



Innovation in research and engineering education:  
key factors for global competitiveness

*Innovación en investigación y educación en ingeniería:  
factores claves para la competitividad global*

# LA EVALUACIÓN OBJETIVA EN INGENIERÍA APORTES EN PROCESOS DE EVALUACIÓN Y MEJORA CURRICULAR

**Olga Rosalba Rodríguez Jiménez**

**Universidad Nacional de Colombia  
Bogotá, Colombia**

## **Resumen**

El objetivo del trabajo es presentar el análisis psicométrico del examen de ciencias básicas para ingeniería – EXIM- realizado por ACOFI y que se aplica desde el año 2007, en distintas universidades y programas de ingeniería. Para hacerlo se estudiaron las calidades psicométricas de la prueba a partir de la Teoría Clásica de los Test y el modelo de Rasch, además de estudiar la fiabilidad y las evidencias de validez. Los resultados señalan adecuados índices psicométricos para los ítems, niveles moderados de fiabilidad y adecuada validez relacionada con el contenido. Los resultados permiten concluir sobre la calidad de la prueba y hacer sugerencias para su mejora, y además presentar reflexiones sobre el uso de la evaluación objetiva en ingeniería y su aporte en los procesos de adecuación y mejora curricular.

**Palabras clave:** evaluación en ingeniería; psicometría; EXIM

## ***Abstract***

*The main purpose is show the psychometric analysis of test EXIM applied for ACOFI from 2007 in different programs and universities of engineering of Colombia. With this purpose we analyzed the psychometric qualities of test with the Classic Test Theory and Rasch Model, in addition we studied the fiability and validity. The results showed good psychometric index for the items, medium level of fiability and good validity of the content. With this information it is possible to conclude on the quality of test and make suggestion for its improvement, in addition it make reflexions about the use of test in engineering assessment and its profit in improving the curriculum.*

**Keywords:** *assessment in engineering; psychometric; EXIM*

## 1. Introducción

El currículo, la didáctica y la evaluación son tres pilares que orientan el proceso de enseñanza-aprendizaje en todos los ciclos escolares; el primero ha cobrado mayor atención en educación superior en consideración a los objetivos que le atañen, aunque se reconocen algunas aproximaciones didácticas, como el uso de aprendizaje cooperativo o el aprendizaje con base en proyectos, estrategias que han tenido un buen desarrollo e implementación en carreras como la ingeniería y que demandan nuevas alternativas de evaluación.

Rodríguez, Casas y Medina (2005) señalan que la evaluación de los procesos y programas educativos cuenta con varios propósitos como el monitoreo, la selección, la certificación y el diagnóstico. En función de estos propósitos Capper (1996) ha establecido la diferencia entre evaluaciones nacionales y exámenes públicos; el autor los diferencia en relación con el currículo, con el número de estudiantes que participan o la periodicidad de la aplicación; de este modo caracteriza las evaluaciones nacionales como aquella cuyo propósito es el monitoreo y la evaluación del sistema educativo, trasciende el currículo y se aplica a una muestra de la población con una periodicidad variable. Por su parte los exámenes nacionales, se basan en el currículo particular, se aplican una o dos veces al año a solicitud de los usuarios interesados. Acorde con estas consideraciones el examen de ciencias básicas de ingeniería –EXIM- es un examen nacional cuyos resultados pueden ser un indicador sobre la calidad de la educación en ciencias básicas en ingeniería y responden a la evaluación denominada externa, según la cual permite *“evaluar con rigor y de manera sistemática lo que se hace en el sistema educativo”* (Gago, 2000, p. 108)

En este orden de ideas se asume que la evaluación surge como un mecanismo para el cambio y el aprendizaje estratégico (Pozo y Monereo, 1999) en tanto permite conocer el nivel de competencia alcanzado, en comparación con el esperado (Graaff y Rompelman, 2004) y una herramienta fundamental en la búsqueda cada vez más reciente en los programas de ingeniería de los currículos integrados (Jeffrey y Matthew, 2005).

