



Innovation in research and engineering education:  
key factors for global competitiveness  
*Innovación en investigación y educación en ingeniería:  
factores claves para la competitividad global*

# MEDICIÓN DE LOS PARÁMETROS IRT DE UNA TAREA DINÁMICA EN LA ASIGNATURA *ESTÁTICA*

Jorge Luis Restrepo Ochoa, Jaime Leonardo Barbosa Pérez, Andrés Restrepo Cadavid

Universidad EAFIT  
Medellín, Colombia

## Resumen

*Estática* es una asignatura básica en la formación de los ingenieros, que presenta continuas dificultades para la comprensión por parte de los estudiantes; por tal motivo, la Escuela de Ingeniería está desarrollando un sistema informático de evaluación en ambiente web, que genera diferentes versiones de una misma tarea para cada estudiante almacenando los resultados para su posterior análisis y valoración. El objetivo final de este desarrollo es mejorar los procesos de enseñanza y aprendizaje. El presente artículo muestra los resultados del análisis realizado a los parámetros de dificultad y discriminación, de los ítems de una evaluación generada utilizando esta herramienta en la asignatura *Estática*, de la escuela de ingeniería de la Universidad EAFIT, en el primer semestre del año 2013. Los ítems de la prueba se han seleccionado de forma que se incluyan tareas en las que se varíen los parámetros en forma controlada. El análisis de este trabajo, se basa en las pruebas aplicadas a un grupo de estudiantes. Cada uno de los ítems utilizados en esta prueba se caracteriza utilizando la *teoría de respuesta al ítem* (IRT, Item Response Theory por sus siglas en inglés). Dicha experiencia ha permitido establecer los diferentes niveles de dificultad y discriminación de las preguntas y su relación con las variables de la tarea. Finalmente este trabajo nos muestra los criterios con los que se deben programar las tareas para controlar la generación automática de pruebas que permitan evaluar estudiantes con diferentes niveles de habilidad en relación al aspecto evaluado y así mejorar los procesos de evaluación del curso.

**Palabras clave:** estática; evaluación del aprendizaje; sistemas de evaluación; IRT

## Abstract

*Statics is a basic subject which presents continuously difficulties for its understanding by the students in their formation as engineers, for this reason, the engineering department is developing an evaluation computer system based on web environment that generates different version of one task for each student, storing the final results for their post-analysis. The final goal of this development is to improve the teaching and learning processes. This article expose the analysis results made to the items parameters (difficulty and discrimination) of an evaluation generated by the computer system in the Statics subject, from the engineering department of*

*the EAFIT University, during the first half of 2013. The test items are selected in a controlled way which is to include tasks and assign varied parameters to each one of them. The analysis of this paper is based in the tests applied to a group of students. Each items of this test is characterized applying the item response theory (IRT). This experience has established different difficulty and discrimination levels for each question of the test and has related them with the task variables. Finally this work shows the criteria by which tasks should be programmed to control the automatic generation of the tests and evaluate students with different ability levels, and then improve the evaluation processes of the course.*

**Keywords:** *statics; learning assessment; evaluation systems; IRT*

## 1. Introducción

Con el fin de mejorar los procesos de aprendizaje de los estudiantes de Estática en los programas de ingeniería, se viene desarrollando e implementando un sistema de evaluación dinámico por medio del uso de plataformas e-learning en el cual se construyen ítems de evaluación a través de una programación en java. El estudiante puede utilizar el sistema para prepararse para las evaluaciones de la materia con tareas similares a las que encontrará en los exámenes del curso (Restrepo, 2012). Aunque la tarea que se le propone al estudiante cada vez que utiliza el sistema es la misma (tanto en su entrenamiento como en el examen), algunas de las variables del enunciado son modificadas automáticamente por el programa, buscando que los estudiantes se centren realmente en aprender los procedimientos o los conceptos necesarios para realizar la tarea y se evite que las respuestas correctas sean memorizadas o copiadas.

