

Revista Digital: Matemática, Educación e

Internet

ISSN: 1659-0643

revistadigitalmatematica@itcr.ac.cr

Instituto Tecnológico de Costa Rica

Costa Rica

Rojas T., Luis

Predicción de la dificultad de la prueba de Habilidades Cuantitativas de la Universidad de  
Costa Rica

Revista Digital: Matemática, Educación e Internet, vol. 13, núm. 1, agosto-febrero, 2013,  
pp. 1-14

Instituto Tecnológico de Costa Rica  
Cartago, Costa Rica

Disponible en: <https://www.redalyc.org/articulo.oa?id=607972989004>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

# Predicción de la dificultad de la prueba de Habilidades Cuantitativas de la Universidad de Costa Rica.

Luis Rojas T.

[luisrojasxtorres@gmail.com](mailto:luisrojasxtorres@gmail.com)

Escuela de Matemática - Instituto de Investigaciones Psicológicas

Universidad de Costa Rica

**Resumen.** El objetivo de este artículo es plantear un modelo estadístico que permita predecir la dificultad de los ítems de la prueba de Habilidades Cuantitativas de la Universidad de Costa Rica. Para esto se plantearon una serie de atributos de los ítems utilizados en la prueba, con el fin de estimar su influencia en la predicción de la dificultad del modelo de Rasch y la probabilidad de acierto de cada uno de los ítems. Para lograr dicho propósito se ha recurrido principalmente al uso del modelo LLTM y al análisis de sus supuestos; sin embargo, también se utilizan otros modelos como el lineal múltiple y el logístico multinivel para encontrar similitudes y diferencias entre los resultados y así poder establecer criterios de comparación.

En este estudio se obtuvo que las dificultades estimadas por el modelo LLTM y las dificultades estimadas por el modelo de Rasch tuvieron un coeficiente de correlación de 0.80, además se estimó que un 63.80% de la variabilidad de la dificultad de los ítems de la prueba es explicada por las variables tomadas en cuenta, de las cuales resultaron significativas: visualización, cantidad de contenidos, planteamiento de relación y cantidad de algoritmos algebraicos.

**Palabras claves:** Razonamiento cuantitativo, atributos del ítem, modelo de Rasch, LLTM, modelo logístico multinivel.

**Abstract.** The aim of this article is to pose a statistical model that allows predicting difficulty of Quantitative Abilities test's items of Universidad de Costa Rica. A series of attributes from items used in the test were posed in order to estimate its influence on Rash prediction of difficulty and the success probability of every single item. For accomplishing this purpose LLTM model was applied and also the analysis of its suppositions, however, other models like linear multiple model and the logistic multilevel model were also applied for finding similitude and differences among the results and therefore establishing comparative criteria.

It was found that the difficulties estimated by LLTM model with those estimated by Rasch model had a coefficient of relation of 0.80, a 63.80% of the variability of the test's items explained by the variables that were taken into account, the attributes with a higher impact on the difficulty of the item are: visualization, content's quantity, the approach of relation and the quantity of algebraic algorithms.

**Key words:** Quantitative reasoning, item's attributes, Rasch's model, LLTM, logistic multilevel model.

## 1.1 Introducción

La prueba de Habilidades Cuantitativas (HC) forma parte de un conjunto de pruebas creadas por el Instituto de Investigaciones Psicológicas (IIP) de la Universidad de Costa Rica (UCR), denominadas pruebas específicas para ingreso a carrera. Estas se desarrollan como parte de una iniciativa para un nuevo sistema de admisión a la UCR, el cual toma en cuenta las habilidades particulares requeridas en cada carrera. En la actualidad, el equipo de investigación a cargo de estas pruebas se encuentra en una etapa de recolección de evidencias de validez, para asegurar su idoneidad como instrumentos de selección específica y diagnóstico.

HC nace con el objetivo de que en las carreras que requieren de destrezas en el campo matemático, se logre "mejorar el proceso de selección de sus estudiantes y garantizar de alguna forma la permanencia y graduación de estos"

(Proyecto de Habilidades Cuantitativas, 2010, p.3). Esta prueba se desea implementar como instrumento de selección en las carreras para las cuales se requiere un buen manejo de información cuantitativa (Matemática, Estadística, Física, entre otras).

En esta prueba no se miden habilidades básicas de razonamiento en contextos numéricos, tal y como se hace en la prueba de ingreso a la UCR (PAA), sino que se mide razonamiento cuantitativo, el cual se define como “la capacidad de analizar la información cuantitativa, incluyendo la determinación de qué técnicas y procedimientos se pueden aplicar a un problema particular para llegar a su solución” (Dwyer, Gallagher, Levin y Morley, 2003, p.12).

Debido a que el objetivo de esta prueba es seleccionar a los y las estudiantes con las mayores probabilidades de éxito en las carreras en cuestión, se debe asegurar que esta cumpla con todos los estándares de calidad que se exigen a las pruebas estandarizadas a nivel mundial. Entre estos, destacan los siguientes: que cada ítem de la prueba mida únicamente el constructo que se pretende evaluar y que la dificultad total de la prueba sea conocida antes de ser aplicada (American Educational Research Association, American Psychological Research Association y National Council on Measurement in Education, 1999).

Dado que es de vital importancia tener claro la dificultad total de la prueba antes de administrarla, se hace necesario la creación de un proceso objetivo que pueda predecir la dificultad de los ítems que van a componer esta. El presente trabajo tiene como principal objetivo plantear un modelo estadístico que permita predecir la dificultad de los ítems de la prueba de Habilidades Cuantitativas de la Universidad de Costa Rica.

Este objetivo conlleva la tarea de analizar detalladamente la estructura de los ítems que componen la prueba. Dicha labor trae consigo varios beneficios, por ejemplo, Romero, Ponsoda y Ximénez (2008) mencionan que este proceso permite obtener “un mayor entendimiento de las habilidades específicas requeridas para resolver el prueba y para el desarrollo de instrumentos válidos” (p. 86), también exponen que esta estructuración permite detectar en cuáles componentes los examinados tienen carencias y fortalezas; por otro lado, Messik (1995) plantea que dicha metodología arroja evidencias sustantivas de validez de constructo.

## 1.2 Referentes teóricos

---

En la presente sección se exponen algunos atributos de los ítems identificados en varias pruebas asociadas con la matemática, posteriormente se presentan los atributos utilizados en la prueba de Habilidades Cuantitativas de la Universidad de Costa Rica y, finalmente, se hace una referencia breve sobre la dificultad de los ítems en pruebas estandarizadas y modelos estadísticos utilizados para predecir estas dificultades.