El currículo en ingeniería tiene un alto componente de fundamentación en matemáticas y ciencias y en la aplicación de este conocimiento a problemas prácticos de ingeniería, tales como el diseño. Y dado que en el currículo se hacen explícitos los propósitos, principios, métodos, estrategias, medios y mediaciones (Pozo y Monedero, 1999) del proceso formativo, en concordancia con la sociedad de la información, el currículo debe orientarse a que el estudiante siga aprendiendo (Pozo y Monereo, 1999), lo que implica que en los currículos se explicita el nivel de conocimientos y de desarrollo de competencias que se esperan lograr en cada semestre o año de formación y la forma de poder obtener información sobre el logro de este propósito es a través de la evaluación.

Ahora bien existen muchas herramientas de evaluación cuyo uso resulta más pertinente según se use en el aula de clase o en aplicaciones masivas, así las pruebas o test objetivos son los más usados dado que son objetivos, de fácil y rápida calificación y permiten evaluación de su calidad (Herrera, 1998).

En la evaluación de la calidad de un test, desde la psicometría se cuenta con dos aproximaciones: la Teoría Clásica de los Test –TCT- y la Teoría de Respuesta al Ítem –TRI-. La TCT analiza los ítems y la prueba en su totalidad. De los ítems se evalúan las características denominadas dificultad y discriminación y de prueba la fiabilidad y la validez, aunque esta última en realidad no es un criterio que pertenezca a ninguna teoría. El índice de dificultad corresponde a la proporción de personas que resuelven correctamente un ítem y el índice de discriminación señala si el ítem tiene alto poder para diferenciar entre las personas que muestran mayor y menor magnitud del atributo evaluado y se expresa como la relación entre las puntuaciones

obtenidas en el ítem con las obtenidas en el test. Uno de los indicadores que se emplean es el Brodgen Clemens (BrCI), el cual estima la correlación máxima entre los puntajes en el ítem y en la prueba, teniendo en cuenta la proporción de personas que contestó correctamente la pregunta (Herrera, 1998). Para estos dos índices existen parámetros de aceptación, en el caso de la dificultad cada investigador o usuario puede fijar el rango de aceptación de acuerdo con los datos obtenidos y los fines de la prueba, aunque generalmente se encuentran entre 0,1 y 0,9 y en el caso de la discriminación se considera valores cercanos o superiores a 0,3.

En relación con la fiabilidad el índice más empleado es el de consistencia interna Alfa de Cronbach. Por su parte en relación con la validez, acorde con Paz (1996) el principal interés en pruebas educativas, es determinar el grado en que los ítems representan el “contenido” disciplinar que se quiere medir, lo cual se logra con juicio de expertos, un proceso que implica la elaboración, revisión y discusión de la estructura de prueba y de los ítems que la componen por parte de expertos en el área.

De forma complementaria, se cuenta con la Teoría de Respuesta al Ítem y dentro de esta, con el Modelo Rasch, acorde con el cual la probabilidad de respuesta en el ítem es función de la diferencia entre la medida del rasgo de una persona y la dificultad del ítem, lo que se representa en la curva característica del ítem (CCI) (Tristán, 1998) y se expresa como una función logística

$P(H) = \frac{1}{1+e^{-(H-b)}}$ . En donde  $H$  es la magnitud de habilidad del evaluado y  $b$  es la dificultad del ítem. El único parámetro del ítem es la dificultad y corresponde al punto donde la probabilidad de acierto es de 0.5. Para determinar la calidad de la estimación y en ese sentido contar con una medida de ajuste, el modelo tiene como referencia dos medidas OUTFIT e INFIT. En el primer caso se trata de una medida sensible al comportamiento inesperado que afecta a los ítems que presentan un nivel de dificultad lejano del nivel de habilidad de una persona y en el segundo caso, la sensibilidad en relación con la cercanía del nivel de habilidad de la persona (Tristan, 1998). Los valores de ajuste de los ítems deben encontrarse entre 0.5 y 1.5 (Linacre, 2003).