Conocer la relación entre las variables de una tarea y los parámetros de la misma puede ser además de mucha utilidad si en el sistema se incluye un algoritmo que, en el proceso de entrenamiento, proponga tareas en forma selectiva, considerando el nivel de habilidad del estudiante, es decir que, se le propongan al estudiante tareas cada vez más complejas de acuerdo con el progreso que él vaya logrando.

A cada tarea se le pueden asociar unos parámetros característicos que miden su dificultad y el grado de discriminación, utilizando la teoría de la respuesta al ítem. Si los parámetros varían significativamente al modificar una variable de una tarea, en los algoritmos del sistema de evaluación, se deben considerar estas variaciones o se deberán generar tareas independientes.

En este trabajo se presentan los resultados de los niveles de dificultad y discriminación de un grupo de tareas programadas para evaluar el logro de las competencias en el cálculo de las fuerzas de reacción y las fuerzas internas (fuerza cortante y momento flector) en problemas estáticamente determinados de vigas con dos apoyos. Para estos cálculos se ha aplicado una prueba a un grupo 90 estudiantes que han cursado la asignatura en el primer semestre de 2013. La prueba consta de 14 preguntas para cada estudiante variando aleatoriamente algunos de los parámetros como se indica más adelante.

## 2. El sistema de evaluación dinámico

El sistema de evaluación dinámico es un módulo de un sistema de gestión de aprendizaje, LMS (Learning Management System) al que se le ha incluido un programa, desarrollado en Java que genera ejercicios de diferentes temas de la asignatura, modificando aleatoriamente algunas variables, y que puede ser utilizado por los estudiantes en su proceso de aprendizaje para entrenarse en la solución de problemas, relacionados

con un módulo específico, que serán iguales a los que utilizará el profesor para evaluar algunos de los logros de ese módulo.

La plataforma le permite al profesor programar los detalles de la prueba que utilizará para evaluar a los estudiantes (tipo y número de ejercicios, porcentajes de peso en la calificación global, tiempos de inicio y duración de la prueba, estudiantes habilitados para presentarla, resultados de la evaluación, etc.). Como la plataforma genera una versión diferente del ejercicio cada vez que un estudiante accede, el profesor puede habilitar el examen para que este se entrene antes de ser evaluado sin temor a que se aprenda o copie las respuestas. Es decir, el estudiante dispondrá de una herramienta para autoevaluarse y conocer su estado en relación a unos objetivos del aprendizaje del curso que están alineados con la forma y tipo de evaluación. El tener el sistema en una plataforma LMS facilita las labores de administración relacionadas con los permisos de acceso a las evaluaciones, el almacenamiento y análisis de la información que resulta del uso del sistema.

Este sistema de evaluación se está implementando como apoyo en Estática, asignatura *que* hace parte de la formación básica de los ingenieros. El objetivo central de esta asignatura es analizar las condiciones de equilibrio de un sistema estructural utilizando modelos matemáticos y los conceptos de la física mecánica. Para evaluar el logro de este objetivo se requiere evidenciar que, dado un sistema estructural, el estudiante puede determinar ciertas variables asumiendo condiciones de equilibrio estático a partir de procesos analíticos (identificar o calcular fuerzas externas, determinar la relación entre las fuerzas y algunas variables geométricas del sistema, etc.). Un ejemplo, que es el que se toma en este trabajo, es el de determinar y analizar las fuerzas en una viga con diferentes condiciones de apoyo y tipos de carga (Bedford, 2008, Beer, 2007 y Boresi, 2011).

### 3. Teoría de respuesta al ítem (IRT)

La teoría de la respuesta al ítem (IRT) es utilizada para medir el desempeño de los estudiantes en una prueba. De acuerdo con esta teoría es posible obtener características de los ítems de una evaluación y puntajes de habilidad de diferentes estudiantes que no dependen del grupo de examinados. La base de las mediciones en la teoría IRT es la función característica del ítem que relaciona la probabilidad de que una persona responda correctamente el ítem en función de la habilidad que esa persona haya desarrollado. La función IRT de tres parámetros se expresa como:

$$P_i(\theta, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{e^{[a_i(\theta - b_i)]}}{1 + e^{[a_i(\theta - b_i)]}}$$

**Ecuación 1 Función IRT de tres parámetros.**

Siendo

$P_i$ , la probabilidad de que un estudiante con habilidad  $\theta$ , responda correctamente el ítem  $i$ .