### 1.2.1 Atributos de los ítems

Cuando se hace una revisión de la literatura psicométrica, se puede encontrar una gran cantidad de estudios en los que se ha tratado de predecir la dificultad de los ítems de alguna prueba (Real, Olea, Ponsoda, Revuelta y Abad, 1999; Romero, Ponsoda y Ximénez, 2006; Romero et al., 2008; Embretson y Daniel, 2008). Dichos estudios presentan una gran semejanza en la metodología, ya que se utilizan como predictores ciertos atributos de los ítems.

Estos atributos de los ítems deben ser definidos mediante la utilización de una teoría subyacente o por medio de la realización de estudios empíricos sobre los procesos involucrados en la resolución de los ítems y las características particulares de los reactivos (Romero et al., 2008). Así, por ejemplo, Embretson et al. (2008) plantean un modelo cognitivo para la resolución de ítems con problemas matemáticos basado en la teoría de resolución de problemas matemáticos de Mayer, Larkin y Kadane (1984). Dicho modelo se presenta como un diagrama de flujo, en el cual se plasman las decisiones que debe tomar el examinado para poder llegar a la solución del ejercicio.

En Real et al. (1999), se describen todos los procesos requeridos para resolver cada ítem de una prueba elaborada para medir el uso correcto de los signos positivo y negativo en las operaciones elementales de la aritmética y de

igual manera que en el estudio anterior, se establecen diagramas de flujo que describen todo el proceso necesario para llegar a las respuestas de los ejercicios, mediante la utilización de los procesos previamente establecidos. En Romero et al. (2008), se definen seis operaciones cognitivas que subyacen en los procesos de resolución de una prueba de aritmética y luego se asigna a cada ítem una clasificación en la que se indica si necesitan el uso de alguno de estos procesos para llegar a la respuesta correcta.

En Attorresi, Picón, Abal, Aguerri y Galibert (2009) se hace una investigación para predecir la dificultad de los reactivos de una prueba de razonamiento deductivo y se establecen como atributos de los ítems a las características específicas de los ítems, como por ejemplo, uso de lenguaje simbólico o textual, presencia de alguna estructura lógica determinada (*modus ponens* o *modus tollens*) y presencia de cuantificadores.

En estas investigaciones se observa que la definición de los atributos de los ítems puede llevarse a cabo con base en una teoría psicológica o pedagógica específica, o a partir de las intuiciones de quienes construyen los ítems.

### **1.2.2 Atributos de los ítems de la prueba de Habilidades Cuantitativas**

En esta sección se enuncian los atributos definidos para los ítems de la prueba de Habilidades Cuantitativas, los cuales son una adaptación de los propuestos para los ítems de la PAA en Villalobos y Molina (2010).

Los atributos propuestos para HC fueron: Niveles de deducción, inducción, visualización y comparación de cantidades que requiere el ítem para su resolución; cantidad de hipótesis, contenidos matemáticos, algoritmos algebraicos y algoritmos aritméticos que están presentes en el ítem; además si el ítem exige para su solución el planteamiento de alguna ecuación o el análisis de cada opción por separado. A continuación se evidencia cómo algunos de estos atributos pueden ser justificados desde el marco teórico que sustenta la prueba.

En la justificación del Proyecto de Habilidades Cuantitativas (2010) se hace hincapié en que para la construcción de la prueba se tomó en cuenta la investigación de Álvarez, Molina y Salazar (2000), la cual buscaba encontrar cuáles habilidades eran necesarias para obtener éxito académico en la Universidad de Costa Rica. Entre las aptitudes encontradas que fueron de interés para el desarrollo de la prueba, destacan: intuición, capacidad de análisis, abstracción, habilidad para interpretar modelos matemáticos, habilidades operativas, interpretación de problemas y la solución de estos. Al analizar estas habilidades, se detectó que pueden ser encontradas, en su mayoría, en la dimensión de la habilidad cuantitativa del razonamiento plateada por Carroll (1993) en su teoría de los tres estratos de la inteligencia.

Posteriormente, coincidiendo con el marco teórico de la prueba, se detectó que en su mayoría, los ítems se asemejaban al factor Razonamiento Cuantitativo (RQ) que propone Carroll, aptitud que requiere del razonamiento basado en relaciones y propiedades matemáticas. Debido a esto, muchos de los atributos de los ítems se pueden deducir de la dimensión del RQ de la teoría de los tres estratos de la inteligencia (Proyecto de Habilidades Cuantitativas, 2010).

La definición de razonamiento cuantitativo sugiere que entre los atributos por medir en los ítems pueden estar la cantidad de contenidos matemáticos que debe dominar el examinado para poder resolver el ítem y la cantidad de datos matemáticos que presenta el ítem. La teoría de Carroll menciona que en ítems de RQ se debe evitar el uso de elementos que impliquen habilidades de visualización espacial, así como también evitar el planteamiento de información con una gran complejidad numérica. Además, se debe velar por que en estos reactivos la computación numérica requerida sea muy pequeña, esto sugiere que entre los atributos de los ítems que han de ser controlados se encuentran: el razonamiento visoespacial, la cantidad de procesos algebraicos y la de procesos aritméticos.

Carroll menciona que el razonamiento general secuencial (RG) y el razonamiento inductivo (RI) también pueden estar incluidos en ítems de razonamiento cuantitativo. En HC los procesos deductivos asociados al RG y los inductivos asociados al RI generalmente son requeridos para la resolución de los ítems. Debido a esto, se pueden definir como atributos de los ítems el nivel de deducción e inducción que se necesita para solucionar los ítems.

Por tanto queda en evidencia que la mayoría de los atributos adoptados por la prueba de Habilidades Cuantitativas pueden ser justificados con base en una de las teorías ampliamente aceptada en el ámbito de la psicometría y la evaluación educativa.

### 1.2.3 Dificultad de los ítems

Para determinar la dificultad de los ítems hay varias teorías psicométricas que establecen escalas basadas en los resultados obtenidos por una población en la ejecución de la prueba, entre las cuales están la Teoría Clásica de los Test (TCT) y la Teoría de Respuesta al Ítem (TRI). Muñiz (2010) menciona que en la actualidad la TRI acapara la mayoría de las investigaciones en el campo psicométrico. En todos los trabajos revisados referentes a la predicción de la dificultad de los ítems, se utilizó la dificultad estimada mediante el modelo de Rasch, ( $b_i$ ).

El modelo de Rasch es un caso particular de la TRI con un parámetro y su fórmula es la siguiente

$$P(x_{ij} = 1/\theta_j, b_i) = \frac{1}{1 + e^{(b_i - \theta_j)}}$$

donde  $b_i$  es la dificultad del ítem  $i$ ,  $\theta_j$  la puntuación de la persona  $j$  en la variable medida y  $P(x_{ij} = 1/\theta_j, b_i)$  la probabilidad de acertar el ítem  $i$  para la persona  $j$  (Muñiz, 2010).