En el marco de lo presentado, el trabajo muestra el estudio psicométrico del examen EXIM 2012, el cual tiene propósito comprobar el proceso de formación en ciencias básicas y comprobar el grado de desarrollo de las competencias “académicas” de los estudiantes que cursan sexto semestre de ingeniería y con ello ser un indicador objetivo que contribuya a la autoevaluación de las instituciones de educación superior Colombianas, en el marco de los procesos de acreditación que se adelanta en el país desde mediados de la década de los noventa (ACOFI, 2007). Se espera a partir de este análisis plantear sugerencias para su mejora y reflexionar sobre los aportes en procesos curriculares.

## 2. Método

Este estudio se considera como instrumental acorde con la clasificación presentada por Montero y León (2007).

### 2.1. Muestra

Los datos corresponden a 1.389 estudiantes de 22 Instituciones universitarias Colombianas de diversos programas de ingeniería: industrial, eléctrica, sistemas, mecánica, biomédica, industrial, entre otras. La información analizada corresponde al año 2012.

## 2.2. Variables

La variable analizada fue el puntaje en la prueba EXIM. El examen consta de 140 preguntas, de las cuales 45 corresponden a Matemáticas, 35 a Física y 30 para Química y Biología. Todas son preguntas de selección múltiple con única respuesta, las cuales se responden a partir de un texto o gráfica. La prueba se estructura para cada área a partir de tres competencia académicas: capacidad de abstracción, análisis y síntesis, aplicación de conocimientos en la práctica e identificación, planeación y resolución de problemas y de dominios o ejes conceptuales que buscan dar cuenta de los aspectos centrales en los que debe formarse un ingeniero, independiente de la institución en la cual se forme.

Estos dominios y competencias fueron definidos en el marco de los lineamientos del Ministerio de Educación de Colombia, la Accreditation Board for Engineering and Technology (ABET) y el proyecto Tuning para América Latina.

## 2.3. Procedimiento

La estructura de prueba y las preguntas fueron construidas por expertos en el tema y posteriormente revisadas y discutidas con otros expertos en cada área y en psicometría. En un segundo momento todas las preguntas y la prueba a aplicar fue revisada por expertos en cada área y en psicometría, ajustando o cambiando las preguntas según cumpliera o no con los objetivos de la evaluación y con los criterios de calidad de las preguntas y la prueba.

La aplicación de la prueba se realizó por personas con entrenamiento previo para la aplicación de este tipo de pruebas, para quienes se diseñó un manual en el cual se presentan las instrucciones antes, durante y después del examen y se acompaña de una lista de chequeo que permita verificar el cumplimiento de cada una de las acciones señaladas. El tiempo de aplicación de la prueba fue de cuatro horas.

## 2.4. Análisis de datos

Toda la información fue capturada por lectora óptica. Se verificaron inconsistencias, como estudiantes que presentaron el examen y aparecían sin respuestas o divergencias en el número de identificación. Para el análisis referido a la Teoría Clásica de los Test (TCT) se empleó el aplicativo en Microsoft Office 2003 - Excel 11.0 desarrollado por Casas (2005) y para el análisis con el modelo de Rasch el programa WINSTEPS.

## 3. Resultados

### Índices de los ítems

En la tabla 1 se encuentran los descriptivos para el índice de dificultad de los ítems. Se aprecia que Matemáticas es el área que cuenta con los ítems extremos, el más fácil y el más difícil. La prueba de Biología es la que cuenta con ítems más cercanos a valores medios de dificultad y con una baja desviación típica. Se presenta en la misma tabla 1 la información del índice de discriminación, se observa promedios adecuados para las áreas de Física, Química y Biología y apenas aceptable para el área de Matemáticas, tanto en Física como en Matemáticas llama la atención la existencia de ítems con discriminación negativa.