$a_i$ ,  $b_i$ , y  $c_i$ , los parámetros que representan la capacidad de discriminación del ítem y la dificultad y la probabilidad de adivinación del ítem respectivamente.

$\theta$ , la habilidad, en relación a lo que se quiere medir, de los estudiantes que toman la prueba.

$e$ , el número de Euler.

La dificultad del ítem,  $b_i$ , coincide con la habilidad de los estudiantes que tienen un 50 % de probabilidad de responder acertadamente el ítem, la capacidad de discriminación,  $a_i$ , está relacionada con la pendiente de la función evaluada en el punto donde la habilidad es igual a la dificultad del ítem y la probabilidad de adivinación del ítem,  $c_i$ . En la figura 3 se presentan las curvas características de ítems con diferentes valores de dificultad ( $b_i = -2.5$ ,  $b_i = 1.0$  y  $b_i = 1.5$ ), capacidad de discriminación ( $a_i = 1.5$  y  $a_i = 3.1$ ) y probabilidad de adivinación ( $c_i = 0.5$ ,  $c_i = 0.25$  y  $c_i = 0.1$ ).

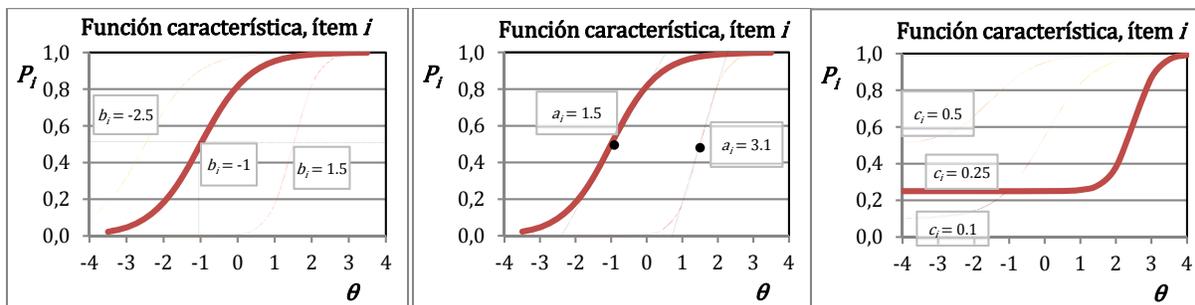


Figura 1 Representación gráfica de los parámetros de la función IRT

Haciendo  $c_i = 0.0$  se elimina la probabilidad de adivinación y así se tiene un caso especial con solo dos parámetros  $a_i$  y  $b_i$  y si además se hace  $a_i = 1.0$ , se elimina el factor de discriminación obteniéndose un caso especial con solo un parámetro.

Para estimar los parámetros de un conjunto de ítems, se debe aplicar una prueba que contenga dichos ítems a una muestra de estudiantes. En el proceso de estimación de los parámetros se halla además la habilidad,  $\theta$ , de cada uno de los estudiantes que tomaron la prueba (Partchev, I, 2004). El procedimiento puede resumirse de la siguiente manera:

- Con las respuestas de los estudiantes se construye la matriz de resultados que se presenta en la tabla 1, donde un “1” significa que el estudiante  $j$  respondió acertadamente al ítem y un “0” que falló en su respuesta.