Este modelo se desarrolla bajo los supuestos de unidimensionalidad e independencia local, el primero indica que la prueba mide un único constructo y el segundo significa que la probabilidad de contestar correctamente un ítem no depende del acierto de algún otro ítem de la prueba (Villareal, 2011).

### 1.2.4 Modelo LLTM

El modelo LLTM (Fischer, 1973) es una variación del modelo de Rasch, donde en vez de estimarse un parámetro  $b$  para determinar la dificultad del ítem, se estiman una serie de parámetros  $n_k$ , asociados a los atributos que están presentes en al menos uno de los ítems examinados. Los parámetros asociados a los atributos, al ser combinados linealmente, sirven para estimar el parámetro  $b$  del modelo de Rasch. La fórmula que describe el modelo LLTM es la siguiente:

$$P(x_{ij} = 1/\theta_j, q, n) = \frac{1}{1 + e^{\sum_{k=1}^K q_{ik}n_k - \theta_j}}$$

donde  $n_k$  es el efecto del atributo  $k$  en la predicción de la dificultad de los ítems,  $q_{ik}$  es el peso del atributo  $k$  en el ítem  $i$ ,  $\theta_j$  la puntuación de la persona  $j$  en la variable medida y  $P(x_{ij} = 1/\theta_j, q, n)$  la probabilidad de acertar el ítem  $i$  para la persona  $j$  (Embretson y Daniel, 2008).

Este método pertenece a la familia de los modelos componenciales que buscan tomar en cuenta los procesos cognitivos que se requieren para resolver un ítem y es en la actualidad uno de los más usados para predecir la dificultad Rasch de los ítems (Romero et al., 2006).

Para poder aplicar este modelo se deben realizar tres contrastes básicos: los datos deben ajustarse a un modelo de Rasch, el modelo debe reproducir de forma satisfactoria las dificultades dadas por el modelo de Rasch y debe validarse el modelo cognitivo en que se basa. Los dos primeros contrastes son relativamente fáciles de verificar; sin embargo, el tercero ha sido poco desarrollado y generalmente se realiza en una segunda etapa de la investigación, la cual no se desarrolla en este estudio (Romero et al., 2008).

Para realizar el primer contraste, el índice de ajuste utilizado es el estadístico *infit* (media de los residuos cuadráticos ponderados con su variancia), cuyo valor esperado es 1. Este se calcula tanto para los ítems como para las personas. Según Prieto y Delgado (2003) para muestras menores a 500 casos el infit –de los ítems y de las personas– debe estar

entre los valores 0.7 y 1.3 para que pueda considerarse que un ítem o persona se ajusta al modelo de Rasch. Es importante recalcar que antes de realizar este contraste, se debe el cumplimiento de los supuestos del modelo de Rasch.

El segundo supuesto se analiza mediante el contraste con el estadístico chi cuadrado de Andersen, cuya hipótesis nula es que el modelo LLTM reproduce adecuadamente los parámetros del modelo de Rasch. Dicho estadístico presenta el problema de que muchas veces obliga a rechazar la hipótesis nula aún cuando los otros contrastes indiquen lo contrario. Otro diagnóstico que se utiliza para verificar este supuesto es la correlación entre las dificultades del modelo de Rasch y las producidas por el modelo LLTM, así como el estudio de un gráfico de dispersión cuyos dos ejes representan, por un lado, la dificultad estimada mediante el modelo de Rasch y, por otro, la estimada con el LLTM.

### 1.2.5 Modelos multinivel

Según Pardo, Ruiz y San Martín (2007) los modelos jerárquicos o multinivel son herramientas que permiten analizar los datos cuando los casos se encuentran agrupados en unidades de información más amplias y en donde se toman medidas tanto en el nivel más bajo (los casos) como en el nivel más alto (grupos).

La fórmula de este modelo es

$$Y_{ij} = \alpha + \beta x_{ij} + \gamma z_j + \varepsilon_{ij}$$

donde  $Y$  es la variable respuesta,  $x$  una variable de individuos,  $z$  una variable de grupos,  $\alpha$  el intercepto,  $\beta$  el coeficiente asociado a la variable de nivel inferior,  $\gamma$  el coeficiente asociado a la variable de nivel superior y  $\varepsilon$  el error de estimación. Además se tienen  $j$  grupos e  $ij$  observaciones (Agresti, 2002).

La razón principal por la que se tomará un modelo multinivel como alternativa de análisis en este estudio se basa principalmente en el incumplimiento de la hipótesis de independencia sobre las que están basados los modelos de regresión tradicionales, al no tomar en consideración los componentes de variancia intra y entre estudiantes.

Los modelos multinivel presentan la ventaja de que además de utilizar la información sobre los individuos (nivel inferior), también hacen uso de las variables relacionadas con grupos de observaciones (nivel superior). Por ejemplo, en Montero, Valverde y Villalobos (2007) se planteó un modelo para predecir el rendimiento académico de los estudiantes de la UCR, con base en la información sobre varios grupos de segundo y tercer año de diversas carreras. Con el fin de mejorar la predicción, no solo se utilizaron variables asociadas a los estudiantes (nivel inferior), sino que también se tomó en cuenta la información recolectada sobre los docentes (nivel superior). Esta metodología permitió determinar qué porcentaje de la variabilidad del rendimiento académico se debió a la variabilidad entre estudiantes y qué porcentaje a la variabilidad entre profesores.

Oliver, Rosel y Jara (2000) indican que la ventaja principal de los modelos multinivel se basa en que permiten obtener mejores estimaciones de los coeficientes de regresión y de su variación que los modelos tradicionales, además de que estiman la contribución de los individuos y los grupos en la variabilidad de la respuesta. En el caso de este estudio se desea estimar qué porcentaje de la variabilidad en los aciertos de los ítems es debido a la variabilidad entre ítems y qué tanto se debe a la variabilidad entre individuos. Por esto, se consideró pertinente utilizar un modelo multinivel, en el cual los examinados sean el nivel inferior y los ítems, el nivel superior.

Es importante mencionar que al tener una variable dicotómica (como en el modelo acá planteado), no se cumple el supuesto de normalidad de la variable respuesta, necesario para modelar una variable linealmente. Sin embargo, es conocido que si  $p$  es la esperanza de una variable dicotómica  $Y$ , entonces  $\ln\left(\frac{p}{1-p}\right)$  puede ser modelada linealmente (Kutner, Neter, Nachtsheim y Wasserman, 1996). Por lo tanto, si se usa esta transformación de la variable respuesta en un modelo multinivel se obtiene un modelo logístico multinivel, cuya fórmula es la siguiente

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha + \beta x_{ij} + \gamma z_j + \varepsilon_{ij}$$

donde  $p$  es la probabilidad de éxito en la variable  $Y$  y los demás componentes son los expresados en la ecuación del modelo multinivel simple.

