [6] González, J., *Evaluación de opción múltiple vs evaluación tradicional. Un estudio de caso en ingeniería*, Ingeniería **7**, 17-37 (2003).
- [7] Frisbie, D. A., *Reliability of scores from teacher-made tests*, Educational Measurement: Issues and Practice **7**, 25-35 (1988).

[8] Ebel, R. & Frisbie, D., *Essentials of Educational Measurement*. (Prentice-Halls, Englewood Cliffs, 1991).

[9] Magnuson, D., *Teoría de los tests*, (Trillas, México, 1982).