	1-1	1-2	2	3	4	5	6	7-1	7-2	8	9	10	11	12
Est01	0	0	0	0	1	1	0	0	0	1	1	0	0	0
Est02	0	1	1	1	1	0	0	0	0	1	1	1	0	0
Est03	1	0	1	0	1	1	1	0	0	1	0	1	0	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Est j	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Est88	1	1	1	1	1	0	0	1	1	1	1	0	0	0
Est89	1	1	1	1	1	0	1	1	0	1	0	0	0	1
Est90	1	1	1	1	1	0	0	0	1	1	1	1	1	0

Tabla 1. Resultados evaluación por parte de los estudiantes

- Se define el modelo a utilizar (uno, dos o tres parámetros)
- Se suponen unos valores iniciales de los parámetros de todos los ítems, es decir, se suponen las funciones características para cada ítem.
- Se calcula la función de probabilidad de que un estudiante haya contestado el test como lo hizo para diferentes valores de habilidad (función de verosimilitud, ec. 2).
- Se determina para qué valor de habilidad la función de verosimilitud toma su máximo valor y esta habilidad se asume como la habilidad del estudiante.
- Se distribuyen los estudiantes en un cierto número de rangos de habilidad y para cada rango se evalúa el porcentaje de estudiantes que respondieron correctamente cada ítem con lo que se obtiene una función de aciertos por cada ítem.

- Si se asume que esta función de aciertos es la función característica del ítem pueden hallarse unos nuevos parámetros por un procedimiento de ajuste de curvas.
- Se comparan estos parámetros con los inicialmente supuestos y este procedimiento se repite hasta obtener una convergencia en los valores.

$$L_j(\theta) = \prod_i [P_i(\theta, a_i, b_i, c_i)^{u_i} Q_i(\theta, a_i, b_i, c_i)^{1-u_i}]$$

**Ecuación 2 Función de verosimilitud.**

Donde

$L_j$ , es la función de verosimilitud (probabilidad de que cada estudiante haya contestado la prueba como lo hizo).

$P_i$ , es la probabilidad de que un estudiante con habilidad  $\theta$ , responda correctamente el ítem  $i$ .

$Q_i$ , es la probabilidad de que un estudiante con habilidad  $\theta$ , responda incorrectamente el ítem  $i$ .

$u_i$ , es igual a uno (1.0) si el ítem  $i$  ha sido contestado correctamente por el estudiante  $j$  ó cero (0.0) en caso contrario.

#### 4. Descripción de la prueba y resultados por estudiante

Se elaboró una prueba con 14 ítems del módulo de vigas que fue aplicada a un grupo de 90 estudiantes del curso de Estática en el primer semestre de 2013, una descripción de los ítems se muestra en la tabla 2.

<b>Ejercicio</b>	<b>N. de cargas</b>	<b>Tiempo</b>	<b>Enunciado</b>
<i>ítem 1-1</i>	1	1 minuto	Determinar la reacción vertical en el apoyo fijo
<i>ítem 7-1</i>	2	2 minutos	
<i>ítem 1-2</i>	1	1 minuto	Determinar la reacción vertical en el apoyo patín
<i>ítem 7-2</i>	2	2 minutos	
<i>ítem 2</i>	1	1 minuto	El diagrama de cuerpo libre, DCL, para esta viga es:
<i>ítem 8</i>	2	1 minuto	
<i>ítem 3</i>	1	3 minutos	La gráfica de cortante, V, correspondiente a esta viga es:
<i>ítem 9</i>	2	3 minutos	
<i>ítem 4</i>	1	3 minutos	La gráfica de momento, M, correspondiente a esta viga es:
<i>ítem 10</i>	2	3 minutos	
<i>ítem 5</i>	1	3 minutos	Determinar la magnitud de la fuerza cortante, V, a una distancia determinada
<i>ítem 11</i>	2	3 minutos	
<i>ítem 6</i>	1	3 minutos	Determinar la magnitud del momento flector, M, a una distancia determinada.
<i>ítem 12</i>	2	3 minutos	

**Tabla 2 Descripción de los ítems de la prueba**

Cada ítem es único para cada estudiante ya que el sistema de evaluación es dinámico y varía los datos, la figura y las dimensiones para cada caso, es decir, en cada ítem el sistema presenta una viga diferente. Como se puede observar en la tabla, la mitad de los ítems corresponden a problemas de vigas con una sola carga y la otra mitad con dos cargas. Cada una de las cargas podía ser una fuerza puntual, un momento flector o una carga distribuida uniforme.