Tabla 1. Descriptivos de los ítems acorde con la Teoría Clásica de los Test y Fiabilidad por áreas

Área	Índice	Mínimo	Máximo	Media	Desv. tip	Coefficiente $\alpha$	Nº de ítems
Matemáticas	Dificultad	0,04	0,79	0,33	0,16	0,53	43
	Discriminación	-0,22	0,65	0,26	0,22		
Física	Dificultad	0,1	0,53	0,3	0,1	0,55	33
	Discriminación	-0,06	0,63	0,32	0,18		
Química	Dificultad	0,16	0,49	0,3	0,08	0,56	29
	Discriminación	0,1	0,61	0,35	0,11		
Biología	Dificultad	0,22	0,61	0,4	0,1	0,79	28
	Discriminación	0,25	0,75	0,47	0,15		

### Análisis de consistencia interna por área

La fiabilidad del test determinada con el coeficiente  $\alpha$  de Cronbach se encuentra entre 0,53 y 0,79 (tabla 1). Se aprecia que la fiabilidad es mayor para el área de Biología. Estos índices señalan niveles moderados de fiabilidad, aceptables para pruebas educativas.

### Análisis de contenido

Los expertos revisaron las preguntas incluidas en la prueba analizada y consideraron que se correspondía con la estructura de prueba diseñada para el examen, reportaron que todas las preguntas eran claras, pertinentes y relevantes y contaban con respuesta correcta.

### Análisis con IRT

Se consideró aceptable la unidimensionalidad de la prueba por área y a partir de ello se obtuvo la calibración de los ítems con el modelo de Rasch, los resultados se presentan en la tabla 2.

Se aprecia que se tiene una buena calibración de los ítems, en las cuatro áreas se tiene el promedio esperado de habilidad y un bajo error de medición. Acorde con los indicadores de ajuste, este modelo ajusta adecuadamente a los datos, con excepción de dos ítems, uno en el área de Matemáticas y uno en el área de Física, los cuales presentan valores superiores a 1,5.

El área que cubre un mayor rango de habilidad corresponde a Matemáticas, seguida por Física, lo cual permite contar con una prueba posible de abordar para todos los estudiantes, quienes tienen bajos y altos niveles de competencia.

Tabla 2. Descriptivos de los ítems acorde con el Modelo de Rasch

	<b>Matemáticas</b>				<b>Física</b>			
	Habilidad	Error de medición	Infit	Outfit	Habilidad	Error de medición	Infit	Outfit
Media	0,00	0,06	1,00	1,04	0,00	0,06	1,00	1,02
Desviación	0,84	0,01	0,06	0,14	0,51	0,01	0,07	0,11
Mínimo	2,51	0,14	1,13	1,53	1,35	0,09	1,13	1,31
Máximo	-2,27	0,06	0,87	0,82	-1,07	0,06	0,87	0,85
	<b>Química</b>				<b>Biología</b>			
	Habilidad	Error de medición	Infit	Outfit	Habilidad	Error de medición	Infit	Outfit
Media	0,00	0,06	1,00	1,01	0,00	0,06	1,00	1,05
Desviación	0,38	0,00	0,05	0,10	0,51	0,00	0,12	0,25
Mínimo	0,84	0,08	1,14	1,37	0,87	0,07	1,34	2,05
Máximo	-0,92	0,06	0,87	0,85	-1,10	0,06	0,78	0,74

#### 4. Conclusiones

El análisis realizado permite conocer las calidades psicométricas del EXIM, de esta manera es posible afirmar que la prueba cuenta con unidimensional aceptable, moderados índices de fiabilidad, adecuada validez de contenido e ítems de distintos niveles de dificultad con predominio de dificultad media, se encuentra en general para todas las áreas ítems discriminativos.

En cuanto a la calidad de la prueba se sugiere revisar con los constructores las preguntas que tiene índices negativos de discriminación en conjunto con el análisis de las respuestas por opción, se buscará también ampliar el rango de medición, incluyendo preguntas de niveles de dificultad extremos, particularmente bajos. También será de utilidad esta revisión para evaluar la fiabilidad de la prueba, así al mejorar los ítems que componen la prueba, se podría obtener un mayor índice de fiabilidad.

En relación con la validez, dado que hasta el momento sólo se tiene evidencias de validez relacionadas con el contenido se sugiere emplear criterios externos que permitan asociar el desempeño el EXIM con el desempeño en la universidad y en las pruebas SABER PRO.