Cabe señalar que en este modelo los coeficientes por sí solos no tienen una interpretación sencilla, no obstante, al ser exponentiados y luego restados en uno, su interpretación se vuelve más comprensible. Lo anterior porque se obtiene el porcentaje de aumento de los *odds* de  $Y$  si se aumenta en una unidad la variable asociada (los *odds* son la razón de las probabilidades de éxito de un evento entre sus probabilidades de fracaso). Por ejemplo, si en la fórmula anterior del modelo logístico multinivel se obtiene que  $e^{\beta} - 1 = 2.3$  unidades, se tiene que *si se aumenta en una unidad la variable  $x$ , los odds de  $Y$  aumentan en 230%*.

Una forma sencilla de interpretar estos coeficientes es la siguiente: si  $e^{\beta} - 1$  es positivo, conforme se aumente el valor de la variable  $x$ , las probabilidades de éxito tenderán a superar las probabilidades de fracaso, en cambio si  $e^{\beta} - 1$  es negativo, las probabilidades de fracaso superarán a las de éxito (Kutner et al., 1996).

## 1.3 Metodología

---

### 1.3.1 Instrumento

La prueba de Habilidades Cuantitativas consta de 40 ítems, divididos en dos categorías: selección única y comparación cuantitativa. Las habilidades cuantitativas las mide utilizando conocimientos básicos de geometría, álgebra, aritmética y análisis de datos, que no sobrepasen los adquiridos por un estudiante que haya aprobado el noveno año de la Educación General Básica de Costa Rica.

Para la aplicación de la prueba se construyen cuatro formularios distintos, con dificultades y contenidos que son equiparados antes de ser utilizados. En este estudio, el análisis se hará con base en la aplicación de la fórmula uno aplicada en 2008.

Este test, junto con uno de razonamiento con figuras y otro de expresión escrita, son construidos con el objetivo de brindar a la Universidad de Costa Rica instrumentos de selección para ingreso a las diferentes carreras ofrecidas (Proyecto de Habilidades Cuantitativas, 2011).

### 1.3.2 Muestra

En el 2008 el programa de Pruebas Específicas realizó una aplicación masiva de sus pruebas a lo largo de todo el territorio costarricense, la cual fue financiada por la Vicerrectoría de Investigación de la UCR. Esta aplicación se realizó en casi todas las sedes utilizadas para la aplicación de la Prueba de Aptitud Académica, de modo que fue posible incorporar a toda la población meta. Particularmente, para HC la población meta fueron todos los estudiantes que cursaban el último año de secundaria y que pretendían ingresar a una carrera de la UCR en la cual se requiriera del razonamiento cuantitativo como herramienta básica de su plan de estudios.

La respuesta a la convocatoria fue relativamente baja situación para la cual hay dos hipótesis fundamentales: 1) la prueba era voluntaria, no obligatoria, como lo es la PAA, y 2) la aplicación se realizó un sábado, es decir, un día en el que los estudiantes no reciben lecciones.

En esta aplicación se utilizaron cuatro formularios de HC, de los cuales se seleccionó la fórmula uno para la realización de este estudio. La fórmula seleccionada fue aplicada a 368 estudiantes, de los cuales 150 (40,9%) eran mujeres y 218 (59,1%), hombres.

### 1.3.3 Procedimiento

Una vez seleccionada la fórmula por analizar, se procedió a seleccionar tres jueces expertos en el campo de las habilidades cuantitativas, los cuales tuvieron como tarea decidir qué atributos estaban presentes en cada ítem. Para esta tarea se les entregó un documento en donde se especificaban las definiciones, niveles y escalas de medición de los atributos planteados para HC (estos serán descritos en la sección 3.4).

La asignación de los atributos se hizo inicialmente de manera individual. Luego, se organizaron varias reuniones entre los tres jueces para comparar los resultados de sus juzgamientos y llegar a consensos sobre los valores más confiables para los atributos.

Con la matriz de atributos de los ítems elaborada, se procedió a correr el modelo LLTM. Además, se calculó un modelo de regresión lineal de las dificultades Rasch de los ítems contra sus atributos y se estimó la probabilidad de acierto de los ítems por medio de un modelo de regresión logística multinivel, en el cual los ítems son el nivel superior y los examinados el nivel inferior. Para estos análisis se utilizaron cuatro paquetes estadísticos: *Winsteps*, *SPSS 19*, *STATA 11* y *R 2.14.0*.

### 1.3.4 Atributos de los ítems

A continuación se presentan los atributos de los ítems, tomados de Villalobos y Molina (2010), con la adaptación de las definiciones hechas por los investigadores e investigadoras de la prueba de Habilidades Cuantitativas (Proyecto de Habilidades Cuantitativas, 2011):

1. Deducción: Proceso por el cual a partir de la veracidad de las premisas se puede obtener una conclusión, pues la información proporcionada por estas es suficiente, desde un punto de vista lógico, para llegar a una conclusión válida.
2. Inducción: Proceso mediante el cual se generaliza una característica que cumple un subconjunto de datos en el contexto del ítem.
3. Ordenamiento: Comparación entre dos expresiones algebraicas o numéricas con el fin de establecer si una expresión es mayor, menor o igual a otra.
4. Visualización: Proceso mediante el cual se extrae información de una representación gráfica dada en el ítem, o bien, en el que se requiera interpretar la situación proporcionada en el ítem mediante diagramas o dibujos.
5. Cantidad de hipótesis: Número de supuestos dados en el encabezado del ítem necesarios para la resolución de este.
6. Número de contenidos: Cantidad de definiciones, teoremas, leyes, etc. que se requieren para resolver el ítem.
7. Número de algoritmos algebraicos: Cantidad de procedimientos algebraicos que se requieren para la solución del ítem.
8. Número de algoritmos aritméticos: Cantidad de procedimientos aritméticos que se requieren para la solución del ítem.
9. Planteamiento de relación: Necesidad de plantear una ecuación o inecuación para resolver el problema.
10. Discriminación de proposiciones: Proceso mediante el cual se analizan exhaustivamente cada una de las opciones para determinar cuál es la clave o las proposiciones (afirmaciones) que se dan en el ítem.
11. Área de contenido: Pertenencia a alguna de las cuatro áreas temáticas de la prueba: Geometría, Aritmética, Análisis de datos o Álgebra.
12. Comparación cuantitativa: Identificación de la relación de orden que presentan dos expresiones matemáticas dadas.

Para la medición de deducción, inducción, ordenamiento y visualización, se elaboró una lista de los procesos necesarios para resolver los ítems. Luego, estos procesos se asociaron a cada uno de los atributos y se clasificaron con los valores 0, 1, 2 y 3, según su nivel de complejidad.