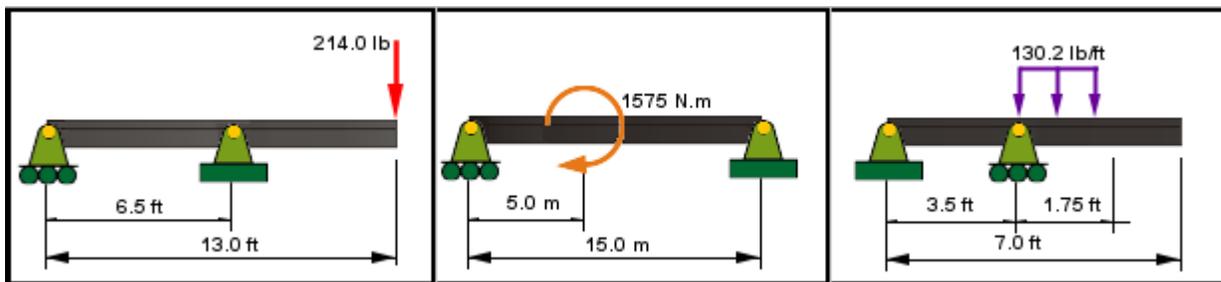


Figura 2. Vigas sometidas a una carga.

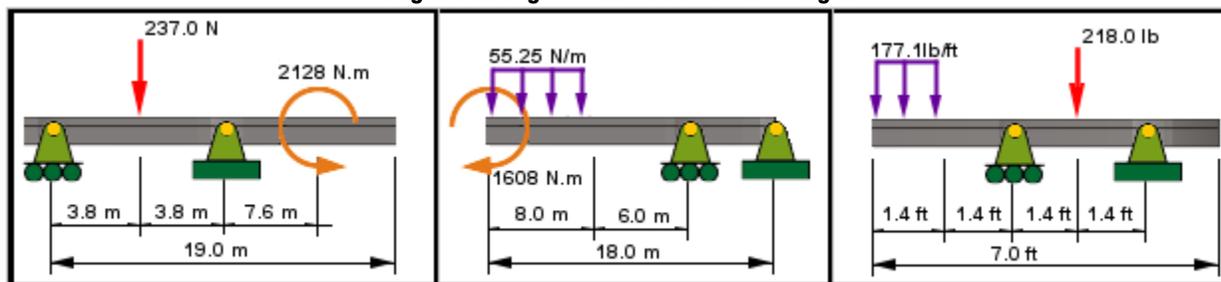


Figura 3. Vigas sometidas a dos cargas

### 5. Estimación de parámetros

Para el análisis de los ítems se utilizó el modelo de 2 parámetros.

$$P_i(\theta, a_i, b_i) = \frac{e^{[a_i(\theta-b_i)]}}{1 + e^{[a_i(\theta-b_i)]}}$$

Ecuación 3 Función IRT de dos parámetros

Para la estimación de los parámetros de dificultad y discriminación se hizo uso de un software de estimación IRT, IRTLDRIF201 (<http://www.unc.edu/~dthissen/dl.html>). Al ingresar los resultados obtenidos por los estudiantes en la prueba se obtuvieron los parámetros de dificultad y discriminación que se presentan en la tabla 3 y en la figura 4.

	Ítem 1-1	Ítem 1-2	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6	Ítem 7-1	Ítem 7-2	Ítem 8	Ítem 9	Ítem 10	Ítem 11	Ítem 12
<i>a, dis</i>	1,69	1,39	1,24	1,33	0,37	0,41	0,39	1,14	0,45	0,60	0,53	0,55	0,64	0,49
<i>b, dif</i>	-0,71	-1,03	-3,02	-1,39	-1,48	-1,08	-0,09	-0,66	-0,52	-4,15	-0,95	-0,16	-0,36	0,14

Tabla 3. Parámetros de dificultad y discriminación obtenidos con IRTLDRIF201.