En relación con el impacto del EXIM en los currículos de las universidades, se sugiere realizar un análisis de los resultados por parte de cada universidad participante, de manera tal que se visibilicen los dominios conceptuales y las competencias que deben potencializarse (Ravela, Wolfe, Valverdey Esquivel, 2001) y se determine la relación entre lo evaluado y lo enseñado. Se espera que este análisis contribuya a retroalimentar el diseño del instrumento en tanto permite valorar la pertinencia de los dominios y competencias definidos y la articulación de estos con el currículo de cada universidad y la relación que guarda con los exámenes nacionales SABER PRO.

Finalmente, es posible afirmar que el EXIM es una posibilidad evaluativa que permite contribuir a la sólida formación en ciencias básicas y a la integración curricular necesaria para los programas de ingeniería

(Jeffrey y Matthew, 2005). Lo anterior implica que este tipo de evaluaciones se emplee de manera complementaria a los ejercicios y trabajos escritos, las presentaciones orales, el desarrollo de proyectos y los portafolios (Escudero, 2001).

## 5. Referencias bibliográficas

- ACOFI. (2007). Manual de prueba. Bogotá (Documento Inédito). ACOFI.
- Capper, J. (1996). Testing to Learn – Learning to Test. USA: International Reading Association.
- Casas, P. (2005). Aplicativo en excel. (Software de uso restringido).
- Escudero, T. (2001). La evaluación del rendimiento de los estudiantes como actividad docente. Programa de formación para la docencia universitaria. Documento de trabajo. Universidad de Oviedo.
- Gago, A. (2000). El CENEVAL y la evaluación externa de la educación en México. Revista Electrónica de Investigación Educativa, 2 (2). Consultado en marzo de 2013 en el World Wide Web: <http://redie.ens.uabc.mx/vol2no2/contenido-gago.html>
- Graaff, E y Rompelman, O (2004). Assessment of learning results. European Journal of Engineering Education, Vol. 29, N° 2, pp. 171–172.
- Herrera, A. N. (1998). Notas de Psicometría. Documento Inédito. Bogotá: Universidad Nacional de Colombia.
- Jeffrey, F. y Matthew, O. (2005). Integrated Engineering Curricula. Journal of Engineering Education, Vol. 94, N°1, pp. 147-164.
- Linacre, M. (2003). A user's guide to Winsteps. Chicago: Mesa Press.
- Montero y León (2007). A guide for naming research studies in Psychology. International Journal of Clinical and Health Psychology. Vol. 7, No. 3, pp. 847-862.
- Paz, M. (1996). Validez. En Muñiz, J (Ed.) Psicometría. Madrid. Universitas.
- Pozo, I y Monereo, C (2001). El aprendizaje estratégico. Alfabeta Ediciones. Madrid. pp.407.
- Ravela, P., Wolfe, R., Valverde, G. y Esquivel, J. (2001). Los próximos pasos. ¿Cómo avanzar en la evaluación de aprendizajes en América Latina?. Programa de Promoción de la reforma educativa en América Latina y el Caribe PREAL. No. 20. Chile - Santiago.
- Rodríguez, O., Casas, P. y Medina, J. (2005). Análisis psicométrico de los exámenes de evaluación de la calidad de la educación superior (ecaes) en Colombia. Avances en Medición, Vol 3, pp. 153-172.
- Tristan, A. (1998). Análisis de Rasch para todos. México: CENEVAL.

### Sobre la autora

- **Olga Rosalba Rodríguez Jiménez:** Psicóloga. Master en educación y en metodología en ciencias del comportamiento y la salud. Doctora en psicología y educación de la Universidad Autónoma de Madrid. Docente de la Universidad Nacional de Colombia y asesora psicométrica de ACOFI. [orrodriquezj@unal.edu.co](mailto:orrodriquezj@unal.edu.co)

Los puntos de vista expresados en este artículo no reflejan necesariamente la opinión de la Asociación Colombiana de Facultades de Ingeniería y de la International Federation of Engineering Education Societies

Copyright © 2013 Asociación Colombiana de Facultades de Ingeniería (ACOFI), International Federation of Engineering Education Societies (IFEES)