Los atributos *Cantidad de hipótesis*, *Número de contenidos*, *Número de algoritmos algebraicos* y *Número de algoritmos aritméticos* se midieron, como su nombre lo sugiere, por medio del conteo de los elementos en cuestión. Por otro lado, las variables *Planteamiento de relación* y *Discriminación de proposiciones* se codificaron con 0 y 1, donde uno significa que se necesita para resolver el ítem y cero que no se necesita.

Para *Área de contenido* se crearon tres variables llamadas *Análisis de datos*, *Geometría* y *Álgebra*, las cuales toman el valor 1 si el ítem está desarrollado en esas áreas de contenido, de lo contrario, toman el valor 0; de esta manera, cuando las tres variables toman el valor 0, se da a entender que el ítem se desarrolla en el área de contenido para la cual no se creó variable (*Aritmética*), lo cual la convierte en la categoría base o de comparación. Finalmente, al atributo *Comparación cuantitativa* se le asigna el valor 1 si el ítem es de este tipo de formato y 0 si no lo es.

## 1.4 Resultados

---

### 1.4.1 Descriptivos de los datos.

En la tabla 1 se presenta el máximo, la media, la desviación estándar (Desv. Est.) y las correlaciones de los atributos con la dificultad Rasch de los ítems (Corr A-R), todos calculados con el software SPSS 19. Al analizar los datos asociados a las variables, se detectó que *Visualización*, *Cantidad de contenidos* y *Algoritmos algebraicos* están correlacionadas positivamente y de manera significativa (a un nivel del 5%) con la dificultad Rasch de los ítems, esto indica que si se aumenta la cantidad de alguno de esos atributos en el ítem, este presentará mayor dificultad.

**Tabla 1.1** Estadísticas descriptivas de los ítems

	Max	Media	Desv. Est.	Corr A-R
Deducción	2	1.22	0.53	0.087
Inducción	3	0.11	0.51	-0.257
Ordenamiento	3	0.65	0.81	0.306
Visualización	3	0.92	0.85	0.462**
Discriminación de opciones	1	0.16	0.37	-0.107
Número de hipótesis	2	0.81	0.73	-0.148
Número de contenidos	3	1.08	0.82	0.422**
Planteamiento de relación	1	0.46	0.55	0.112
Algoritmos algebraicos	3	1.30	0.96	0.487**
Algoritmos aritméticos	3	0.78	0.66	0.076
Algebra	1	0.30	0.46	0.014
Geometría	1	0.32	0.47	0.289
Análisis de datos	1	0.19	0.39	-0.391**
Comparación Cuantitativa	1	0.16	0.37	0.043

\*\*Significativas a un nivel del 5%

Por otro lado, la categoría *Análisis de datos* correlaciona significativamente con la dificultad, solo que de manera negativa. De lo anterior se deduce que los ítems desarrollados en esta área de contenido son más fáciles que los ítems de las restantes áreas.

Además, se puede apreciar que, en promedio, los ítems necesitan para ser solucionados de un proceso de deducción y otro de visualización. También se puede concluir que, en promedio, todos los ítems requieren para su solución de al menos un algoritmo algebraico y el conocimiento de un contenido de secundaria.

#### 1.4.2 Análisis de supuestos del modelo LLTM

Para que un modelo LLTM pueda ser aplicado en el análisis de una prueba, este debe cumplir los supuestos del modelo de Rasch, principalmente que sus ítems presenten un patrón unidimensional. Para evaluar este supuesto se examinó el gráfico de sedimentación obtenido al hacer un análisis factorial exploratorio con rotación promax. Dicho gráfico se obtuvo con la ayuda del software SPSS 19.

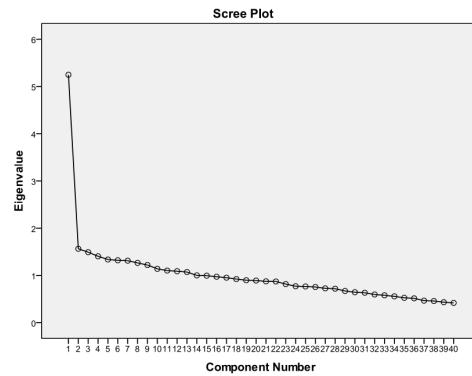


Figura 1.1 Gráfico de sedimentación de los ítems de HC

La figura 1 muestra un “codo” claramente marcado en el componente número 2, lo cual es evidencia de que la prueba presenta un carácter unidimensional; además, el análisis factorial indicó que el primer factor explica un 13,1% de la variabilidad de los ítems, mientras que el segundo apenas alcanza a explicar un 3,9% de varianza, lo cual sugiere que hay una diferencia sustantiva entre la variabilidad explicada por el primer factor con respecto al segundo. Por estas razones, se aceptó el supuesto de la unidimensionalidad.

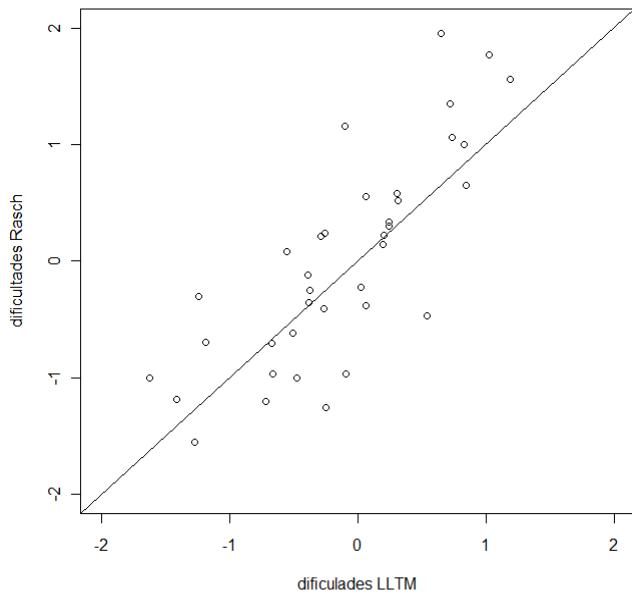
Además se omitieron de la base de datos las personas y los ítems que no se ajustaban al modelo de Rasch (infit mayor a 1.2 o menor a 0.8), con lo cual la base de datos se redujo a 356 individuos y 37 ítems.

El segundo supuesto del modelo LLTM es que estime dificultades similares a las dadas por el modelo de Rasch. Para evaluar este supuesto se usó la prueba chi cuadrado de Andersen, la cual, como en la mayoría de investigaciones a nivel internacional, produjo un estadístico significativo y el correspondiente rechazo de la hipótesis de igualdad entre las dificultades estimadas por los dos modelos. Sin embargo, al calcular la correlación de Pearson entre las dificultades estimadas por el modelo LLTM con las estimadas por el modelo de Rasch, se obtuvo un coeficiente de correlación de 0.80, lo cual indica que las dificultades se asocian fuertemente de manera lineal.