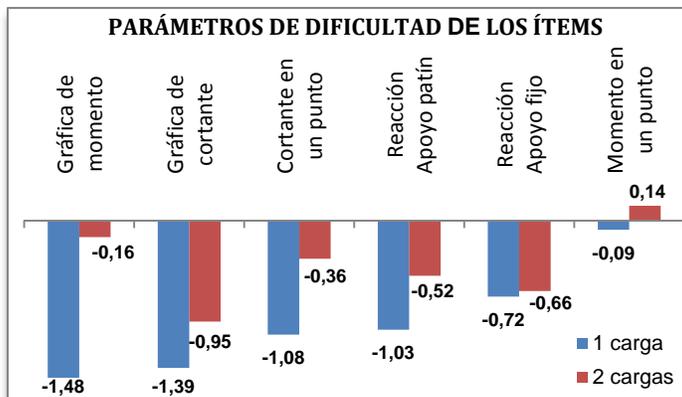


Figura 4. Gráficas de las dificultades de los ítems

Las funciones características de los ítems se muestran en la figura 5.

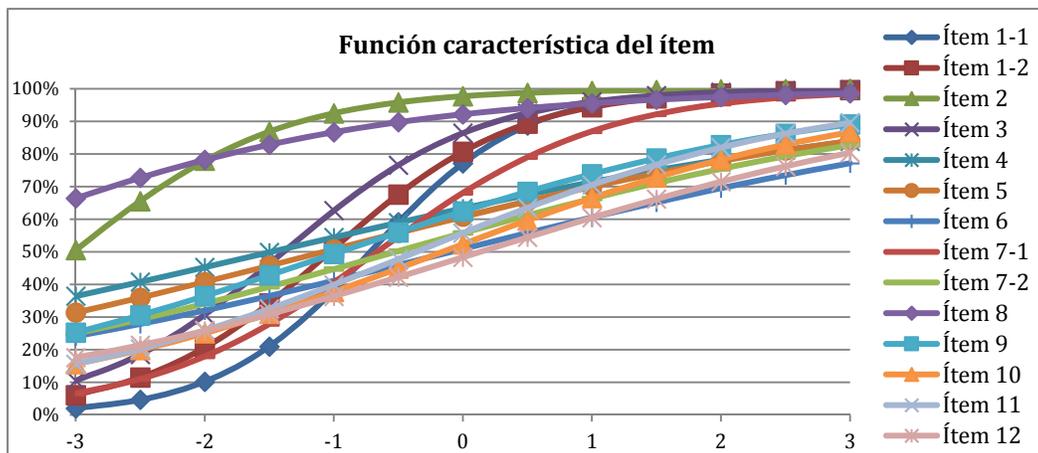


Figura 5. Gráficas de la función característica de cada ítem

A partir de los resultados de la tabla 1 y las funciones características de los ítems es posible calcular una función de probabilidad de que un estudiante haya contestado el test como lo hizo, para diferentes valores de habilidad con la ecuación 2, como se observa en la tabla 4.

	-3.00	-2.50	-2.00	-1.50	-1.00	-0.50	0.00	0.50	1.00	1.50	2.00	2.50	3.00	H.P.E
e2	0.24%	0.18%	0.09%	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	-3.00
e35	0.10%	0.07%	0.04%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	-3.00
e85	0.09%	0.13%	0.12%	0.07%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	-2.50
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
e77	0.00%	0.00%	0.00%	0.00%	0.02%	0.04%	0.05%	0.04%	0.02%	0.01%	0.00%	0.00%	0.00%	0.00
e89	0.00%	0.00%	0.00%	0.01%	0.03%	0.05%	0.05%	0.03%	0.02%	0.01%	0.00%	0.00%	0.00%	0.00
e15	0.00%	0.00%	0.00%	0.01%	0.03%	0.09%	0.15%	0.16%	0.11%	0.07%	0.03%	0.01%	0.01%	0.50
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
e76	0.00%	0.00%	0.00%	0.00%	0.00%	0.04%	0.18%	0.55%	1.19%	2.06%	3.01%	3.91%	4.63%	3.00
e82	0.00%	0.00%	0.00%	0.00%	0.01%	0.07%	0.29%	0.77%	1.50%	2.30%	3.01%	3.49%	3.69%	3.00
e90	0.00%	0.00%	0.00%	0.00%	0.01%	0.06%	0.32%	1.14%	2.97%	6.16%	10.83%	16.91%	24.08%	3.00

Tabla 4. Función de probabilidad de que un estudiante haya contestado el test como lo hizo

La probabilidad más alta para cada estudiante estima la habilidad de éste, como por ejemplo el e2 (estudiante 2) tiene más probabilidad de tener una habilidad de -3.