En la figura 2, se presenta un gráfico de dispersión construido con el software R 2.14.0, en el que se contrastan las dificultades de cada ítem dadas por los dos modelos. Se aprecia una fuerte asociación lineal entre ambas dificultades, de hecho la asociación es dada por la recta identidad, lo cual indica que las dificultades son bastante similares.

#### 1.4.3 Modelos obtenidos

Con el fin de tener un parámetro de comparación con el LLTM, se decidió aplicar previamente un modelo de regresión lineal múltiple considerando los atributos de los ítems como variables independientes y la dificultad Rasch como variable dependiente. Los coeficientes asociados a este modelo se observan en la tabla 2. En esta regresión lineal se debe obtener la dificultad Rasch a priori, a diferencia del LLTM, en el cual se estima la dificultad a partir de las contribuciones de cada uno de los atributos de los ítems. Este modelo se ejecutó con STATA 11, mientras que



**Figura 1.2** Dificultades dadas por los modelos LLTM y Rasch de los ítems de la prueba de HC

la dificultad Rasch se calculó con Winsteps.

En este modelo, solo la variable *Discriminación de opciones* no resultó significativa a un nivel del 5%. Asimismo, el modelo presentó coeficientes positivos en las variables *Ordenamiento*, *Visualización*, *Cantidad de contenidos*, *Planteamiento de relación* y *Cantidad de algoritmos algebraicos*, lo cual indica que contribuyen a aumentar la dificultad de los ítems.

Además, los ítems de las áreas *Análisis de datos*, *Geometría* y *Álgebra* son en promedio más fáciles que los ítems de *Aritmética*; por otro lado, los reactivos de selección única son en promedio más sencillos que los de comparación cuantitativa. Finalmente, cabe señalar que al aumentar el número de procesos deductivos e inductivos, la cantidad de hipótesis y procesos aritméticos, la dificultad de los ítems disminuye.

#### Coeficientes de la regresión lineal contra la dificultad Rasch de los ítems

Este modelo de regresión presentó un porcentaje de variancia explicada de un 63,80%, lo cual indica que un 63,80% de la variabilidad de la dificultad de los ítems de la prueba de Habilidades Cuantitativas es explicada por las variables tomadas en cuenta para el modelo.

En la tabla 3 se presentan los resultados arrojados por el modelo LLTM, el cual fue estimado con la ayuda del software R 2.14.0. En este paquete los coeficientes negativos indican que las variables contribuyen a aumentar la dificultad Rasch y los positivos contribuyen a la disminución de la dificultad.

En este modelo se presentaron solamente tres variables no significativas a un nivel del 5%: *Ordenamiento*, *Discriminación de opciones* y *Algoritmos aritméticos*. Los atributos de los ítems que contribuyen a aumentar la dificultad son los siguientes: *Visualización*, *Cantidad de contenidos*, *Planteamiento de relación*, *Cantidad de algoritmos algebraicos* y el hecho de que los ítems sean de comparación cuantitativa. Por otro lado, los atributos que contribuyen a hacer más fáciles los ítems son la presencia de procesos deductivos e inductivos y la cantidad de hipótesis; además, se observó que en promedio los ítems de las áreas *Análisis de datos*, *Geometría* y *Álgebra* son más fáciles que los de aritmética.

Finalmente, se decidió estimar un modelo multinivel cuya variable dependiente fuera el acierto en el ítem, el nivel inferior fuera el examinado y el superior, el ítem. Este modelo se planteó con dos fines: 1) examinar la relación entre la dificultad del ítem y el acierto de este, y 2) determinar cuánto de la variabilidad de las respuestas en los ítems es explicada por la variabilidad de los ítems. Este modelo se estimó con STATA 11 y sus coeficientes se presentan en la tabla 4. Este modelo presentó solo tres variables no significativas, las cuales fueron las mismas que aparecieron

**Tabla 1.2**

	Coef	E.E.	t	P>t	[95% Int.]	confianza]
Deducción	-0.539**	0.012	-44.110	0.000	-0.563	-0.515
Inducción	-0.752**	0.014	-55.440	0.000	-0.778	-0.725
Ordenamiento	0.075**	0.011	6.800	0.000	0.053	0.096
Visualización	0.419**	0.009	48.280	0.000	0.402	0.436
Discriminación de opciones	-0.018	0.017	-1.030	0.302	-0.051	0.016
Número de hipótesis	-0.108**	0.010	-11.310	0.000	-0.127	-0.090
Número de contenidos	0.346**	0.008	42.770	0.000	0.330	0.362
Planteamiento de relación	0.449**	0.015	30.900	0.000	0.420	0.477
Algoritmos algebraicos	0.597**	0.014	42.300	0.000	0.569	0.625
Algoritmos aritméticos	-0.035**	0.010	-3.600	0.000	-0.053	-0.016
Algebra	-1.768**	0.025	-70.720	0.000	-1.817	-1.719
Geometría	-1.173**	0.024	-48.030	0.000	-1.221	-1.126
Análisis de datos	-0.327**	0.023	-14.050	0.000	-0.373	-0.282
Comparación Cuantitativa	0.791**	0.024	32.470	0.000	0.743	0.839
Constante	-0.093**	0.027	-3.380	0.001	-0.147	-0.039

\*\* Significativas a un nivel del 5%

**Tabla 1.3** Coeficientes del modelo LLTM

	Coef	E.E.	[95% Int.]	confianza]
Deducción	0.490**	0.050	0.392	0.587
Inducción	0.693**	0.056	0.582	0.803
Ordenamiento	-0.085	0.044	-0.171	0.000
Visualización	-0.406**	0.035	-0.476	-0.337
Discriminación de opciones	0.069	0.066	-0.060	0.199
Número de hipótesis	0.097**	0.038	0.022	0.172
Número de contenidos	-0.313**	0.032	-0.377	-0.250
Planteamiento de relación	-0.395**	0.057	-0.507	-0.283
Algoritmos algebraicos	-0.573**	0.058	-0.686	-0.460
Algoritmos aritméticos	0.011	0.038	-0.064	0.085
Algebra	1.693**	0.102	1.493	1.893
Geometría	1.112**	0.099	0.918	1.306
Análisis de datos	0.260**	0.094	0.077	0.444
Comparación Cuantitativa	-0.676**	0.096	-0.864	-0.488

\*\* Significativas a un nivel del 5%

como no significativas a un nivel del 5% en el LLTM: *Ordenamiento, Discriminación de opciones y Algoritmos aritméticos*

Se puede inferir de la tabla 4 que si se aumenta la cantidad de procesos deductivos e inductivos y de hipótesis, entonces se aumentan los *odds* (división de la probabilidad de éxito por la probabilidad de fracaso) de acertar el ítem. Sin embargo, si se aumentan los pesos de las variables *Visualización, Cantidad de contenidos, Planteamiento de relación* y *Cantidad de procesos algebraicos*, entonces los *odds* de acertar el ítem disminuyen. Finalmente, los *odds* de acertar un ítem de selección única son mayores que los de contestar correctamente un ítem de comparación cuantitativa, mientras que los *odds* de acertar un ítem de álgebra, de análisis de datos o de geometría son mayores que los de acertar uno de aritmética.

Este modelo indicó que solamente un 13,61% de la variancia en el acierto de los ítems es debido a características de los ítems y que más de un 85% de la variancia del acierto en los ítems es debido a características de las personas.

**Tabla 1.4** Coeficientes de la regresión logística multinivel contra el acierto en los ítems

	A. Odds	E.E.	t	P>t	[95% Int.]	confianza]
Deducción	0.631**	0.081	9.820	0.000	1.479	1.798
Inducción	1.001**	0.113	12.280	0.000	1.792	2.236
Ordenamiento	-0.082**	0.040	-1.970	0.049	0.842	1.000
Visualización	-0.334**	0.023	-11.550	0.000	0.622	0.714
Discriminación de opciones	0.071	0.071	1.040	0.299	0.941	1.220
Número de hipótesis	0.101**	0.042	2.530	0.011	1.022	1.187
Número de contenidos	-0.269**	0.024	-9.670	0.000	0.686	0.779
Planteamiento de relación	-0.327**	0.039	-6.930	0.000	0.601	0.752
Algoritmos algebraicos	-0.435**	0.033	-9.910	0.000	0.505	0.633
Algoritmos aritméticos	0.011	0.038	0.290	0.773	0.938	1.089
Algebra	4.421**	0.552	16.600	0.000	4.440	6.619
Geometría	2.036**	0.300	11.250	0.000	2.502	3.684
Análisis de datos	0.302**	0.122	2.820	0.005	1.084	1.565
Comparación Cuantitativa	-0.491**	0.049	-7.050	0.000	0.422	0.614

A.Odds: aumento en los odds de acertar el ítem si se aumenta en una unidad la variable.

\*\* Significativas a un nivel del 5%

## 1.5 Discusión

---

Al comparar el modelo lineal simple con el LLTM, se puede observar que en ambos la variable *Discriminación de opciones* no es significativa y que los atributos que ayudan a aumentar la dificultad de los ítems son la visualización, la cantidad de contenidos, el planteamiento de relación y la cantidad de algoritmos algebraicos.

El hecho de que la visualización y el planteamiento de relación contribuyan a aumentar la dificultad de los ítems de HC coincide con lo hallado en Embretson y Daniel (2008). En esta investigación se obtuvo que la visualización y la generación de una nueva ecuación son predictores significativos de la dificultad de los ítems del GRE (*Graduate Record Examination*).

Estos resultados también coinciden con la Teoría de los Tres Estratos, ya que, según Carroll (1993), el componente visoespacial y la cantidad de computación algebraica colaboran en aumentar la dificultad de los ítems, lo cual es coherente con lo obtenido en este estudio, ya que en los dos modelos estimados para la dificultad se evidenció que la visualización y los algoritmos algebraicos tienden a aumentar la dificultad de los ítems.

Sin embargo, en el estudio de Embretson y Daniel (2008), la variable “conteo de computaciones” relacionada con la cantidad de algoritmos necesarios para hallar la solución del ítem no fue significativa, contrario a lo hallado para HC, a la teoría de Carroll y a lo esperado por los autores, ya que estos presentan el “conteo de computaciones” como una variable asociada a la dificultad de los ítems.

Una situación similar a la de Embretson y Daniel (2008) se presentó en los resultados obtenidos para HC. Los modelos LLTM y de regresión lineal simple coincidieron en que los atributos que ayudan a disminuir la dificultad de los ítems son la presencia de procesos deductivos e inductivos en los ítems, lo cual contradice los resultados de las investigaciones sobre el razonamiento (De Juan-Espinoza, 1997), en particular las asociadas a la Teoría de los Tres Estratos, que es la base sobre la cual se construyó la prueba de Habilidades Cuantitativas.

La primera hipótesis que se puede plantear para explicar estos coeficientes es el error de medición. Debido a que los procesos deductivos e inductivos que realizan los examinados para llegar a la solución de un ítem son muy variados, y que por lo general un juez experto no razona igual que un examinado, podría suceder que los pesos

asignados por los jueces a los procesos de razonamiento de los ítems no coincidan con la realidad de los examinados. No obstante, si esta fuera la razón por la cual se obtuvieron estos valores, ¿por qué los coeficientes son significativamente diferentes de cero? Una explicación alternativa podría ser que los procesos de inducción y deducción estén involucrados en el planteamiento de relación, la discriminación de opciones, el ordenamiento o la realización de algoritmos, de modo que el efecto de ambos podría estar siendo “absorbido” por otras variables.

Ante esta nueva hipótesis, Romero et al. (2008) mencionan que una “limitación del modelo LLTM es su carácter compensatorio pues la descomposición de los parámetros de dificultad es lineal, en la práctica esto implica que la ejecución correcta de una operación compensa la ejecución incorrecta en otra”. Los autores exponen esta idea sobre la estructura del modelo debido que, al igual que con HC y con el GRE, el efecto significativo de una variable resultó ser inverso a la dirección esperada.

Con base en las ideas anteriores, no se puede concluir que el incremento de los procesos deductivos e inductivos en los ítems de HC tienda a facilitar los ítems. Se recomienda para próximos estudios donde se tomen en cuenta estas variables no incluir aquellas que estén asociadas tan directamente a los procesos de razonamiento deductivo e inductivo.

En los modelos donde la variable de respuesta es la dificultad de los ítems se mostró que los ítems de álgebra, análisis de datos y geometría son en promedio más fáciles que los de aritmética. Este resultado parece erróneo, pero al analizar la prueba de HC, este resultado se explica porque los ítems de aritmética incluidos en la prueba son menos familiares para los examinados que los de las otras áreas de contenido.

También se obtuvo que los ítems de comparación cuantitativa son en promedio más difíciles que los de selección única, lo cual coincide con lo expuesto anteriormente sobre el uso de formatos diferentes en los ítems. Además, se evidenció que la cantidad de hipótesis dadas en el problema contribuyen a facilitar el ítem, lo cual era esperado debido a que el examinado tiene más información para llegar a la conclusión.