## 6. Conclusiones

A partir del parámetro de dificultad encontrado se puede observar como los ítems que involucraban dos cargas se hacen más difíciles que los que involucraban solo una, a pesar de que se pregunta lo mismo, lo cual coincide con el sentido común. A manera de ejemplo se puede ver como el ítem 3; donde se pregunta por La gráfica de cortante,  $V$ , correspondiente a la viga pero con solo una carga; tiene un valor de dificultad de -1.39, mientras que el ítem 9; donde se hace la misma pregunta pero para una viga sometida a dos cargas diferentes; tiene un valor de dificultad de -0.95, la diferencia entre ambos valores de dificultad de estos dos ítems es de 0.44.

Los ítems 2 y 8 deben ser revisados o eliminados de esta prueba pues han sido respondidos por más del 96 % de los estudiantes evaluados lo que se refleja en los niveles de dificultad encontrados, -3.02 y -4.15. Por esta razón en la construcción de la figura 4 se han eliminado los estos ítems, ya que sus resultados no son aptos para este estudio.

Los valores determinados para los parámetros de los ítems con la IRT requieren una muestra mayor para obtener conclusiones estadísticamente válidas. Sin embargo con este trabajo se ha comenzado a implementar una herramienta que va a permitir evaluar los ítems del sistema de evaluación dinámico.

Es claro que el sistema de evaluación de generación de preguntas dinámicas que se ha venido implementando en la Universidad EAFIT está en constante mejoramiento, aunque la prueba realizada muestra que algunos de los puntos programados en el sistema deben ser mejorados y reprogramados para que sean realmente útiles tanto en los procesos de evaluación como para el entrenamiento en el proceso de aprendizaje de los estudiantes que hacen uso de la herramienta.

## 7. Referencias

- Bedford. (2008). *Mecánica para ingeniería ESTÁTICA*. Pearson, México, pp. 486
- Beer, y Johnston. (2007). *Estática y Mecánica de Materiales*. McGraw-Hill, Inc. ER.
- Boresi P, and Schmidt R (2011). *Engineering Mechanics: Statics General Engineering*. Brooks/Cole, Pennsylvania State University, pp. 553.
- Partchev, I. (2004). A visual guide to item response theory [Document]. Friedrich-Schiller-Universität Jena. Consultado el 14 de febrero de 2013 en: <http://www.metheval.uni-jena.de/irt/VisualIRT.pdf>
- Restrepo, J. L. (2012). Generador automático de tareas como apoyo a los procesos de evaluación, asignatura estática, Reunión Nacional ACOFI.
- Thissen, D. (2001). IRTLRF v.2.0b [Software]: Software for the Computation of the Statistics Involved in Item Response Theory Likelihood-Ratio Tests for Differential Item Functioning [Document], The University of North Carolina (UNC). Descargado el 5 de mayo de 2013 desde <http://www.unc.edu/~dthissen/dl.html>

## Sobre los autores

- **Jorge Luis Restrepo Ochoa:** Ingeniero Mecánico, Doctor en ingeniería. Profesor titular Departamento de Ingeniería Mecánica. [jrestrep@eafit.edu.co](mailto:jrestrep@eafit.edu.co)
- **Jaime Leonardo Barbosa Pérez:** Ingeniero Mecánico, Magister en Ingeniería, Especialista en Diseño Mecánico. Profesor Asistente Departamento de Ingeniería Mecánica. [jbarbosa@eafit.edu.co](mailto:jbarbosa@eafit.edu.co)
- **Andrés Restrepo Cadavid:** Ingeniero Mecánico, Estudiante de maestría en ingeniería. Profesor cátedra Departamento de Ingeniería Mecánica. [arestr44@eafit.edu.co](mailto:arestr44@eafit.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)