Por otro lado, cuando se comparan los resultados del LLTM con los obtenidos en el modelo de regresión logística multinivel, se puede notar que se llega a las conclusiones esperadas: las mismas variables que ayudan a aumentar la dificultad son las que provocan una disminución en la probabilidad de acierto y aquellas que reducen la dificultad son las que contribuyen a aumentar los *odds* de acierto en un ítem.

Otro resultado importante para el Proyecto de Habilidades Cuantitativas es el hecho de que la variabilidad de los aciertos sea explicada en menos del 14% por los atributos de los ítems. Este coeficiente está indicando que, en su mayoría, el acierto de los ítems es debido a características propias de las personas, entre las cuales se espera que su habilidad cuantitativa tenga un papel preponderante, lo cual indicaría que el acierto del ítem no depende tanto de cómo es el ítem sino de la habilidad cuantitativa de la persona. En este sentido, surge la pregunta sobre por qué realizar un estudio en dónde se intente predecir la dificultad de los ítems si se quiere que la variabilidad del acierto no dependa de la variabilidad de los atributos de los ítems. La respuesta es sencilla y es que, según lo establecido por el modelo de Rasch, la dificultad de un ítem está en la misma escala que el nivel de habilidad de una persona, produciendo de esta manera mayores probabilidades de acierto para las personas que tengan un nivel de habilidad mayor al nivel de dificultad del ítem en comparación con las que tienen niveles de habilidad inferiores (Prieto y Delgado, 2003). Entonces, aunque la variabilidad del acierto del ítem es explicada mayoritariamente por la de las características de las personas, las probabilidades de acierto de los ítems pueden ser controladas de manera que favorezcan a la población meta.

Así las cosas, según los resultados obtenidos en el modelo de regresión lineal simple, casi un 65% de la variabilidad de la dificultad Rasch es explicada por los atributos planteados. Por tal motivo, se estaría alcanzando de manera satisfactoria el objetivo inicial de plantear un modelo estadístico que permita predecir la dificultad de los ítems de la prueba de Habilidades Cuantitativas de la Universidad de Costa Rica.

Queda ahora la tarea para el Proyecto de Habilidades Cuantitativas emplear los modelos acá establecidos para construir ítems que permitan seleccionar a la población con el nivel de habilidad cuantitativa requerido para cada carrera. Además, en próximas investigaciones se probarán dichos modelos al contrastarlos contra otros formularios

de la prueba que han sido aplicados recientemente.

## Bibliografía

---

- [1] Álvarez, A., Molina, M. y Salazar, A. (2000). *Competencias Académicas y otras dimensiones asociadas en estudiantes y según profesores e investigadores de diversas áreas de la Universidad de Costa Rica* (Informe final de investigación). Universidad de Costa Rica.
- [2] Agresti, A (2002). *Categorical data analysis*. New York: Wiley.
- [3] American Educational Research Association, American Psychological Research Association y National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Estados Unidos. American Educational Research Association. Attorresi, H., Picón, J., Abal, F., Aguerri, M., y Galibert, M. (2009). Aplicación del modelo LLTM de Fischer al análisis de las fuentes de dificultad de ítems de razonamiento deductivo. *Interdisciplinaria*, 26 (1), 77-93.
- [4] Carroll, J. B. (1993). *Human Cognitive Abilities: A survey of factor analytic studies*. USA. Cambridge University Press.
- [5] Dwyer, C., Gallagher, A., Levin, J. y Morley, M. (2003). *What is Quantitative Reasoning? Defining the Construct for Assessment Purposes*. Estados Unidos. ETS.
- [6] Embretson, S. y Daniel, R. (2008). Understanding and quantifying complexity level in mathematical problem solving items. *Psychology Science Quarterly*, 50 (3), 328-334.
- [7] Fischer, G., H. (1973). The linear logistic test model as an instrument in educational research. *Acta psychologica*, 37, 359-374.
- [8] Kutner, M., Neter, J., Nachtsheim, C. y Wasserman, W. (1996). *Applied linear statistical models*. Estados Unidos: McGraw-Hill.
- [9] Mayer, R. R., Larkin, J., y Kadane, J. B. (1984). A cognitive analysis of mathematical problem solving ability. En R. Sternberg (Ed.), *Advances in the psychology of human intelligence*, V2 (pp 231-273).
- [10] Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, 50, 741-749.
- [11] Montero, E., Valverde, A. y Villalobos, J. (2007) Factores institucionales, pedagógicos, psicosociales y sociodemográficos asociados al rendimiento académico en la Universidad de Costa Rica: Un análisis multinivel. *Relieve*, 13(2), 215-234.
- [12] Muñiz, J. (2010). Las teorías de los test: la teoría clásica de los test y la teoría de respuesta al ítem. *Papeles de psicólogo*, 31 (1), 57-66.
- [13] Oliver, J., Rosel, J y Jara, P (2000) Modelos de regresión multinivel: aplicación en psicología escolar. *Psicothema*, 12 (3), 487-494
- [14] Pardo, A., Ruiz, M. y San Martín, R (2007). Cómo ajustar e interpretar modelos multinivel con SPSS. *Psicothema*, 19 (2), 308-321
- [15] Prieto, G. y Delgado, A. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15 (1), 94-100.
- [16] Proyecto de Habilidades Cuantitativas. (2010). *Construcción y validación de un banco de ítems y pruebas de habilidades cuantitativas para diagnóstico e ingreso a carrera en la Universidad de Costa Rica*. Documento oficial del proyecto.
- [17] Proyecto de Habilidades Cuantitativas. (2011). *Categorización de los ítems de la prueba de Habilidades Cuantitativas*. Documento oficial del proyecto.
- [18] Real, E., Olea, J., Ponsoda, V., Revuelta, J. y Abad, F. (1999). Análisis de la dificultad de un test de matemáticas mediante un modelo competencial. *Psicológica*, 20, 121-134.
- [19] Romero, J., Ponsoda, V. y Ximénez, C. (2006). Validación de la estructura cognitiva del test de signos mediante modelos de ecuaciones estructurales. *Psicothema*, 18 (4), 835-840.
- [20] Romero, J., Ponsoda, V. y Ximénez, C. (2008). Análisis de un test de aritmética mediante el modelo logístico lineal de rasgo latente. *Revista Latinoamericana de Psicología*, 40 (1), 85-95.
- [21] Villareal, M. (2011). *Construcción de un modelo psicométrico cognitivo para una prueba de inteligencia fluida*. Trabajo final de graduación para optar al grado de máster en evaluación educativa de la Universidad de Costa Rica.
- [22] Villalobos, J. y Molina, M. (2010). *Ánalisis de la estructura subyacente de la Prueba de Aptitud Académica (PAA) de la Universidad de Costa Rica* (Informe final de investigación). Universidad de Costa Rica.